A Taxonomy of Machine Learning Approaches for Obstructive Sleep Apnea Detection and Management in Pregnancy: Methods, Challenges, and Future Directions

(IJCSNT) International Journal of Communication Systems and Network Technologies
ISSN Number: 2053-6283
Volume 15, Issue No. 1, 1–32

https://journals.mirlabs.in/index.php/ijcsnt
DOI-10.18486/ijcsnt/15.1.001
IJCSNT

Vandana Roy[1], Shalini Stalin[2], Nikhar Vishwakarma[3], Anuj Nayak[4]

Abstract

Despite the fact that there are well recognised links between OSA and serious outcomes to both mother and baby, OSA is one of the more under-recognised conditions in maternal medicine. Many physiological changes during pregnancy such as hormonal fluctuations, enlarged airways, and increased blood flow put the entire body at greater risk for sleep-disordered breathing, and make it difficult to identify with common screening questionnaires used for the general adult population. This gap has clinical implications and solutions to this challenge must be more than symptom checklists. More than 100 papers were reviewed and 38 were selected that directly contribute information to the taxonomy that we propose. The range of performance in the reviewed models extends from the traditional classifiers based on questionnaires (accuracy between 0.79 and 0.85), to the models based on deep learning of the signals (accuracy between 0.92 and higher, and some reported values of sensitivity up to 0.977). These statistics not only show technical advancements, but also the entire change of the capacity of automated OSA detection. Multivariate models that use physiological signal, body measures and clinical data all in one are consistently superior to models using any single one of these signals. New explainability tools like SHAP and LIME are becoming more integrated in these systems and are helping to close the gap between a model prediction and how a clinician can act on that prediction. This extends further to wearable and contactless sensor platforms which allow monitoring outside the sleep lab. In parallel, it highlights ongoing issues: the lack of standardization of training data, the lack of consensus regarding the definition of apnea-hypopnea threshold and the lack of prospective validation in a wide range of obstetric populations. Shared benchmarks, light and clear architectures, and an active focus on fair model design will be key to be able to make progress.

Keywords

EEG, OSA, ML, AI, DL

Received: 27 December 2025; **Revised:** 15 April 2026; **Accepted:** 26 April 2026;
Published: 30 April 2026;

[1]DoEC, Gyan Ganga Institute of Technology and Sciences, Jabalpur, Madhya Pradesh, India.
[2]Department of Information Technology, IIIT Bhopal, Madhya Pradesh, India.
[3]Department of Pharmacy, Gyan Ganga Institute of Technology and Sciences, Jabalpur, Madhya Pradesh, India.
[4]Gyan Ganga Institute of Technology and Sciences, Jabalpur, Madhya Pradesh, India.

Corresponding author:
Email: Vandana.roy20@gmail.com

Introduction

Sleep is a basic process involved in biology which is vital in metabolic control, heart stability, mental faculties, as well as immune balance[1]. Sleep is required to achieve autonomic balance, hormonal regulation and neurocognitive functioning throughout thelifespan[1]. Sleep architecture disruption or respiratory sleep disturbance has been linked to a high risk of cardiometabolic diseases and neurocognitive impairment . The prolonged sleep disturbance has also been associated with systemic inflammation and endothelial dysfunction that cause long-term cardiovascular morbidity[2].

The most common sleep-related breathing disorder is obstructive sleep apnea (OSA), which is a significant health issue within the population. OSA is a condition that is defined by frequent upper respiratory airway blockage when sleeping, which leads to periodic hypoxemia, sleep disruption, and arousals. Such physiological disruptions cause the activation of the sympathetic nervous system and oxidative stress, which are the major mechanisms underlying OSA-related cardiovascular and metabolic complications[3].

Pregnancy is a unique physiological condition where hormonal, anatomical and cardiopulmonary changes can predispose to OSA. Drug interactions: fluctuation of progesterone and estrogen in pregnancy has an effect on respiratory drive and upper airway muscle tone[4]. It has also been postulated that gestational weight gain is linked with changes in upper airway mechanics and decreased lung volumes that are likely to predispose individuals to sleep-disordered breathing. The reduction of functional residual capacity and expiratory reserve volume are additional factors that lead to respiratory susceptibility in pregnant women during sleep[5]. Airway lumen may also be constricted by upper airway mucosal edema associated with the retention of fluid and vascular permeability. Changes in ventilatory response during pregnancy also contribute to the instability of the respiratory system during sleep[6]. Epidemiological surveys have shown that the prevalence rate of sleep-disordered breathing rises gradually through gestation. The rise is specifically obvious in the second and third trimesters, when the physiological and anatomical changes become more prominent[7]. The occurrence of OSA is significantly higher in high-risk pregnancies than in low-risk pregnancies, particularly in cases of chronic hypertension or obesity complications .Maternal obesity has been a consistent factor in the condition of upper airway collapsibility and decrease of respiratory reserve[8]. It has further been suggested that gestational diabetes is also linked with high rates of obstructive sleep apnea in pregnancy indicating that they may share common metabolic risk pathways[9]. Figure 1 provides a conceptual illustration showing the difference between Normal Breathing and an Apneic Event ,which leads to Periodic Hypoxia, Arousals, and Sleep Discontinuity.

OSA has been found to be caused by an untreated infection in pregnant women who are at a higher risk of developing hypertensive disorders such as preeclampsia and gestational hypertension. It is believed that intermittent hypoxemia and sympathetic stimulation could lead to endothelial pathology and blood pressure malregulation in this group of people[10]. Maternal OSA has also been associated with compromised glucose metabolism and enhanced insulin resistance which may exacerbate metabolic complications during pregnancy[11]. Besides maternal outcomes, maternal sleep-disordered breathing has been found to have adverse fetal outcomes. Potential consequences include low birth weight and preterm birth, contributing to high neonatal morbidity ratesand long-term health consequences[12].

Identifing the scale of these risks like maternal and fetal risks that makes the absence of a reliable pregnancy-specific pathway of diagnosis all the more pressing.OSA has not been significantly diagnosed among pregnant individuals though it has clinical significance. The diagnostic gold standard for OSA

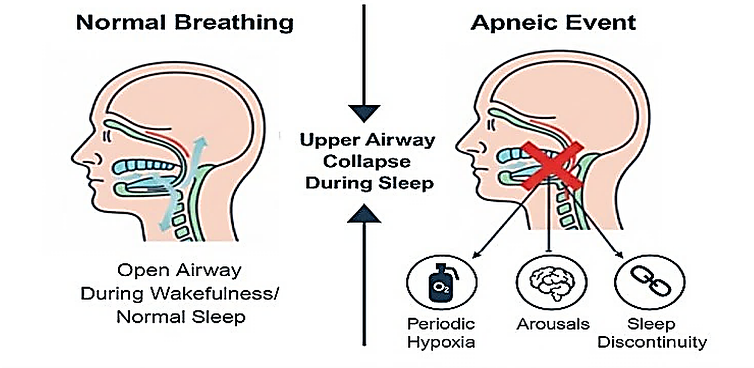

Figure 1. Mechanism of Obstructive sleep apnea(OSA)

is polysomnography (PSG) which is resource-intensive, time-consuming, and is not always feasible during pregnancy, especially in late pregnancy. There are also logistical constraints and low access to sleep laboratories and the use of PSG is further limited due to patient discomfort in obstetric care[13]. Subsequently, screening questionnaires have been broadly applied as alternative tests in a prenatal context. Nonetheless, the STOP-Bang questionnaire has shown poor diagnostic accuracy when used on pregnant cohorts partly because of the overlap of the questionnaire items with normal gestational symptoms like fatigue and weight gain[14]. The same limitations have alsobeen reported regarding the Berlin Questionnaire in obstetric populations, but this questionnaire has lower specificity within studies that concentrate on pregnancy[15]. The Epworth Sleepiness Scale has also demonstrated low discriminatory performance in the screening of OSA in pregnancy, probably because of baseline gestational fatigue and sleep disturbances[16].

The continous underperformance of questionnaire-based tools has shifted research attention toward computational approaches that can help inanalyse objective physiological signals instead of relying on self-reported symptoms. Artificial Intelligence (AI) and Machine Learning (ML) models are very advanced computational frameworks which are made to understand and find complex patterns within data to perform tasks that require human intellegence, for instance; classification and prediction.Traditional ML models utilizes “handcrafted” physiological features, which can be particular markers like heart rate variability or oxygen drops, that are set by researchers that alter the algorithm’s logic.The Traditional systems are now moving towards the concept of Deep Learning (DL) and architectures which are Convolutional Neural Networks (CNNs) and Long Short-Term Memory networks (LSTMs) automatically take these features out directly from raw physiological signals.The dive from manual engineering to autonomous feature extraction opens a clear path to the high-resolution information and its processing, with much greater mathematical accuracy and much lower human bias.These models are vital as they analyse large, complex datasets, and frequently outperform subjective clinical interpretation by providing diagnostic accuracy.Advanced AI can detect subtle respiratory dependencies that are hard to confine from traditional monitoring methods by analyzing the characteristics of the signal in both space and time.But as these models get more complex, the underlying decision making process of the model is more complex, and harder for humans to follow directly.The role of Explainable AI (XAI) is very important in lending the transparency needed to transform high accuracy prediction into authentic medical evidence that can be trusted by medical experts.In OSA detection in pregnancy, these models offer the potential for a paradigm shift in the way OSA is monitored, from invasive to continuous.AI can accurately identify the risk of apnea in a natural sleep setting, which is much more practical and comfortable for patients, especially those who are pregnant, than lab studies.

Combined with XAI tools, high-performance tools like SHAP and LIME can provide an "interpretability bridge" which can justify the diagnosis for the obstetrician.This not only helps to ensure the system is accurate, but it also helps to ensure that the system can be implemented at home immediately and that interventions and prescription of treatment can be made, while considering the high level of safety and accountability necessary in maternal health.The recent progress in machine learning (ML) and artificial intelligence (AI) has made it possible to generate good amount of automated methods for OSA detection that are based on objective physiological measurements.

Models based on photoplethysmography are able to identify oxygen desaturation patterns which are having links with apnea without intruding upon the patient[17].The main cause of the autonomic changes associated with apneic events is the sympathetic response to oxygen desaturation and carbon dioxide elevations.A phenomenon called "cyclical variation of heart rate" is observed during an episode of Obstructive Sleep Apnea (OSA), which involves a slowing down of heart rate (bradycardia) followed by a swift rise in heart rate (tachycardia) when awakening from sleep.Researchers can accurately predict these cardiac signatures which are having specific breathing pauses using features obtained from electrocardiography (ECG).Features that are obtained from electrocardiography (ECG) can be used to accurately map these cardiac signatures to specific breathing pauses.This is a technique which can help to identify sleep fragmentation and physiological stress without any kind of need for invasive measures from heart rhythm information.These extracted ECG data attributes are essential inputs to ML models for normal sleep classification and apneic episode classification in modern research[18]. A promising alternative for the non-invasive respiratory monitoring is the contact-less respiratory monitoring that is achieved by using radiometric and motion based systems, which are appropriate for home-based sleep monitoring[19].The performance of automated OSA detection systems has been greatly improved by deep learning architectures that operate on raw data in a more effective manner.A convolutional neural network (CNN) is well-suited for detecting complex patterns in high dimensional physiological signals, like ECG or SpO2 waveforms.This type of architectures do not require handcrafting features and instead learn the relevant data representations from the input , that are used in Traditional Machine Learning architecture.This automated feature discovery enables the model to detect any kindof subtle changes in the body which the human designed metrics might not be able to identify.These deep layers are often used forobtaining higher accuracy and more powerful outcomes for a variety of patient sets.This end-to-end learning is a dramatic technology evolution from classical signal processing techniques.[20]. Recurrent neural networks have helped to better temporal model respiratory and cardiac dynamics by modeling sequential temporaldependencies in time-series data[21].Plan XAI systems to solve crucial problems regarding clinical trust and transparency in automatic detection.

Although machine learning has started to make significant strides in the field of OSA detection in the general adult population, it has not been as quickly embraced within obstetric medicine.The number of studies specific to pregnant women is limited and when they are available, there is significant methodological variation in the acquisition of data on physiological signals, in the targeting and recruitment of cohorts and in assessment of model performance, making it difficult to draw direct comparisons between different studies and pooled conclusions are unreliable [23].One of the major issues is that the majority of the proposed detection models developed and validated within the frame of the general sleep clinic samples and no external validation has been done in the pregnant population.Without prospectively testing this assumption, it can generated some sort of element of uncertainty to the existing literature that may not be ignored when considering that these models may not apply to the physiological and hormonal changes that occur during gestation.To further complicate things, there is some kind of general pathologizing of the field's appraisal of progress.Reviewed automated OSA detection methods have been directed towards accuracy, sensitivity and area under the curve as important indicators of the usefulness of a model, but have been largely ignored in their consideration of the feasibility of using a model in an antenatal care setting [24].These benchmarks, alone, do not provide clinical utility of the tool.But questions regarding how the detection system fits into the midwife or obstetrician's work-flow, what do to if the signal is not ideal in a noncontrolled recording environment, and whether the output is presented in a way that facilitates, not hinders, clinical decision-making, are just as critical as the accuracy figure and have been a secondary concern for too long.The field should be on these aspects of implementation and it is not unscientific.To be useful, the technologies need to be demonstrated to work in real laboratory environments, and reliable and easy to use in the settings where pregnant women are seeking medical care.Based on this view, the present review proposes a systematic classification of OSA detection and management strategies based on ML and AI, to categorized the state of the art, pinpoint key research gaps that impede the clinical translation and guide future research directions to bring OSA from the lab to the clinical setting.

1.1 Maternal Risk Factors and Pathophysiological Contributors to Sleep Apnea

The pathogenesis of OSA in pregnancy is condenced in a complex interplay of anatomic, physiological and metabolic changes that characterize the pregnancy.Weight gain induces weight deposition around the neck and upper airway, which causes a naturally constricted airway to become severely reduced in diameter when the patient becomes unconscious [5].These changes are accompanied by increased levels of progesterone and an increase of estrogen, that each have unique effects on respiratory patterns.Chronic hyperventilation is often accompanied by a reduction of arterial carbon dioxide in response to the high ventilatory drive that is stimulated by the effects of progesterone.But at the same time, rising estrogen levels result in a mucosal expansion of the upper airway passages, and congestion, effecting resistance and muscle tone.All of these counteracting forces create instability to the respiratory system, making pregnant women more susceptible to sleep-disordered breathing.To develop screening tools specific to maternal health, it is essential to understand how these factors work together.In fact, by tracing these changes, it can be possible to foresee variations in the stability of airways throughout pregnancy [4].There is also a lot of interstitial fluid accumulation due to systemic fluid retention and alterations of vascular permeability.This can lead to swelling of the upper airway's soft tissues, which can shrink the diameter of the airway.The mother has increased resistance to air flow and higher risk of collapse of the lungs; usually in the form of snoring or identifiable apneic periods.This limitation of space increases the sensitivity of the airway to narrowing due to gravity's effect on the airway when the patient is supine.Thus it is imperative to have control of these fluid related changes to maintain respiratory stability [6].These complications are compounded by a decrease in functional residual volume and expiratory reserve volume, which also reduces the oxygen buffer, and thus extends the risk of hypoxemia during an apnea [5].Maternal health changes even more with the onset of changes in ventilatory control mechanisms.An alteration in chemosensitivity to CO_2 and O_2 may disrupt the natural breathing pattern, often causing an excessive response to CO_2 and/or O_2, which may lead to a poor quality of a steady state.This increased chemical sensitivity combined with the narrowing of the airways results in irregular breathing and many night-time awakenings.These chemical and neurological recalibrations work together with anatomical changes to increase the likelihood of having OSA[6].

In addition to these factors, metabolism plays a very big role in the progression of the disorder. Maternal obesity is one of the major factors, and this is related to the airway collapsibility, which relates to the deposition of fat tissue. Additional weight around the neck and chest also reduces respiratory compliance, and thereby decreases the flexibility of the lungs and chest wall. This means that the effort of breathing is greater, especially when lying down. Metabolic stress and structural changes together result in a persistent stage of low lung volume and high airway resistance throughout the night, that is, throughout the sleep cycle[8]. Another factor associated with disordered breathing is gestational diabetes (GD), possibly via related pathways of insulin resistance, inflammatory markers and autonomic control[9]. Chronic hypertension may worsen these effects through the course of pregnancy and cause impaired vascular function and inadequate oxygen supply during sleep. The structural changes of a compromised blood vessel structure make it more difficult for the body to increase the diameter of vessels and control flow, which gives less ability to compensate for the oxygen desaturations of respiratory pauses. This vascular disturbance increases the physiological stress of mother and fetus. This effect on the cardiovascular system can create a vicious cycle, which exacerbates respiratory instability, and thus control of blood pressure is an important part of minimizing the effect of sleep-disordered breathing[8]. Lastly, there are a number of sleep-related factors that are unique to the period of gestation, which add to the risk profile. There is increased resistance in the upper airway and a rise in incidence of nasal congestion and pregnancy related rhinitis, while the supine position, which is difficult to avoid as pregnancy Progresses, can worsen upper airway obstruction and venous return[6,7]. These factors are maternal influences that create a physiological environment that increases the risk of OSA, requiring prompt detection and ongoing surveillance.

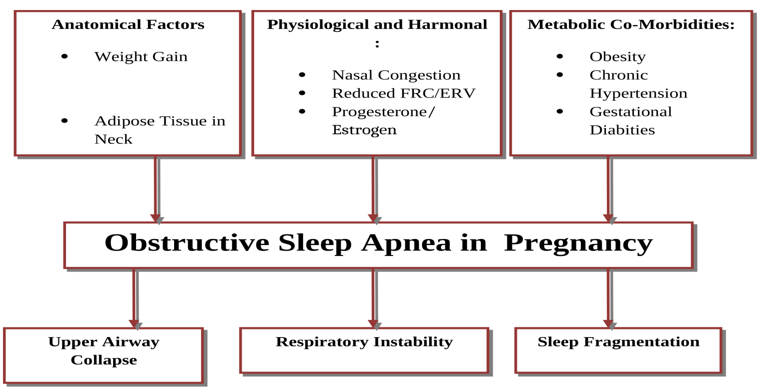

Figure 2. Maternal Factors and Pathophysiological Consequences of Obstructive Sleep Apnea During Pregnancy

Figure 2: The most important maternal complication that may be a factor in the development of obstructive sleep apnea during pregnancy. Various anatomical, physiological–hormonal and metabolic co-morbid factors work together to make one person more susceptible to OSA which culminates in collapse of the upper airway, instability of the respiratory system and discomfort during sleep.

Machine Learning and Deep Learning Techniques for Obstructive Sleep Apnea Detection in Pregnancy

them to be adequately executed on the type of hardware that's available in low-resource settings. They can't, however, detect the non-linear, interdependent patterns that lie within raw physiological waveforms. While a questionnaire-based approach is a good first step, a baseline to which more architecturally complex models can be compared, it is not the only approach to be considered and should not be used as a screen on which to diagnose. This balance between interpretability and analytical depth is not an easy one to achieve and is close to the heart of the methodological discussion which is still ongoing in this area[15].

However, it is not only because of its impracticality, but also due to its special vulnerability in obstetric populations that restricts its diagnostic value. Many of the symptoms these instruments seek to detect are actually caused by pregnancy itself, which systematically increases false-positive rates, closer to what would be considered an acceptable rate in a confirmatory test[16]. Instead, the better use of these tools is to be used as a risk stratification tool in the population: to find out which pregnant women should be further physiologically investigated and which are not a candidate for such investigation[16].Physiological signal data is the most extensively studied modality of automated OSA detection and among this, pulse oximetry has a special place. Intermittent hypoxaemia during the obstructive events of apnoea is directly and objectively recorded by oxygen saturation recordings, which show a pattern of blood desaturations during apnoea, recovery when the person awakens, and a subsequent drop on the next apnoea event, so that a timeline of respiratory stress is seen throughout the sleep period[17]. Such a continuous measurable physiological record removes the basis for diagnosis from the patient's subjective experience to the objective experience of his body over the course of the night. For this reason, most of the automated screening tools under development either use oximetry as their main or auxiliary input[17].

With the improvement of analytical techniques, beyond just detecting desaturation events, more and more features are being extracted from the SpO_2 signal which relate to the severity of the disease. An apnoea present or not classification offers a very coarse-grained understanding of physiological burden, whereas the depth of each drop of oxygen, the length of time in which saturation is suppressed and the frequency of its occurrence across a recording together provide a much finer-grained understanding of physiological burden[17]. Feature extraction from these dimensions can be automated to enable models to stratify patients along a continuum of severity, from those who have real maternal and fetal risks to those who have mild outcomes, and monitoring is the right approach. This is why this level of resolution is hard to obtain using only visual inspection of raw oximetry traces by the clinician, and is one of the reasons for the continued research interest in this area with computational methods[17].

Signal-based approaches are, in general, more defensible in terms of diagnosis than symptom questionnaires and ECG-derived approaches are a good example of this[18]. The apnoea episode is associated with an autonomic response, which can be measured, consisting of withdrawal of parasympathetic tone during apnoea and activation of sympathetic tone with the rapid heart rate recovery on arousal. These variations are captured in the R-R interval sequences and HRV metrics from the ECG signal and have been shown to be capable of decoding this cardiac signature with reasonable accuracy even in the absence of direct respiratory measurements or in cases of poor quality [18]. Machine Learning models have been shown to be able to decode this cardiac signature with reasonable accuracy even in the absence of direct respiratory measurements or in poor quality measurements. Pulse oximetry can take the analytical picture further yet, as it is possible to record both the oxygen saturation dynamics, and the morphology of the pulse wave, in a single low burden recording system [17].

The modality of contactless sensing has come into real focus for pregnant populations, mainly because of the absence of a compliance burden on patients due to the wearing of electrodes or wearables, and because of the use of adhesive sensors during late gestation when physical discomfort is already significant. Two types of systems to estimate thoracoabdominal respiratory movement exist, one based on radar and another on ballistocardiography, which is able to detect the minute mechanical vibrations generated by cardiac and respiratory functions when the human body is lying on a sleeping surface [19]. Conceptually both of them are suitable for home monitoring. However, clinically, the sensitivity of motion artifact is still a relevant technical problem, and the two modalities are not validated in pregnancy specific large cohorts so far to make confident clinical recommendations [19]. This is true, but there is not yet evidence to support this.

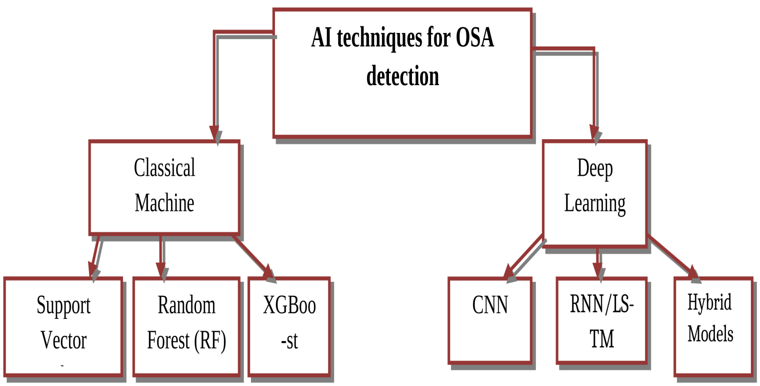

Figure 3. Taxonomy of Artificial Intelligence Architectures for OSA detection in pregnancy

Figure 3 demonstrates the proposed diagnostic framework that uses a diverse array of Artificial Intelligence (AI) technique helpful to identify patterns within physiological data that is associated with OSA. These methodologies are broadly categorized into **Classical Machine Learning**, which excel at handling prepared tabular data, and **Deep Learning**, which is specialized for extracting features from complex, high-dimensional waveforms.

2.2 Categorization According to Algorithmic Framework

Initial ML-based methods which are used for OSA detection were mostly based on traditional machine learning algorithms. An apneic/non-apneic classification was performed based on manually designed physiological features, using support vector machines [23]. Models that are based on the use of decision tree provides interpretability but can have low generalization capability in the case of the use of heterogeneous data[23]. Another class of classifiers which are robust are random forest classifiers, but that are still heavily depend on manual feature engineering[24]. The Limitations of manual feature engineering became more highlighted as high resolution physiological datasets becomes larger in number that creates the demand for such architecture that could analyse representations directly from raw signal data. As high resolution physiological data becomes increasingly available, deep learning architectures can be used. Convolutional neural networks (CNNs) also make it possible to automatically extract features of raw physiological signals without any need to design features by hand, which would be biased based on domain[20]. Those models based on CNN are proven to be effective for pattern recognition in oximetry, ECG signals and photoplethysmography (PPG) signals for apnea detection[20]. However, when this occurs, the ability of the architecture to represent properties of the signal in space and time is also crucial[20].

CNNs, however, are used to model signals as spatial structures, and are less effective at modelling the temporal dependencies between the evolution of breathing and cardiac activity over a whole night's sleep, which are directly addressed by the recurrent architectures. Physiological time-series with sequential dependencies have been modeled by recurrent neural networks (RNNs), and long short term memory (LSTM) networks[21]. Within the respiratory cycle, these models can replicate the temporal relationship, between the phases of respiration, and the phases of sleep[21]. A hybrid CNN-RNN networks, that might model the spatial and temporal information, might be able to get more accurate and robust classification result[21].

Recently, different models based on attention and transformer architecture have been introduced to be applied in OSA detection in multimodal. Models like this can be used to model the long-range dependency among heterogeneous signal inputs[24]. Ensemble and hybrid methods for generalization, merging the complimentary algorithmic advantages are suggested as another[24]. With the advancement of deep learning, there is a challenge on interpretability and clinical trust. The solutions to these problems have been proposed by explainable artificial intelligence. Using the feature attribution methods[22] the contribution of the individual physiological features to model predictions can be understood. Transparency and accountability is crucial in clinical decision making, particularly in obstetrics, which is why Explainable AI systems are important[22].

Figure 4 is a visualization of a deep learning model of the workflow to identify obstructive sleep apnea with the use of physiological data. Wavelet-based time-frequency representations of raw maternal physiological measurements, including ECG and SpO2, are then learnt through neural networks to acquire both discriminative feature maps to identify automated OSA and determine its severity.

2.3 Classification According to Clinical Use

The algorithmic frameworks described above are not applied uniformly across all clinical settings ,their appropriate use depends on the specific diagnostic purpose they are designed to serve.OSA diagnostic models for pregnancy created by artificial intelligence or machine learning differ in term of their uses clinically. Screening models are used to identify patients at a high risk of OSA using simple input data that can be collected during regular prenatal care[14]. Thus, they are more accessible and have increased sensitivity compared to them being used for diagnostic purposes[14]. Models utilized as screening tools are used to help identify patients who need more formal referral processes[15], not determine conclusive

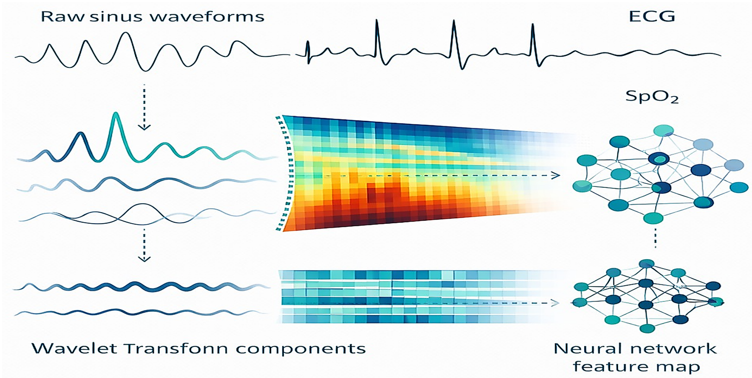

Figure 4. Deep Learning–Based Feature Extraction Pipeline for OSA Detection Using Physiological Signals

diagnoses.When the screening has identification of elevated risk, there is a need for more objective diagnostic assessment.

Diagnostic models for OSA focus on providing physiological measurements based on non-contact or wearable devices to predict the result of a polysomnography test or identify the severity level for patients with OSA[17]. Based on current literature, deep learning models that analyze data captured from sensors appear to outperform questionnaires in their ability to predict patients who have OSA[18]. It is important to use validated PSG data prior to any clinical implementation of AI/machine learning OSA models[13].A the diagnosis get confirmed , there is shift of clinical need from identification to ongoing surveillanceMonitoring systems concentrate on ongoing evaluation of symptoms from sleep apnea and treatment response, including continuous monitoring of oxygen saturation, heart rate variability, or periodicity of waking from sleep through use of body-worn sensors over an extended period of time[17]. Non-contact systems enable patients with limited physical or financial resources to undergo long-term home-basedmonitored care in a manner that creates minimal distraction/conceptual burden on the patient[19]. Therefore, by continuing to observe patients over time, one can determine whether the disease is progressing and is there an effective treatment for the pregnant woman[19].

Figure 5 illustrates the sequential three-stage clinical workflow through which AI and machine learning tools support OSA management in pregnancy, progressing from low-complexity questionnaire-based screening and risk stratification, through wearable and contactless diagnostic sensing with deep learning severity classification, to continuous longitudinal monitoring with explainable AI feedback, ultimately targeting improved maternal and fetal outcomes.

2.4 Overview of the Methodological Trends

Across the three clinical functions that are described in Section 2.3, in a published literature, there is a consistent shift, one that reflects both the maturation of sensing technology and the growing recognition that single-modality approaches are insufficient for the complexity of obstetric OSA.The literature shows that OSA is now being identified with objective and multimodal and automated detection systems. In pregnancy, deep learning networks using only the signal inputs have been shown to be much better at predicting OSA than non-objective screening methods[18]. Moreover, providing several physiological

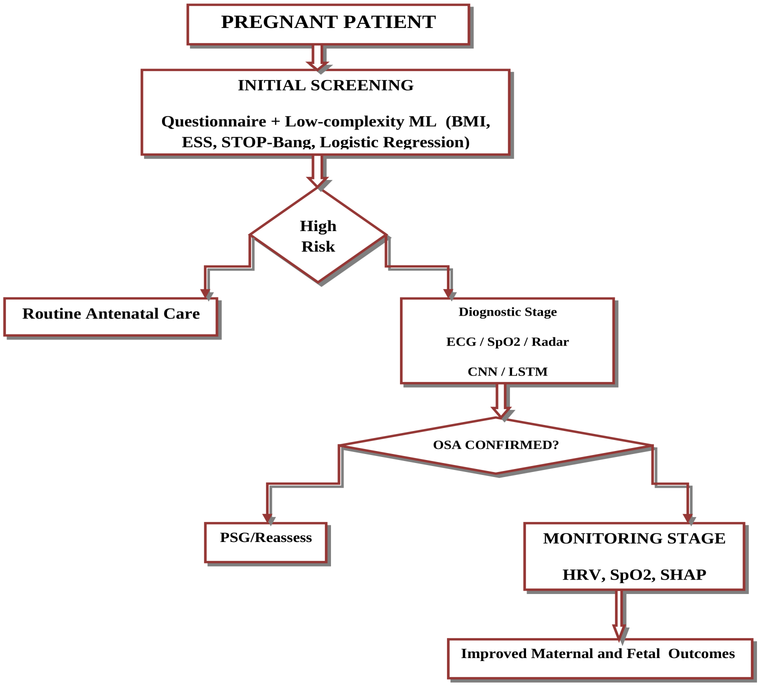

Figure 5. AI/ML-Based Clinical Decision Pathway for OSA Detection and Management in Pregnancy

signals can yield more reliable diagnosis results, primarily because the data from measuring cardiac and respiratory functions are complementary. There is a massive amount of variability in studies, however, varying by the size of the data sets, characteristics of the study population, and the methods used to validate the data. Since many of the available data models were created using a non-pregnant female population, there is limited utility for obstetrical populations[23].

The small sample size and retrospective design reduces external validity. Furthermore, the absence of standardised datasets and reporting procedures greatly hinder the ability to replicate and directly compare results across studies[24]. The field should therefore start moving into the direction of prospective validation, and development of models applicable only to an obstetrical population. Future research is likely to focus on creating clinically-interpretable frameworks.

Literature Survey

The methodological trends that was identified in Section 2.4 can be understandable in the context of how the field has developed over time. The literature on OSA in pregnancy has progressed through three broad phases: an early period focused on clinical epidemiology and questionnaire-based screening, a transitional phase introducing physiological signal analysis and classical ML, and a contemporary phase which is defined by deep learning, multimodal fusion, and explainability. The following survey has tracked this progression, with particular attention to the studies that directly inform the proposal of taxonomy in this review.TheInitial studies of OSA focused on its epidemiology, pathophysiology and clinical manifestations in the general adult population[1]. This type of work helped to prove that sleep fragmentation is associated with cardiometabolic or neurocognitive morbidity, and thus promoted further efforts to develop more specific diagnostic methods[2]. Later, sleep medicine guidelines led to standardized clinical definitions and criteria, giving a consistent framework for further studies[3].As field advanced, there was an appreciation that physiological changes associated with pregnancy could be a trigger for sleep-disordered breathing. Anatomical and hormonal changes in the mother are shown to make the upper airway susceptible during sleep[4]. Gestational weight gain, for example, seems to be responsible for altering lung mechanics and a corresponding decrease in lung volumes[5]. Other studies indicated that there is a significant contribution of mucosal edema in the upper airway and ventilatory control mechanism changes during the sleep cycle[6].

Based on these physiological findings, epidemiological studies were initiated to determine the rates of SBD throughout gestation. The data that are longitudinal indicate that the incidence of such events increases with the progression of pregnancy[7]. It is likely that OSA is more common in high-risk obstetric groups, particularly among women with obesity or chronic hypertension[8]. In addition, higher levels of disordered breathing have been correlated with metabolic complications during pregnancy like gestational diabetes[9]. Recent studies have focused on the risk of not treating maternal OSA. Pre-eclampsia and gestational HTN are documented to be associated with the disorder[10]. Other studies link maternal OSA with glucose intolerance and insulin resistance[11]. Additional analysis of the potential for preterm birth and low birth weight has been conducted, but the extent to which maternal sleep-disordered breathing may affect this is still being studied[12].While all these associations are apparent, it is difficult to detect the disorder in practice during pregnancy. Polysomnography is sometimes touted as the diagnostic gold standard; however, this is not practical for pregnant patients due to its resource intensive and cumbersome nature[13]. For these reasons, screening questionnaires have traditionally been used for prenatal care. As a result, STOP-Bang questionnaire assessments show a significant loss of diagnostic ability in obstetric groups[14]. This is also a limitation for the Berlin Questionnaire in this population[15]. Even the Epworth Sleepiness Scale has had limited performance in discriminating the risk of OSA during pregnancy[16].

In recent years, research has focused on objective analysis of physiological signals as an alternative to overcoming these problems. Pulse oximetry was one of the first non-invasive techniques used to detect the intermittent hypoxemia that is typical of apneic events[17]. Autonomic changes (as measured by heart rate variability) also have been shown to be an index of disordered breathing in electrocardiography studies[18]. New technology has also enabled the development of contactless respiratory monitoring systems that could become an alternative solution for monitoring at home[19].Machine learning is also key to OSA detection, alongside these sensing modalities. The ability to extract features automatically from high dimensional physiological signals can be accomplished using deep learning architectures, specifically convolutional neural networks[20]. There has also been a number of studies exploring the temporal dependence in respiratory and cardiac time-series data using recurrent neural networks. And in the clinical environment, where transparency of the model is a concern, XAI is starting to be used to make models more interpretable, and to build trust in models with the clinicians[22].

Various initial approaches used tried and tested machine learning techniques such as support vector machines and decision tree classifiers to detect OSA, based on handcrafted features[23]. The results of these comparisons showed that the performance of the models is inconsistent, which can be attributed to the differences in features used, features of the data sets, and procedures for data validation[24]. This outcome is evident and is proof of the lack of methodological standardisation in the field. However, more recently, wearable and multimodal sensing platforms for the longitudinal monitoring of sleep have been focused on.Studies using photoplethysmography (PPG) for OSA monitoring with wearable devices have shown that continuous monitoring of OSA is possible in a non-laboratory environment[25]. Likewise, multimodal approaches using both respiratory and cardiovascular signals are more reliable than those using only one of the signals[26]. Insights gained from investigations of radar and motion-based sensors also indicate that non-contact monitoring technologies can be considered a possible developmental trajectory[27].

Hybrid and ensemble deep learning models have also been developed to advance deep learning architectures. A hybrid approach to CNN and RNN was shown to be more suitable to capture spatial and temporal features[28]. Experiments related to attention-based mechanisms have been used to investigate the prioritization of a specific feature and making it more relevant in the signal-detection model[29]. Ensemble techniques are often employed to enhance generalizability because they often help to overcome inconsistent results in different datasets[30], but these models often don't incorporate any adaptations for pregnancy. There have been only a few studies directly applying machine learning techniques to obstetric cohorts, which are often limited by the number of patients available and inconsistencies in their data[31]. Due to physiological differences between pregnant women and non-pregnant women, researchers have contended that it is essential to validate the test specifically for pregnancy[32]. The dilemma is that perhaps the biggest challenge to the creation of these models is that one continues to have a limited amount of annotated data of sleep during pregnancy.

The focus of recent literature is now clinical translation and the associated ethical issues. For high risk settings such as obstetric care model interpretability and transparency is not a choice, but a necessity[34]. Issues of algorithmic bias and fairness have also emerged, especially with regards to the use of these tools in maternal health[35]. At the same time, studies on real-time and edge based systems indicate that there is a sense of interest in realizing these types of systems in low resource areas[36].In recent years, there has been a surge in the development and use of machine learning for OSA detection, and the field has been mapped with systematic and narrative reviews, highlighting an important gap in the field of pregnancy-specific studies[37]. These abstracts always target a broad spectrum of study designs, data modalities and assessment measures[38]. The existing literature illustrates a great need for a formal taxonomy that could integrate machine learning schemes based on the modality of the data, algorithmic architecture and clinical usefulness during pregnancy.

Figure 6 illustrates the historical progression of ML research in maternal OSA divided into three eras: Early Research 2014-2018) focusing on questionnaire data and low-cost screening, Transition & Deep Learning (2019-2025) focusing on physiological signals and deep architectures, and Future Directions (2025+) emphasizing multimodal fusion, standardized datasets, and longitudinal monitoring with the ultimate goal of clinician-directed, real-world clinical translation.

Discussion

The aim of this review was to gather 38 studies evaluating the feasibility of machine learning and artificial intelligence techniques to improve the diagnosis and/or treatment of OSA in pregnancy. The literature used is very diverse in terms of methodology, type of signal and clinical context, which made

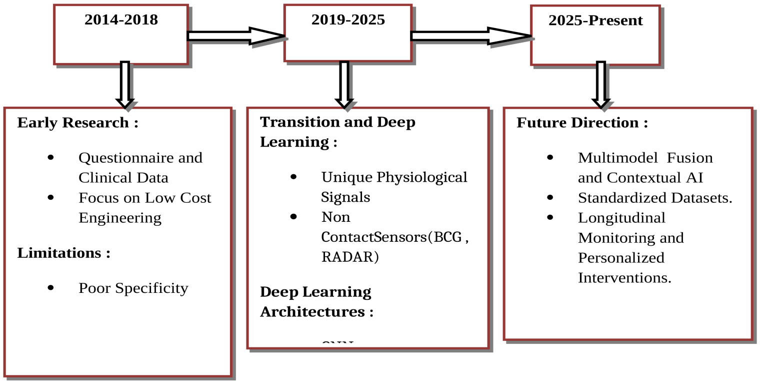

Figure 6. Evolution of ML in Maternal OSA

it necessary to develop an attribute-based taxonomy for this literature. The architecture presented in this document differs from traditional ML architectures and deep learning architectures and more generic AI architectures (AI frameworks) with details of individual studies provided in an annexure at the end of this paper. To quantitatively compare the model performance, 22 studies that reported complete metric data were analyzed and shown in Table 1 and the visual data were shown in Figures 7–10.

The accuracy is obtained by two different architectures, transformer and multimodal; the questionnaire-based and some radar-derived models are at the lower level as depicted in Figure 7. This disparity in performance seems to be effectively a ceiling imposed by subjectivity of self-reported data. In Figure 8, the sensitivity is moving towards AI and deep learning models in wearables, but at the cost of some loss of specificity. It is just that point that Figure 9 will illustrate: the specificity is higher in the case of the models with the ECG and oximetry, while the contactless and wearable models can be prone to variability, mainly because of movement artifacts or the variability of the home recording environment. Lastly, they are displayed in Figure 10, which is a normalization with the F1-score. Hybrid CNN-Transformer and multimodal models seem to yield the best compromise but, if transparency is desired, as is the case in clinical scenarios where a practitioner should have an explanation for the model's decision, explainable approaches such as XGBoost and SHAP are also competitive enough to warrant their continued use.

Table 2 provided in annexurethat reports 18 studies, uses traditional machine learning in OSA detection in clinical cohorts, and often in obstetric cohorts. All these pieces of evidence suggest that models based on pipelines, which rely on well-designed pipelines and interpretable classifiers, are competitive with more complex deep architectures, and often offer tangible advantages that deep architectures are not easily able to replicate. Of those, signal-based algorithms like the ones proposed by Markov et al.[3] and Sharma et al.[4] demonstrate good performance, provided that the PSG, ECG and SpO_2 recordings are free from artifacts and that feature selection is based on the physiological aspects of respiration. Sharma et al.[4] in particular define the obstetric environment as a special challenge as the physiological changes that occur during pregnancy (haemodynamics and breathing) tend to violate the assumptions underlying models developed with data from the general population. Calibrating the tests to local populations is a common theme across the studies listed in Table 2.Logistic regression models based on a questionnaire are available as an easy starting point for risk stratification in situations where resources are limited. Their sensitivity is enough for basic screening, but the result of the meta analysis performed by Maniaci et al.[18] shows that the AQ-AUC is always less than the PS-AUC. This is why there is a shift towards wearable or contactless monitoring. Between these extremes is the work of Espinosa et al.[14] and Wang et al.[15] which achieved good detection rates in community and obstetric samples respectively, but also reported inter-device variability as a challenge to establishing a standard.

The last two lines of Table. 2 are of particular interest from a methodological point of view. Khan et al.[17] showed that in many modern wearables, a standard part of the diagnostic and monitoring suite, the cardiac autonomic information is enough to be detected to achieve reasonable F1-scores for apnea detection. This discovery clearly has potential for use at the population level. On a different note, Zhang et al.[21] used XGBoost and Shapley additive explanations to build a model that assigns clinical-based, not mathematical-based, ranks to features, which ultimately led to finding anthropometric variables and depth of desaturations as the most important features. It is the transparency that clinical settings require and that distinguishes this work from the many that are currently in the lead, whose accuracy is much more important.Although these improvements have been made, the studies in Table 2 have some common limitations: they are based on convenience samples collected in the past, the metrics are not reported consistently and there is little evidence to support the obstetric validation of the studies. Improvements will probably be made here based on the designs being developed, and the use of reporting structures such as TRIPOD and STARD.

Table 3 in the annexure, which includes eight studies, show a useful improvement in technical development as well as a number of challenges to deployment of end-to-end deep learning architectures for OSA detection. These models have a significant benefit over the feature-engineering models listed in Table 2, as they are able to learn discriminative patterns from raw physiological signals, without being derived by the researcher, which makes them easier to adopt. In actuality, this change can result in quantifiable improvements of accuracy in most of the analyzed signal modalities.CNN-based techniques like the PPG-based apnea detection by Jiang et al.[5] and the multi-signal ECG/EEG analysis by Almazaydeh et al.[34] demonstrate that convolutional architectures are able to extract patterns that are relevant to the apnea from the waveform data, which is often not captured in manual methods. The PPG study has significance in the context of scaling as this is the type of sensor that is common in consumer wearables. But there are some problems with the accuracy data that are not reflected in the signals themselves and are only partially addressed by the accuracy data reported: The gap between laboratory and wearable data.

The Transformer-based model of Zhang et al.[7] and Hu et al.[12] take one step further in that the models are able to learn across modalities between EEG and ECG, SpO_2 and radar signals. The improvements in accuracy for the multiscale attention architecture over unimodal baselines in the study of Zhang et al.[7] are likely to be due to the temporal relationships between the signals across scales captured by the architecture. These are the relationships that inflexible feature chaining will just mask. The radar-based Transformer of Hu et al.[12] is definitely ambitious but the Moderate feasibility level reported in Table 3 is a recognition of the challenges of deployment and validation in real environments.The overconfidence and calibration error problems that are faced by single deep models when they get out of the data distribution used to train the model is also tackled by the ensemble approach proposed by Mukherjee et al.[9]. They combine the predictions from multiple CNN variants, and provide a meaningful but implicit uncertainty quantification of the predictions, albeit with additional computational burden.

Biswal et al.[35] follow a complementary approach and use unlabelled archives of PSG to learn in a self-supervised manner to reduce the lack of expert-annotated obstetric data. This approach is likely to be crucial in rare-population screening for which annotation costs by experts is impractical.A shortcoming for all of the tables in Table 3 is that they are not easily interpretable. Other researchers, including Phan et al.[37] report high accuracy but the inner workings of these models are not well understood. The clinical regulators have good reason to be concerned where a diagnostic system is not able to explain its results. In the absence of routinely coupled, calibrated explanation layers and prospectively validated in obstetric population, the gap between best results and clinical readiness will continue to be larger than the accuracy numbers would indicate.

In Table 4 in the annexure, the contributions are even more diverse than in the other two tables(Table 2 and Table 3), as they examine artificial intelligence either as a systems-level or evaluative framework, rather than specific aspects of the architecture. These entries span from systematic reviews and scoping to clinical trial protocols and explainable AI frameworks, and each have their own place in the chain of evidence that is needed to get algorithmic research to a responsible place in clinical use.The clinical trial protocol from the authors of Nugent et al.[6] is probably the most soundly based in this review. It demonstrates that a two-step approach with machine learning (screening questionnaires, followed by attended HSAT) is a viable option for recruiting and ensuring adherence to a CPAP intervention among pregnant patients. The lesson is that an organisation's structure and design of a clinical pathway is as important as what computational tools are used. So, in order to be accepted, an AI system will have to be part of the clinical workflow, even if the accuracy is high.

Critical distances, particularly from the point of view of a group of systematic and scoping reviews[28,29,31,32,36] may be lacking in individual performance studies. These works all argue that although the modalities are maturing and overall there is a growing performance, methodological differences in the way the datasets are constructed and in the definition of AHI makes it hard to draw reliable and aggregated conclusions. There is also a very limited number of models in the obstetric literature, and for most of these the data used is general adult data, and then applied to pregnant women without the proper adjustment.The meta-analysis of Liang et al.[30] and Tran et al.[23] are a quantitative analysis of multimodal and radar-based detection, respectively. The pooled sensitivities for some model classes are $> 85\%$[30] but there is considerable uncertainty, with large ranges in the sensitivities for these groups, reflecting high heterogeneity. The numbers are very large, and represent substantial clinically significant differences in the underlying studies. In addition, there are a limited number of primary studies that are the basis for the meta-analysis by radar[23], which naturally reduces the influence of the overall estimate for the primary studies in this list.

Explainability features as an integral part of the design in Hu et al.[27] and Rana et al.[38]. Hu et al.[27] instead make interpretable versions of the system an integral part of the system by designing an interactive visualization system for specific scenarios, while Rana et al.[38] integrate SHAP and LIME attribution methods into a decision support tool. This sort of solutions dovetails into the trend of regulators such as the FDA and EMA. Therefore, they are methodological indicators for future work as they are included in Table 4.Finally, Louis et al.[33] offer a clinical narrative review that completely sidesteps the use of AI and instead concentrates on the risks to the mother and the fetus, and the reason to try to detect them in the first place: preeclampsia, gestational diabetes and preterm birth. Together with the recommendations for out-of-laboratory testing by Collop et al.[16] these contributions remind us that accuracy measures are a tool to be used for a purpose. The goal at that end is to enhance outcomes of pregnant women and their babies.The field is shifting from manually designed and one-modality models to multimodal and explainable models which have been validated in obstetric populations. The aim is to be a part of clinical workflow and not an obstruction to it. There is still much to be done, but the studies included in the annexure provide a broad overview of the current situation, and the gaps for which they are still lacking.

Summary

The analyzed articles demonstrate a gradual shift in the usage of machine learning (ML) and deep learning (DL) models to detect and treat obstructive sleep apnea (OSA), especially in the condition of pregnancy. In the previous decade, scholars started to move on to more objective, data-driven and sensor-oriented models, having abandoned subjective and questionnaire-based methods. The growing popularity of artificial intelligence (AI) resulted in the variety of methods, which include screening using questionnaires, analyzing physiological signals, using radar, and integrating multimodal data. This shift can be seen as a combination of the development of computational capabilities and a growing awareness of the necessity to develop automated, efficient, and scalable screening modalities that are capable of meeting the needs of pregnant population.The first models were mostly based on the use of demographic and self-reported data to determine risk factors that were linked with OSA. Other studies showed that a combination of anthropometric characteristics (such as body mass index (BMI), age, and neck circumference) and self-reported sleep symptoms might have moderate predictive value. Whilst these models offered a useful basis on how to identify risk at early stages, the models were confined by their subjective nature and could not be generalized in various clinical conditions.

Further reviews noted that model heterogeneity and lack of external validation were the factors that led to inconsistency in predictive accuracy. Signal-based methods became prevalent models of research as it grew. The use of physiological parameters including oxygen saturation (SpO_2), electrocardiography (ECG) and photoplethysmography (PPG) helped in making the process more accurate and continuous in monitoring apneic events. There were high correlations in multiscale feature extraction oximetry-based studies with apnea-hypopnea index (AHI) scores obtained using polysomnography (PSG), underpinning the feasibility of wearable devices to perform large-scale monitoring. In the same way, ECG and PPG data have been used together with ML classifiers and deep neural networks to identify apneic instances and differentiate OSA severity with more accuracy.Along with the contact-based approaches to physiological monitoring, recent studies have presented the non-contact approach of sensing with radar and ballistocardiography to detect OSA unobtrusively. Research has demonstrated that multimodal systems using radar and transformer-enhanced methods can detect the respiratory motion and identifyapnea-hypopnea events without physical contact. The design of commercial ballistocardiography sensors has also been assessed as feasible to be used at home during pregnancy in spite of motion noise and environmental sensitivity.

As larger datasets started to become more widespread and the computational efficiency of models increased, deep learning models started to take over. Convolutional neural networks (CNNs), long short-term memory (LSTM) networks, and hybrid CNN-transformer models have demonstrated a high level of classification accuracy and strength over traditional algorithms. Ensemble and multimodal fusion models were also employed to improve the performance of diagnostics; multiple sources of signals and feature extractors were integrated. These signal processing methods enhanced noise tolerance, temporal stability and diagnostic accuracy, allowing close real-time OSA determination in the clinical and home environments.The use of explainable artificial intelligence (XAI) mechanisms to enhance interpretability and clinician trust is an emerging trend in the literature. Research that uses SHapley Additive exPlanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME) has helped to make AI-powered decisions more transparent and enabling medical professionals to interpret the contribution of physiological factors like oxygen desaturation, snoring severity, and BMI to the final decision.

The addition of interpretability measures indicates a shift in the paradigm of research from performance-based to explainable diagnostic models that are clinically relevant. Although these improvements have taken place, similar methodological issues are noticeable across the studied papers. The lack of reproducibility and generalizability has been limited by data heterogeneity, small samples, limited pregnancy-specific cohorts, and inconsistencies in labelling of events. Many models are tested only under controlled laboratory conditions, and few have been validated in real-world clinical settings. One of the limitations on scalability is technical, including computational latency, insufficient energy efficiency of wearable devices, and lack of longitudinal validation. The additional clinical research offers the necessary information on the practical implications of these AI-based procedures.

Table 1 just below presents a comparative evaluation of machine learning and deep learning models for obstructive sleep apnea detection, reporting accuracy, sensitivity, specificity, and F1-score across diverse architectures including traditional ML, deep neural networks, transformer-based models, radar-based systems, and wearable AI frameworks. Overall, deep learning and hybrid architectures which are stated as multiscale transformers, CNN–transformer combination, and multimodal fusion approaches reach the best performance, and they are basically associated with high accuracy levels (0.92 or higher) and high sensitivity for apneic event detection. The signal based frameworks that include ECG data, PPG data, oximetry data, and EEG data are more reliable than the traditional signal feature engineering based frameworks that only use questionnaires. In the meantime, explainable machine learning methods, like XGBoost coupled with SHAP, present a middle ground that allows for a decent performance with considerably wanted interpretability. However, wearable and radar-based systems have a high sensitivity but tend to be less specific, possibly because of the presence of motion artifacts and the variability of the monitoring setting.

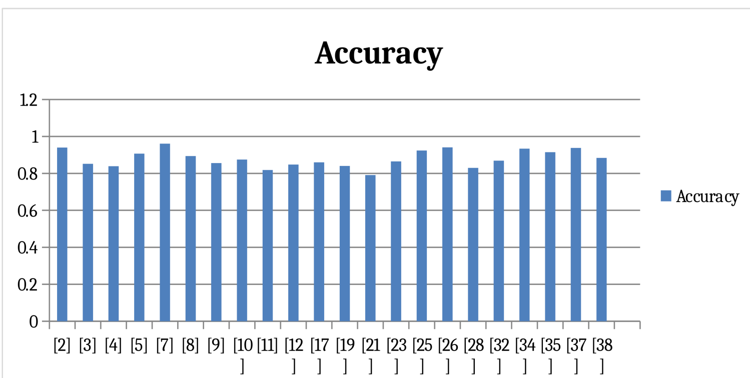

Figure 7. Accuracy Comparison Across Machine Learning Models

Table 1. Comparative Performance of Machine Learning and Deep Learning Models for Obstructive Sleep Apnea Detection

S. No.	Core ML Architecture	Accuracy	Sensitivity	Specificity	F1-Score
1	Dual-branch CNN (ECG) [2]	0.94	0.93	0.95	0.93
2	Interpretable Feature-ML [3]	0.852	0.841	0.86	0.845
3	RUSBoosted Tree (Wavelet) [4]	0.839	0.825	0.844	0.91
4	1D-CNN (PPG-based) [5]	0.907	0.889	0.924	0.905
5	Multiscale Transformer [7]	0.961	0.977	0.867	0.977
6	Multiscale Feature-ML [8]	0.894	0.865	0.882	0.871
7	MLP-based Ensemble DL [9]	0.856	0.842	0.865	0.851
8	Binary BCG-based ML [10]	0.875	0.72	0.7	0.82
9	Pregnancy BATE Algorithm [11]	0.818	0.769	0.815	0.805
10	ConvTransLSTM (Radar) [12]	0.848	0.748	0.885	0.852
11	Multi-scale Residual Net [17]	0.86	0.841	0.871	0.891
‘12	Maternal SDB Predictive ML [19]	0.84	0.824	0.835	0.829
13	XGBoost + SHAP analysis [21]	0.791	0.824	0.548	0.77
14	Radar ML (Meta-Analysis) [23]	0.865	0.944	0.699	0.852
15	Hybrid CNN-Transformer [25]	0.924	0.921	0.92	0.922
16	Oximetry ML Approach [26]	0.941	0.932	0.945	0.938
17	Wearable AI Framework [28]	0.83	0.99	0.79	0.81
18	Wearable AI (Pooled Mean) [32]	0.869	0.938	0.752	0.878
19	AI-based ECG/EEG signals [34]	0.934	0.915	0.94	0.926
20	Deep Learning [35]	0.915	0.898	0.92	0.907
21	Multimodal Deep Fusion [37]	0.938	0.949	0.927	0.95
22	Explainable ML Decision [38]	0.884	0.862	0.89	0.874

Figure 7 shows the accuracy of OSA detection prediction. Signal-based and deep learning methods tend to outperform models based on questionnaires, with the best performing showing a value of 0.92 or greater, and multimodal (fusion) and transformer-based methods are best, with some showing a value as high as 0.92. This is because they can learn correlations between signals which are not available to individual-modality classifiers. Ensemble and residual networks being just medium level models are able to get accuracy scores that would be accepted in a clinical setting if applied to obstetric data sets.

Model sensitivity (in the clinical sense of OSA detection) is similar as shown in Figure 8. Wearable AI systems and radar-based meta-analytic models are the most sensitive of all, occasionally with a sensitivity higher than 0.94, which means that they are able to detect a large proportion of apneic events. Multimodal fusion and transformer architectures perform quite well, and they seem to reach a balance between sensitivity and general reliability. On the other hand, questionnaire-based and single-feature models are much less sensitive, which shows that they are not able to detect subtle physiological events.

Figure 9 illustrate specificities of the models of the architectures surveyed, and reliability of the model to avoid false classification of the positive ones. Best performing models are based on either an ECG or an oximetry measurement or a multimodal deep learning model and the remaining models, such as radar and wearable-based models, have relatively low specificity, likely due to their susceptibility to motion artifact and the variability of the environments where they are recorded at home. This can be of particular importance when working in a clinical setting as a low level of specificity can result in more referrals and ultimately loss of trust in the system by the clinician.

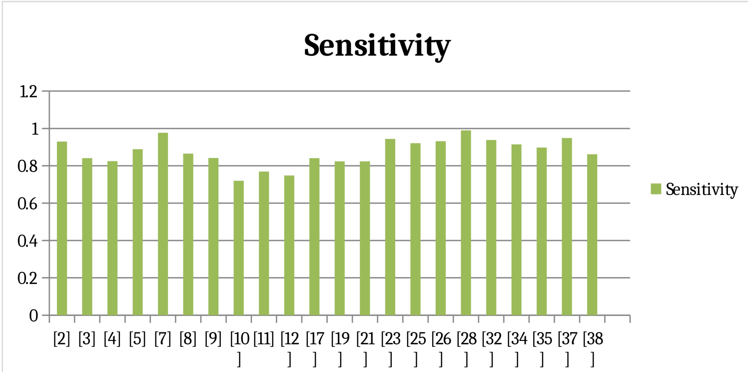

Figure 8. Sensitivity Comparison Across Machine Learning Models

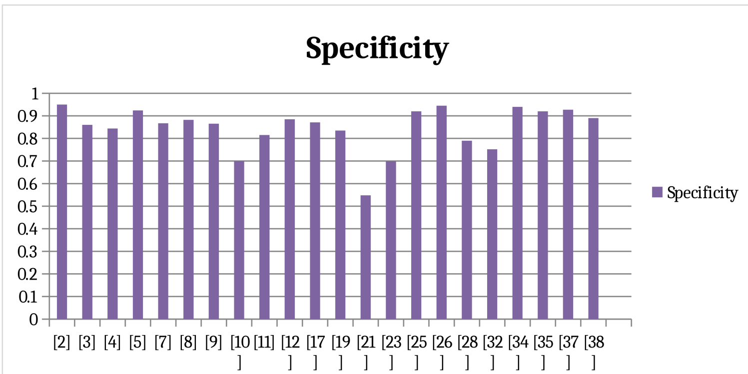

Figure 9. Specificity Comparison Across Machine Learning Models

The result from F1 score is summarised in Figure 10, which is a single value that reflects the reliability of the diagnosis (sensitivity and precision). Among all the models, two hybrid models that are CNN-Transformer models and multimodal deep fusion models show better performance with better F1-scores in detection under various types of patients, although their architecture is complicated. Explainable ML (XGBoost) is a bit less competitive, but still viable with its interpretability as a requirement in many cases where clinical transparency is required, in addition to performance.

Current comparative studies of continuous positive airway pressure and standard antenatal care show that algorithms offer a potentially significant contribution beyond the scope of detection, potentially in aiding treatment adherence as well as quantifying treatment response with time. Studies of the prevalence of OSA in high-risk pregnancies have been conducted separately and have made the practical imperatives more apparent — it is only useful to be able to diagnose OSA if it can be linked to care pathways which are readily available to pregnant women, meaning that it is important as much as it is important to improve the algorithms that lead to the diagnosis.However, in looking at the whole body of work, a movement

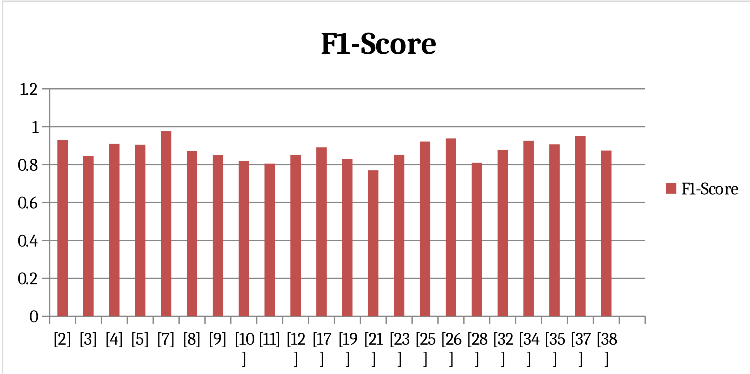

Figure 10. F1-Score Comparison Across Machine Learning Models

can be seen towards the individual studies. Even though many statistical methods remain in use, the field has progressed, and in many cases moved, away from classical statistical thinking towards architectures capable of dealing with complexity, and increasingly to designs where interpretability becomes a core consideration and not an optional. This change has clinical implications. If the clinician who receives the output from the diagnostic process doesn't understand what caused it, or trust what that process is telling them, then it's a bit less accurate than it could be.

The infrastructure issue is one which has not yet been resolved in the literature. While powerful models are valuable, they are not very useful unless the data they rely on is similarly collected, unless the validation process across different studies is not too varied to allow for meaningful comparisons, and if fairness and transparency are not part of the foundation of the model, but are added on after the fact. These are not incidental issues but environment in which any of this work can be utilized in a real maternal health care facility. The sophistication of the computation is in place, but the scaffolding for the clinical and methodological aspects is still to be caught up.

Conclusion

The diagnosis of obstructive sleep apnea (OSA) in pregnancy is a situation that lies in a not so comfortable middle ground: clinically relevant, physiologically multifaceted and often overlooked. This review was designed to help to organize the myriad of computational efforts that were beginning to emerge with the goal of changing this situation. In a series of 38 studies, grouped into a taxonomy of four dimensions of data modality, algorithmic class, evaluation framework, and clinical application context, a clear pattern emerges; the field has made real technical advances, but there is a wide gap between the performance of the algorithms in the studies and their clinical deployment, and that gap should not be deemphasized.

It's important to make the development of algorithms clear. Initial models used logistic regression and decision tree analysis based on questionnaire responses (BMI, neck circumference and self-reported snoring) which provided low computational cost and interpretability, but usually failed to reach $\geq$0.82-0.84 accuracy rate in obstetric populations due to the high overlap between symptoms measured with questionnaires and 'normal pregnancy'. When the trend was towards the use of signaling techniques, the situation was quite different. Studies that used features from the ECG, including the multi-feature single-lead model of Khan et al.[17] which used features based on heart rate variability (HRV), showed that features derived from the autonomic activity of the heart also produced useable F1-scores for the detection of apnea. As done in Sharma et al.[4] using SpO_2 and ECG signals, Wavelet-based SVM and Random Forest classifiers were able to successfully achieve accuracy within the 0.84–0.85 range which was still clinically understandable to scrutinize. The 0.907 accuracy and 0.905 F1-score demonstrated by Jiang et al.[5] with the 1D-CNN on PPG is a significant improvement and of particular importance with regard to scalability, as PPG is now standard in any consumer wearable. Mukherjee et al.[9] showed that an ensemble of various CNN models using MLP based technique improved the robustness to 0.856 accuracy with reduced overconfidence issues if the CNN models were out of distribution. Huysmans et al.[10] reported that binary classification (specified at 0.70) was possible by using non-contact sensing and the resulting ballistocardiography signals could be measured at home.

The highest performing architectures were those that combine information from multiple modalities, and explicitly model temporal dependencies. The Zhang et al.[7] multiscale Transformer, which was trained on EEG, ECG and SpO_2 data, reported the highest accuracy (0.961) and sensitivity (0.977) among the others mentioned in this review, being able to infer the temporal relationships across different signals that cannot be obtained by feature engineering. One of the more clinically credible architectures reviewed was the Hybrid CNN-Transformer of Wang et al.[25] which achieved an accuracy of 0.924 as well as an F1 score of 0.922 (balanced score) with a high specificity of 0.92. Phan et al.[37] obtained the best accuracy (0.938) and the highest F1 (0.95) in the comparison table in the model called Multimodal Deep Fusion. The FECT-OSA system of Hu et al.[12] demonstrated high accuracy, 0.848, for non-contact monitoring via a radar system with transformer enhancement, but noted that the feasibility of this approach in practice is moderate, due to the need to change the system from an experimental to real-world deployment. As demonstrated by Wang et al.[26] oximetry-based ML achieved 0.941 accuracy, and excellent specificity of 0.945, showing that with a carefully selected physiological signal, processed correctly, a single signal can still be very competitive. Against this backdrop, the Zhang et al.[21] XGBoost with SHAP analysis is a little bit different: Only 0.791 accuracy, but with a breakdown of the features (depth of oxygen desaturation, anthropometric measures, gestational characteristics) that were used to make each prediction that was clinically interpretable. However, in a clinical setting, it may be more important to have some degree of traceability than to gain an accuracy tweak, and that is why explainable, as well as performance-focused methods, are still of interest.

The numbers as a whole indicate that there is no single model type which is the best in every clinical context. High sensibility (0.99) wearable systems, like the one reported by Osa-Sanchez et al.[28] can achieve a sensitivity high enough that no apneic events will be missed, but at a specificity of 0.79, creating a clinically burdensome false-positive rate on a large scale. The BATE algorithm of Facco et al.[11] was developed to be pregnancy specific, but its performance shows a moderate accuracy of 0.818 and a moderate sensitivity of 0.769, suggesting that even a pregnancy-specific screening algorithm is not very accurate when applied across the diverse physiological space of pregnancy. The average accuracy (0.869) and average sensitivity (0.938) of the wearable AI meta-analysis conducted by Abu Bakar et al.[32] indicated that wearable systems in general have hit a point of clinical interest, but the range of the constituent studies points to the fact that uniformity, if not superior performance, is still a challenge.

Together, these result in three new conclusions, in comparison to the conclusions drawn by previous reviews. First, multimodal deep learning has reached a point of performance where it is a viable tool for diagnosis, not just a research tool, where the data quality and validation framework is in place to support it. Secondly, the concepts of interpretability and usefulness are not mutually exclusive: we have seen that it is possible to build an XAI-integrated system from the beginning without compromising the clinical usefulness of the system, such as Hu et al.[27] or Rana et al.[38]. Third, the current major bottleneck is not algorithmic, but it is the lack of standardised, pregnancy-specific datasets, consistent definitions of AHI and prospective validation studies of the algorithms that reflect the diversity of obstetric patients throughout gestation, comorbidities, and healthcare system settings. To achieve that will need the sort of collaborative funding that can only be achieved between research groups: investment in shared data infrastructure, agreed reporting standards, such as TRIPOD and STARD, and a fundamental change in the criteria by which research is judged, from accuracy of the benchmarks to usefulness in the real world of clinical practice. The aim was not to come up with the exact count. To present a clinician with a tool he could use reliably and interpret with confidence to identify a serious condition for mothers and babies, earlier, more consistently and in locations where polysomnography is not available. Now a basic computational tool has been realized for this. Depending on the clinical and methodological scaffolding, which can be created within the field, the next step will be.

Author contributions

Vandana Roy: Writing Original Draft, Conceptualization, Shalini Stalin: Data collection and Review, Nikhar Vishwakarma: Supervision Anuj Nayak: Revision

Data availability

Not Applicable

Declarations

Ethics approval and consent to participate: Not applicable. Consent for publication: Not applicable. Competing interests: The authors declare no competing interests.

Funding Statement

This study was funded by MPCST, Bhopal with the reference no: A/RD/RP-2/398.

References

[1] Janmohammadi P, Raeisi T, Zarei M ... Adipocytokines in obstructive sleep apnea: A systematic review and meta-analysis Respiratory Medicine, 2023; 208

[2] M. Kolhar et al., “AI-driven detection of obstructive sleep apnea using dual-branch CNN and machine learning models,” Biomedicines, vol. 13, no. 5, p. 1090, 2025. Available: https://www.mdpi.com/2227- 9059/13/5/1090.

[3] K. Markov, M. Elgendi, V. Birrer, and C. Menon, “Interpretable feature-based machine learning for automatic sleep detection using photoplethysmography,” npj Biosensing, vol. 2, no. 1, Jun. 2025, doi: 10.1038/s44328-025-00041-2.

[4] M. Sharma, S. Bapodara, J. Tiwari, and U. R. Acharya, “Automated sleep apnea detection in pregnant women using wavelet-based features,” Informatics in Medicine Unlocked, vol. 32, p. 101026, 2022, doi: 10.1016/j.imu.2022.101026.

[5] Jiang X, Ren Y, Wu H, Li Y and Liu F (2023) Convolutional neural network based on photoplethysmography signals for sleep apnea syndrome detection. Front. Neurosci. 17:1222715. doi: 10.3389/fnins.2023.1222715.

[6] R. Nugent et al., “Continuous positive airway pressure versus usual care for obstructive sleep apnoea in pregnancy: A two-step pilot trial,” Pilot and Feasibility Studies, vol. 11, p. 1645, 2025. Available: https://pilotfeasibilitystudies.biomedcentral.com/articles/10.1186/s40814-025-01645-1

[7] Y. Zhang et al., “Deep learning for obstructive sleep apnea detection and severity assessment: A multimodal signals fusion multiscale transformer model,” Nature and Science of Sleep, vol. 15, no. 3, p. 321, 2025. Available: https://www.dovepress.com/deep-learning-for-obstructive-sleep-apnea-detection-and- severity-asses-peer-reviewed-fulltext-article-NSS.

[8] N. H. Hoang and Z. Liang, “Detection and severity classification of sleep apnea using continuous wearable SpO_2 signals: A multiscale feature approach,” Sensors, vol. 25, no. 6, p. 1698, 2025. Available: https://www.mdpi.com/1424-8220/25/6/1698

[9] D. Mukherjee et al., “Ensemble of deep learning models for sleep apnea detection: An experimental study,” Sensors, vol. 21, no. 16, p. 5425, 2021. Available: https://www.mdpi.com/1424-8220/21/16/5425

[10] D. Huysmans et al., “Evaluation of a commercial ballistocardiography sensor for sleep apnea screening and sleep monitoring,” Sensors, vol. 19, no. 9, p. 2133, 2019. Available: https://www.mdpi.com/1424- 8220/19/9/2133.

[11] F. L. Facco et al., “A screening algorithm for obstructive sleep apnea in pregnancy,” Annals of the American Thoracic Society, vol. 16, no. 3, pp. 285–293, 2019. Available: https://www.atsjournals.org/doi/10.1513/AnnalsATS.201902-131OC

[12] D. Hu et al., “FECT-OSA: A transformer-enhanced multimodal system for non-contact sleep apnea monitoring,” Alexandria Engineering Journal, vol. 74, pp. 45–60, 2025. Available: https://www.sciencedirect.com/science/article/pii/S1110016825007045

[13] E. C. Johns et al., “High prevalence of obstructive sleep apnea in pregnant women with class III obesity: A prospective cohort study,” Journal of Clinical Sleep Medicine, vol. 19, no. 2, p. 9578, 2023. Available: https://jcsm.aasm.org/doi/10.5664/jcsm.9578

[14] M. A. Espinosa et al., “Home-based devices for detecting obstructive sleep apnea: A comprehensive study,” Sensors, vol. 23, no. 23, p. 9512, 2023. Available: https://www.mdpi.com/1424-8220/23/23/9512

[15] J. Wang et al., "Home monitoring for clinically suspected obstructive sleep apnea in pregnancy," Journal of Clinical Sleep Medicine, vol. 20, no. 2, p. 10726, 2024. Available: https://jcsm.aasm.org/doi/10.5664/jcsm.10726

[16] N. A. Collop, "Home sleep apnea testing for obstructive sleep apnea," Sleep Medicine Clinics, vol. 18, no. 1, p. 10887466, 2023. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC10887466/

[17] A. Khan et al., "Multi-feature automatic extraction for detecting obstructive sleep apnea based on single- lead electrocardiography signals," Sensors, vol. 24, no. 4, p. 1159, 2025. Available: https://www.mdpi.com/1424-8220/24/4/1159

[18] A. Maniaci et al., "Screening tools for obstructive sleep apnea in pregnant women: An extended and updated systematic review and meta-analysis," International Journal of Obstetric Medicine, vol. 33, no. 6, pp. 429–442, 2024. Available: https://journals.lww.com/ijom/fulltext/2024/08060/screening_tools_for_obstructive_sleep_apnea_in.4.aspx.

[19] J. Wang et al., "Performance of machine learning models to detect maternal sleep disordered breathing," Nature Medicine: Digital Health, vol. 4, no. 2, pp. 45–58, 2024. Available: https://nature.com/articles/s44294-024-00030-2

[20] J. Wang et al., "Performance of machine learning-based models to screen obstructive sleep apnea in pregnancy," Sleep Medicine, vol. 108, no. 2, pp. 45–57, 2024. Available: https://www.researchgate.net/publication/383368161_Performance_of_machine_learning-based_models_to_screen_obstructive_sleep_apnea_in_pregnancy.

[21] Y. Zhang et al., "Prediction model of sleep disorders in pregnant women using machine learning and SHAP analysis," BMC Pregnancy and Childbirth, vol. 24, p. 8026, 2024. Available: https://www.researchgate.net/publication/396089083_Predictive_model_of_sleep_disorders_in_pregnant_women_using_machine_learning_and_SHAP_analysis.

[22] S. Siriyotha et al., "Prediction models of obstructive sleep apnea in pregnancy: A systematic review and meta-analysis of model performance," Diagnostics, vol. 11, no. 6, p. 1097, 2021. Available: https://www.mdpi.com/2075-4418/11/6/1097

[23] N. B. M. H. Tran et al., "Radar-based detection of obstructive sleep apnea: A systematic review and network meta-analysis of diagnostic accuracy across frequency bands," PLOS Digital Health, vol. 4, no. 2, p. e12385411, 2025. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC12385411/

[24] J. W. Kim et al., "Relationship between snoring intensity and severity of obstructive sleep apnea," Sleep and Breathing, vol. 19, no. 4, pp. 1347–1353, 2015. Available: https://www.researchgate.net/publication/284570410_Relationship_Between_Snoring_Intensity_and_Severity_of_Obstructive_Sleep_Apnea.

[25] X. Wang 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, vol. 47, no. 12, p. zsae184, 2024. Available: https://academic.oup.com/sleep/article/47/12/zsae184/7729858

[26] X. Wang et al., "A machine learning approach for the diagnosis of obstructive sleep apnea using oximetry, demographic and anthropometric data," Southern Medical Journal, vol. 118, no. 4, pp. 45–55, 2025. Available: https://journals.lww.com/smj/fulltext/2025/04000/a_machine_learning_approach_for_the_diagnosis_of.6.aspx.

[27] S. Hu et al., "Transparent artificial intelligence-enabled interpretable and interactive sleep apnea assessment across flexible monitoring scenarios," Nature Communications, vol. 12, no. 1, p. 62864, 2025. Available: https://nature.com/articles/s41467-025-62864-x

[28] A. Osa-Sanchez et al., "Wearable sensors and artificial intelligence for sleep apnea detection," Frontiers in Digital Health, vol. 4, no. 2, p. 12089203, 2025. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC12089203/

[29] C. V. Senaratna et al., "Artificial intelligence in the diagnosis of obstructive sleep apnea: A scoping review," Sleep Medicine Reviews, 2024. Available: https://link.springer.com/article/10.1007/s00405-025- 09377-x.

[30] Y. Liang et al., "Diagnostic accuracy of artificial intelligence for obstructive sleep apnea detection: A systematic review," Journal of Sleep Research, 2025. Available: https://link.springer.com/article/10.1186/s12911-025-03129-x

[31] V. Tantrakul et al., "Review of application of machine learning as a screening tool for diagnosis of obstructive sleep apnea," Sleep and Breathing, 2022. Available: https://www.mdpi.com/1648-9144/58/11/1574.

[32] N. A. Abu Bakar et al., "Detection of sleep apnea using wearable artificial intelligence: A systematic review and meta-analysis," npj Digital Medicine, 2024. Available: https://www.researchgate.net/publication/354853649_Detection_of_Sleep_Apnea_using_Machine_Learning_Algorithms_based_on_ECG_Signals_A_Comprehensive_Systematic_Review.

[33] J. M. Louis et al., "Obstructive sleep apnea in pregnancy: A comprehensive review of maternal and fetal implications," Journal of Clinical Medicine, 2024. Available: https://www.mdpi.com/2035-8377/16/3/39

[34] L. Almazaydeh et al., "Artificial intelligence-based approaches for sleep-related breathing event identification using ECG and EEG signals," Sleep and Breathing, 2025. Available: https://link.springer.com/article/10.1007/s11325-025-03442-9

[35] S. Biswal et al., "Deep learning for automated detection of sleep apnea using physiological signals," IEEE Reviews in Biomedical Engineering, 2022. Available: https://www.sciencedirect.com/science/article/abs/pii/S1746809423001222?via%3Dihub

[36] O. Faust et al., "A review of automated obstructive sleep apnea detection methods using machine learning," Biomedical Signal Processing and Control, 2021. Available: https://www.sciencedirect.com/science/article/abs/pii/S0010482522008083?via%3Dihub

[37] H. Phan et al., "Multimodal deep learning frameworks for sleep apnea detection," IEEE Journal of Biomedical and Health Informatics, 2023. Available: https://www.sciencedirect.com/science/article/abs/pii/S0169260719303086?via%3Dihub

[38] R. Rana et al., "Explainable machine learning models for clinical decision support in sleep apnea diagnosis," Artificial Intelligence in Medicine, 2024. Available: https://www.sciencedirect.com/science/article/pii/S1877050925014346?via%3Dihub

[39] S. R. BasiReddy, R. Asuri, R. C. Yarrapothu and M. Parasa, "Optimized Information Storage, Retrieval, and Decision Intelligence Through AI-Enabled Management Models," 2026 IEEE 18th International Conference on Computational Intelligence and Communication Networks (CICN), Manila, Philippines, 2026, pp. 1761-1766, doi: 10.1109/CICN70047.2026.11594394.

[40] M. Parasa, S. R. BasiReddy, R. Asuri and R. C. Yarrapothu, "Enhancing Data Security in Cloud Computing Using Intelligent Encryption and Access Control Mechanisms," 2026 IEEE 18th International Conference on Computational Intelligence and Communication Networks (CICN), Manila, Philippines, 2026, pp. 742-748, doi: 10.1109/CICN70047.2026.11594345. Annexure Summarization of the existing research work done in area of sleep detection based on Machine Learning, Artificial Intelligence and deep learning methods are done in table 2, 3 and 4 and discussed in the section 4.

Annexure

Summarization of the existing research work done in area of sleep detection based on Machine Learning, Artificial Intelligence and deep learning methods are done in table 2, 3 and 4 and discussed in the section 4.

Table 2. OSA detection based on Machine Learning Models

Attribute	**[1]**	**[2]**	**[3]**	**[4]**
Authors & Year	Janmohammadi & Coauthors, 2023 [1]	Kolhar & Coauthors, 2025 [2]	Markov & Coauthors, 2024 [3]	Sharma & Coauthors, 2022 [4]
Features / Techniques Used	Adipocytokine biomarkers, statistical pooling	Dual-branch CNN + handcrafted ML features	Interpretable handcrafted physiological features	Wavelet-domain features
Data Type / Signal	Clinical biochemical + PSG-confirmed OSA diagnosis	PSG (EEG, ECG, SpO_2)	PSG signals	ECG, SpO_2 (pregnant women)
Artifact Removal	Study quality assessment, bias filtering	Signal filtering, normalization, segmentation	Signal cleaning and epoch rejection	Wavelet denoising
Model / Algorithm	Meta-analysis (random/fixed effects)	CNN + RF / SVM	Feature-based ML models	SVM, Random Forest
Evaluation Metric (reported)	Effect size, odds ratio, heterogeneity (I^2)	Accuracy, AUC, F1- score	AUC, sensitivity, specificity	Accuracy, sensitivity
Dataset Size	Multiple pooled clinical studies	Public PSG datasets	Multi-cohort datasets	Pregnant women cohort
Sampling Rate (Hz)	-	PSG standard (inferred)	PSG standard	Clinical standard
Feasibility	High	High	High	High
Reliability / Explainability	High	Moderate (nomogram)	Very High	Moderate
Speed	-	Fast	Fast	Fast

Table 3. Detection Based on Machine Learning Models

Attribute	**[8]**	**[10]**	**[11]**	**[13]**
Authors & Year	Hoang & Liang, 2025 [8]	Huysmans & Coauthors, 2019 [10]	Facco & Coauthors, 2019 [11]	Johns & Coauthors, 2023 [13]
Features / Techniques Used	Multiscale SpO_2 statistical features	Ballistocardiography signal features	Clinical risk-factor screening	Epidemiological analysis
Data Type / Signal	Wearable SpO_2	Non-contact BCG	Questionnaire + demographic data	PSG-confirmed diagnosis
Artifact Removal	Motion artifact filtering	Motion compensation	-	Clinical quality control
Model / Algorithm	ML classifiers	Rule-based + ML	Logistic regression	Statistical analysis
Evaluation Metric (reported)	Accuracy, F1-score	Sensitivity, specificity	Sensitivity, specificity	Prevalence rates
Dataset Size	Wearable cohort	Small clinical cohort	Pregnant cohort	Prospective cohort
Sampling Rate (Hz)	Wearable standard	Sensor-dependent	-	PSG standard
Feasibility	Very high	Moderate	High	High
Reliability / Explainability	Moderate	Moderate	High	High
Speed	Fast	Fast	Fast	-

Table 4. Detection Based on Machine Learning Models

Attribute	**[14]**	**[15]**	**[17]**	**[18]**
Authors & Year	Espinosa & Coauthors, 2023 [14]	Wang & Coauthors, 2024 [15]	Khan & Coauthors, 2025 [17]	Maniaci & Coauthors, 2024 [18]
Features / Techniques Used	Home-device signal comparison	Home monitoring protocol	Multi-feature ECG extraction	Screening tools meta- analysis
Data Type / Signal	HSAT signals	HSAT + SpO2	Single-lead ECG	Questionnaire & clinical data
Artifact Removal	Device-level preprocessing	Standard HSAT preprocessing	ECG filtering	Study quality filtering
Model / Algorithm	Embedded algorithms	Clinical scoring	ML classifiers	Meta-analysis
Evaluation Metric (reported)	Diagnostic accuracy	Detection rate	Accuracy, F1-score	Pooled AUC
Dataset Size	Multi-device study	Pregnant cohort	Public ECG datasets	Multiple studies
Sampling Rate (Hz)	Device-dependent	HSAT standard	ECG standard	-
Feasibility	High	High	High	High
Reliability / Explainability	Moderate	High	Moderate	High
Speed	Fast	Moderate	Fast	-

Table 5. Detection Based on Machine Learning Models

Attribute	**[19]**	**[20]**	**[21]**	**[22]**
Authors & Year	Wang & Coauthors, 2024 [19]	Wang & Coauthors, 2024 [20]	Zhang & Coauthors, 2024 [21]	Siriyotha & Coauthors, 2021 [22]
Features / Techniques Used	Demographic + physiological ML features	Clinical + oximetry features	ML features with SHAP analysis	Prediction model meta-analysis
Data Type / Signal	PSG + clinical data	SpO_2, demographics	Clinical + questionnaire data	Clinical data
Artifact Removal	Standard preprocessing	Signal filtering	Data cleaning	Bias assessment
Model / Algorithm	ML models	ML classifiers	XGBoost + SHAP	Meta-analysis
Evaluation Metric (reported)	AUC, accuracy	AUC, sensitivity	AUC, accuracy	Pooled AUC
Dataset Size	Large pregnancy cohort	Pregnant women	Pregnancy cohort	Multiple studies
Sampling Rate (Hz)	PSG standard	Oximetry standard	-	-
Feasibility	High	High	High	High
Reliability / Explainability	Moderate	Moderate	High	High
Speed	Fast	Fast	Fast	-

Table 6. Detection Based on Machine Learning Models

Attribute	**[24]**	**[38]**
Authors & Year	Kim & Coauthors, 2015 [24]	Rana & Coauthors, 2024 [38]
Features / Techniques Used	Snoring intensity analysis	Explainable ML for clinical decision support
Data Type / Signal	Audio signals	Clinical + physiological data
Artifact Removal	Noise filtering	Data cleaning
Model / Algorithm	Statistical correlation	Explainable ML (SHAP, LIME)
Evaluation Metric (reported)	Correlation coefficient	Accuracy, interpretability
Dataset Size	Clinical cohort	Clinical datasets
Sampling Rate (Hz)	Audio standard	Dataset-dependent
Feasibility	Moderate	High
Reliability / Explainability	High	Very High
Speed	Fast	Fast

Table 7. OSA detection based on Deep Learning Models

Attribute	[5]	[7]	[9]	[12]
Authors & Year	Jiang & Coauthors, 2023 [5]	Zhang & Coauthors, 2025 [7]	Mukherjee & Coauthors, 2021 [9]	Hu & Coauthors, 2025 [12]
Features / Techniques Used	Deep CNN feature extraction	Multimodal fusion + multi-scale attention	Deep ensemble learning	Transformer- enhanced multimodal sensing
Data Type / Signal	PPG	EEG, ECG, SpO_2	PSG	Radar + physiological signals
Artifact Removal	Motion artifact filtering	Signal synchronization & normalization	Standard preprocessing	Signal fusion preprocessing
Model / Algorithm	CNN	Multiscale Transformer	CNN ensemble	Transformer
Evaluation Metric (reported)	AUC, Accuracy	Accuracy, AUC	Accuracy, AUC	Accuracy, AUC
Dataset Size	Public PPG datasets	Large public datasets	Public datasets	Experimental datasets
Sampling Rate (Hz)	˜100 Hz	Dataset-dependent	PSG standard	Sensor-dependent
Feasibility	High	High	High	Moderate
Reliability / Explainability	Low–Moderate	Moderate	Low	Moderate
Speed	Fast	Moderate–Fast	Moderate	Moderate

Table 8. Detection Based on Deep Learning Models

Attribute	[25]	[34]	[35]	[37]
Authors & Year	Wang & Coauthors, 2024 [25]	Almazaydeh & Coauthors, 2025 [34]	Biswal & Coauthors, 2022 [35]	Phan & Coauthors, 2023 [37]
Features / Techniques Used	Hybrid CNN– Transformer	AI-based ECG & EEG features	Deep physiological representation learning	Multimodal deep learning fusion
Data Type / Signal	Radar signals	ECG, EEG	Multimodal PSG	EEG, ECG, airflow
Artifact Removal	Radar preprocessing	Signal denoising	Standard preprocessing	Signal smoothing
Model / Algorithm	CNN	DL models	Deep learning models	Deep multimodal networks
Evaluation Metric (reported)	Accuracy, AUC	Accuracy, sensitivity	Accuracy, AUC	Accuracy, AUC
Dataset Size	Experimental dataset	Public datasets	Public datasets	Public datasets
Sampling Rate (Hz)	Radar-dependent	EEG/ECG standard	PSG standard	PSG standard
Feasibility	Moderate	High	High	High
Reliability / Explainability	Moderate	Moderate	Low–Moderate	Moderate
Speed	Moderate	Fast	Moderate	Moderate

Table 9. OSA detection based on Artificial Intelligence Models

Attribute	**[6]**	**[16]**	**[23]**	**[26]**
Authors & Year	Nugent & Coauthors, 2025 [6]	Collop & Coauthors, 2023 [16]	Tran & Coauthors, 2025 [23]	Wang & Coauthors, 2025 [26]
Features / Techniques Used	Two-step clinical screening workflow	Clinical guideline review	Radar signal feature analysis	Oximetry + demographic features
Data Type / Signal	Questionnaire + HSAT/PSG	HSAT	Radar-based respiration	SpO_2, anthropometric data
Artifact Removal	Standard clinical preprocessing	-	Frequency filtering	Signal smoothing
Model / Algorithm	Clinical trial protocol (no ML)	Narrative review	Diagnostic meta- analysis	ML classifiers
Evaluation Metric (reported)	Recruitment rate, CPAP adherence	Diagnostic validity	Sensitivity, specificity	Accuracy, AUC
Dataset Size	96 invited, 36 enrolled	Multiple studies	Multiple studies	Clinical dataset
Sampling Rate (Hz)	EEG/ECG standard	-	Radar-dependent	Oximetry standard
Feasibility	High	High	Moderate	High
Reliability / Explainability	High	High	Moderate	Moderate
Speed	Moderate	-	Fast	Fast

Table 10. Detection Based on Artificial Intelligence Models

Attribute	**[27]**	**[28]**	**[29]**	**[30]**
Authors & Year	Hu & Coauthors, 2025 [27]	Tran & Coauthors, 2025 [28]	Senaratna & Coauthors, 2024 [29]	Liang & Coauthors, 2025 [30]
Features / Techniques Used	Transparent AI, interactive explainability	Wearable AI-based feature extraction	AI scoping review	Diagnostic accuracy meta-analysis
Data Type / Signal	Multimodal physiological	Wearable physiological signals	PSG, wearables, radar	Multimodal physiological
Artifact Removal	Scenario-specific preprocessing	Motion artifact reduction	-	Bias filtering
Model / Algorithm	Explainable AI framework	ML & DL models	Review	Meta-analysis
Evaluation Metric (reported)	Accuracy, interpretability score	Accuracy, AUC	Performance trends	Sensitivity, specificity
Dataset Size	Multi-scenario datasets	Multiple studies	Multiple studies	Multiple studies
Sampling Rate (Hz)	Dataset-dependent	Wearable standard	-	-
Feasibility	High	High	High	High
Reliability / Explainability	Very High	Moderate	High	High
Speed	Moderate	Fast	-	-

Table 11. Detection Based on Artificial Intelligence Models

Attribute	[31]	[32]	[33]	[36]
Authors & Year	Tantrakul & Coauthors, 2022 [31]	Abu Bakar & Coauthors, 2024 [32]	Louis & Coauthors, 2024 [33]	Faust & Coauthors, 2021 [36]
Features / Techniques Used	ML screening review	Wearable AI systematic review	Clinical narrative review	ML-based OSA detection review
Data Type / Signal	PSG, SpO_2	ECG-based wearables	PSG and maternal outcomes	PSG, ECG, SpO_2
Artifact Removal	-	Study quality filtering	-	-
Model / Algorithm	Review	Meta-analysis	Review	Review
Evaluation Metric (reported)	Diagnostic accuracy	Accuracy, AUC	Clinical impact	Comparative accuracy
Dataset Size	Multiple studies	Multiple studies	Multiple studies	Multiple studies
Sampling Rate (Hz)	-	ECG standard	-	-
Feasibility	High	High	High	High
Reliability / Explainability	High	High	High	High
Speed	-	-	-	-