A Machine Learning Approach to Sleep Study Selection for Diagnosing Obstructive Sleep Apnea

BUSINESS PROBLEM

Obstructive sleep apnea (OSA) affects an estimated one billion adults worldwide, yet most remain undiagnosed. Diagnosis requires a clinical sleep study, either a Home Sleep Test (HST) or an in-lab Administered Sleep Test (AST). Matching each patient to the right modality is a complex decision influenced by clinical, operational, and logistical factors. Resmed's cloud-based diagnostic platform is uniquely positioned to leverage its rich historical data to enhance this process. This project explored whether machine learning can augment clinician decision-making to improve test completion rates, reduce retests, and accelerate time-to-diagnosis.

DATA SOURCES

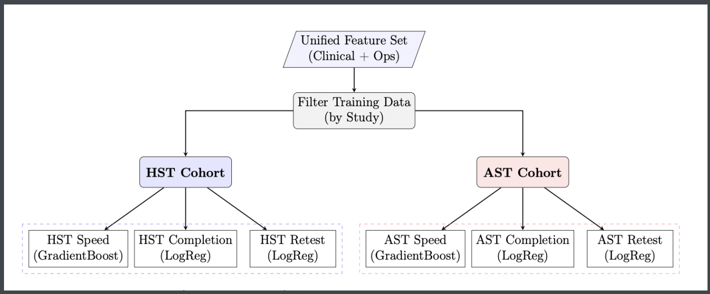
De-identified records from Resmed's diagnostic platform spanning 2021 to 2025. The dataset includes patient clinical histories, anthropometric measurements, facility-level operational metrics, insurance and referral records, and scheduling data. Features were engineered to capture the clinical, operational, and logistical factors that influence diagnostic outcomes.

Data Types and Format

Boolean, Categorical (nominal), Count (integer), Numeric, Ordered (ordinal), Relational Data, Structured, Text

APPROACH

Six predictive models were trained on historical patient and facility data from Resmed's diagnostic platform, estimating three outcomes for each modality: the probability of test completion, expected time-to-result, and retest risk. Logistic Regression classifiers were used for the binary outcomes to ensure clinical interpretability, while Gradient Boosting Regressors captured non-linear patterns in time-to-result prediction. These model outputs feed into a composite utility function that ranks HST and AST for each patient, producing a configurable recommendation tunable to clinical priorities.



John Martins
MIT Leaders for Global Operations, Class of 2026

Leaders for Global Operations

IMPACT

The recommendation policy was evaluated on a held-out test set by simulating what outcomes would have occurred had the model's recommendations been followed instead of historical selections. The primary deployment candidate, the Balanced Strategy, projected a 7.8 percentage point increase in test completion rates and an 8.2 percentage point reduction in diagnostic retests over the historical baseline. These gains translate directly to more patients reaching diagnosis on their first attempt and fewer consuming pipeline capacity through repeated testing. Five strategic scenarios demonstrated the policy's configurability. The Max Completion strategy achieved an 8.8 percentage point completion lift, while the Retest Averse strategy achieved a 16.3 percentage point retest reduction, illustrating how the utility function's tunable weights allow Resmed to align the tool with evolving clinical priorities. Even the Speed First scenario, which most closely mirrored historical modality volumes, projected improvements across all three objectives, confirming that value comes from smarter patient-to-modality matching rather than shifting volume between pathways. The underlying models demonstrated strong predictive validity. The HST Completion model achieved an ROC-AUC of 0.94 and the AST Completion model achieved 0.80, confirming that diagnostic outcomes can be reliably predicted from data available at intake. Feature importance analysis revealed that facility-level operational metrics and patient engagement history are stronger predictors of diagnostic success than clinical factors alone, giving Resmed actionable insight into the levers that drive outcomes.

DRIVERS

Resmed's cloud-based diagnostic platform processes high volumes of sleep study data, creating a strong foundation for machine learning. Growing demand for OSA diagnosis and an industry-wide push toward data-driven clinical decision support motivated the development of a tool that enhances clinician workflows while improving patient outcomes.

BARRIERS

Privacy-preserving data protocols excluded demographic features such as biological sex, a known predictor of OSA presentation. Missing anthropometric data required careful imputation strategies. Retrospective evaluation cannot fully capture behavioral changes upon deployment, motivating a phased rollout to validate performance in production.

ENABLERS

Resmed's diagnostic platform provided a rich, centralized dataset spanning four years of patient and facility records. Close collaboration between the Product Management division and clinical stakeholders ensured the models reflected real-world operational workflows. Engineering and management advisors helped shape a solution balancing technical rigor with deployment practicality.

ACTIONS

Historical patient and facility data were extracted into a unified feature set spanning five functional categories. Six predictive models were trained for each modality and outcome, then combined through a composite utility function to produce patient-level recommendations. The policy was evaluated across five scenarios on a held-out test set, quantifying improvements in completion, retest reduction, and speed over the clinician baseline.

INNOVATION

Rather than predicting OSA diagnosis, this approach models operational outcomes of the diagnostic pathway, treating modality selection as an optimization problem. A composite utility function enables configurable trade-offs between competing objectives, adapting to clinical priorities without retraining. Feature importance analysis revealed that facility operations and patient engagement outperform clinical indicators as predictors of success.

IMPROVEMENT

The Balanced Strategy projected a 7.8 percentage point increase in test completion and an 8.2 percentage point reduction in diagnostic retests over the historical baseline. The model achieved an ROC-AUC of 0.94 for HST completion prediction and 0.80 for AST, confirming reliable performance. Across all five scenarios, the policy demonstrated that smarter patient-to-modality matching improves outcomes without additional clinical resources.

BEST PRACTICES

Probability calibration is essential when model outputs feed into a decision function. Time-based train/test splits better simulate deployment than random shuffling. Prioritizing model interpretability builds clinician trust and accelerates adoption. Configurable weights let stakeholders tune priorities without retraining.

OTHER APPLICATIONS

The framework generalizes to any clinical triage decision where multiple pathways have distinct operational constraints. Within Resmed, it could extend to CPAP adherence prediction or therapy selection. More broadly, the approach applies to healthcare scheduling, radiology routing, or any domain where patient-to-resource matching drives outcomes.