Artificial Intelligence Does Not Always Win

Noam Shomron PhD[1] and Yariv Yogev MD[1,2]

[1]Gray Faculty of Medical and Health Sciences, Tel Aviv University, Tel Aviv, Israel
[2]Department of Obstetrics and Gynecology, Lis Hospital for Women's Health, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel

KEY WORDS: artificial intelligence (AI), fetal weight estimation, Hadlock formula, machine learning

IMAJ 2026; 28: 260–261

For editorial see page 262

Artificial intelligence (AI) has become the emblem of progress. We are told it learns faster, sees patterns invisible to the human eye, and will soon outthink us in every domain, from finance to philosophy, from language to life. In medicine, where decisions carry the weight of saving lives, this narrative has gained traction. Algorithms promise precision without fatigue, accuracy without bias, and reproducibility without emotion.

Yet, sometimes, the data tell a quieter story.

In our recent multicenter study of fetal weight estimation [1], we tested whether machine learning (ML) could outperform the venerable *Hadlock formula,* a method derived nearly four decades ago from fewer than 300 pregnancies. Using data from almost 10,000 births across Israel and Canada, our machine learning models–LightGBM, XGBoost, and neural networks–achieved superior numerical performance. Errors were smaller and predictions were smoother. It seems that, statistically, AI has effectively surpassed the former technique

Clinically, it barely mattered.

When evaluated on clinically meaningful classification tasks, identifying fetuses at risk of being small for gestational age (SGA) or large for gestational age (LGA), the apparent statistical advantage of ML largely vanished. Across the multicenter cohort, sensitivity for SGA detection ranged from 73% to 75% for ML models versus 74–% to 76% for the Hadlock formula, with comparable accuracy (91–92% vs. 92%, respectively). Similar patterns were observed for LGA classification, where sensitivity ranged from 52–60% for ML and 65–68% for Hadlock, while accuracy remained nearly identical (approximately 91–92% across methods). Thus, despite lower root mean square error (RMSE) and smoother predictions, AI did not translate its numerical superiority into a clinically meaningful improvement in risk stratification. In practice, both approaches flagged largely the same pregnancies for clinical follow-up, underscoring that improved regression accuracy did not meaningfully alter management decisions.

This distinction matters. In medicine, the race is not for decimal points but for decisions. Predicting 100 grams closer to the true birth weight changes little if both models classify the same fetus as normal. The obsession with numerical supremacy, lower RMSE, and higher F1 risks confusing statistical superiority with clinical significance.

AI's appeal lies in its ambition to optimize everything. However, in that ambition lies its blind spot. Machine learning does not *understand* context; it recognizes patterns. It cannot be intuitive when a measurement feels inconsistent with a mother's history, when anxiety demands reassurance more than another scan, or when the human conversation itself is the intervention. The Hadlock formula does not endure because it is perfect, but rather because it is transparent. Clinicians can see *why* it works and *when* it might fail.

There is humility in simplicity. Each layer of technological abstraction, each *black box*, adds a veneer of sophistication while eroding interpretability. In an age that celebrates disruption, our study reminds us that stability can be equally revolutionary. Sometimes, the old formula wins, not despite its age, but because it was built on human understanding rather than mathematical recursion.

This discussion is not an argument against AI, but rather against its uncritical worship. AI is reshaping medicine, as it already has in imaging, pathology, and drug discovery. Yet it must do so through collaboration, not conquest. The goal is not to replace clinicians with code, but to *enhance* their intuition, to illuminate patterns that human eyes cannot see, while still deferring to the judgment that only humans can make.

Science progresses through iteration, not infatuation. The true test of innovation is not whether it dazzles in benchmarks, but whether it endures in practice. As researchers, we must resist the cultural pressure to crown AI as the inevitable successor of human reasoning. In many cases, like ours, it is merely the next apprentice in a lineage of tools, awaiting careful supervision.

In fields such as oncology, cardiology, and genomics, AI has indeed demonstrated tangible clinical benefit in specific, well-defined tasks, particularly in imaging-based screening, pattern recognition, and high-throughput data interpretation [2-4]. Notable success include screening of electronic medical records and selected radiological applications, where decision boundaries are narrow and outcomes well constrained. However, these successes cannot be generalized universally. As our obstetric study illustrates, in domains where clinical decisions hinge on contextual judgment, threshold-based classification, and downstream management rather than pure prediction accuracy, improved model performance does not necessarily confer improved care.

AI does not always win. The lesson from obstetrics is therefore not that AI fails, but that its value is domain-specific and inseparable from clinical context. It is a reminder that progress is not linear, that the human element remains the indispensable core of scientific discovery [5]. In the contest between the formula and the model, between intuition and automation, medicine finds its most valuable truth: that reason and compassion still belong to the same equation.

Correspondence

Dr. N. Shomron
Gray Faculty of Medical and Health Sciences, Tel Aviv University, Tel Aviv 69978, Israel
Email: nshomron@tauex.tau.ac.il

References

1. Dor O, Ashwal E, Cohen M, Rottenstreich O, Yogev Y, Shomron N, Rottenstreich M. Machine learning versus traditional formulas for fetal weight estimation: an international multicenter study evaluating prediction accuracy across birth weight percentiles. *Int J Gynaecol Obstet* 2025; 00: 1-7.
2. Schwammenthal Y, Rabinowitz T, Basel-Salmon L, Tomashov-Matar R, Shomron N. Noninvasive fetal genotyping using deep neural networks. *Brief Bioinform* 2024; 26 (1): bbaf067.
3. Petrick LM, Shomron N. AI/ML-driven advances in untargeted metabolomics and exposomics for biomedical applications. *Cell Rep Phys Sci* 2022; 3 (7): 100978.
4. Zoabi Y, Kehat O, Lahav D, Weiss-Meilik A, Adler A, Shomron N. Predicting bloodstream infection outcome using machine learning. *Sci Rep* 2021; 11 (1): 20101.
5. Shomron N. Don't Devoid Emotions from Science. *J Mol Neurosci* 2025; 74 (4): 111.

Concentrate all your thoughts upon the work at hand.
The sun's rays do not burn until brought to a focus.

Alexander Graham Bell (1847–1922), Scottish-born Canadian-American inventor, scientist, and engineer who is credited with patenting the first practical telephone

Capsule

Single-cell atlas of the developing Down syndrome brain cortex

Lattke and co-authors performed single-cell transcriptomic and chromatin accessibility profiling of approximately 250,000 cells from 15 DS and 15 control human fetal cortices (10–20 weeks postconception). The analysis revealed a subtype-specific reduction in RORB- and FOXP1-expressing excitatory neurons and widespread disruption of neurodevelopmental transcriptional programs. Chromosome 21 transcription factors BACH1, PKNOX1, and GABPA emerged as dosage-sensitive hubs regulating genes linked to intellectual disability. Antisense oligonucleotide-mediated normalization of these transcription factors in human neural progenitors in vitro partially rescued target gene expression. Benchmarking a humanized in vivo model captured additional molecular and cellular signatures of DS, complementing the in vitro model. Together, the authors presented a resource defining the gene-regulatory landscape underlying cortical development in DS and highlight molecular pathways for further investigation.

Nature Medicine 2026; 32: 1061

Eitan Israeli