AI in Healthcare: Case Studies

SUCCESS

Diabetic Retinopathy AI

Google Health / Screening Tool

90%	98%	50K+
Sensitivity	Specificity	Patients

- Strong clinical validation across populations
- Clear regulatory pathway (FDA approved)
- Effective provider training program

LESSONS

Watson for Oncology

IBM / Treatment Recommendations

Limited	High	Variable
Adoption	Cost	Results

- Insufficient local data adaptation
- Over-promised capabilities
- Poor integration with workflows

EVOLVING

Sepsis Prediction Model

Epic Systems / Risk Alert

63%	Wide	Mixed
PPV	Deploy	Reviews

- Alert fatigue concerns addressed
- Continuous model improvement
- Need for local calibration

LANDMARK

IDx-DR

First Autonomous AI / FDA De Novo

1st	87%	2018
FDA AI	Accuracy	Approved

- Clear intended use definition
- Robust clinical trial design
- Set regulatory precedent

Detailed Case Analysis

Google Health: Diabetic Retinopathy AI System

A Success Story in AI-Driven Screening and Early Detection

Implementation Workflow

1. Image Capture

Retinal fundus images captured using standard cameras in primary care settings

↓

2. AI Analysis

Deep learning model processes images in real-time (< 30 seconds)

↓

3. Risk Classification

Binary output: Referable vs. Non-referable diabetic retinopathy

↓

4. Clinical Action

Immediate referral to ophthalmologist if positive, routine monitoring if negative

Performance Metrics

90.3% Sensitivity	**98.1%** Specificity	**87.2%** AUC
50,000+ Patients Screened	**11** Countries	**2016** First Deployment

Success Factors

- **Rigorous Validation:** Trained on 128,000 images and validated across diverse populations in India, Thailand, and the United States
- **Clinical Integration:** Designed to work with existing screening infrastructure without requiring specialist equipment
- **Clear Use Case:** Focused on binary classification (refer/don't refer) rather than complex diagnosis
- **Regulatory Approval:** CE marked in Europe and received regulatory clearance in multiple countries
- **Provider Training:** Comprehensive training programs for non-ophthalmologist healthcare workers
- **Quality Control:** Built-in image quality assessment to ensure reliable results

Clinical Impact

- **Access Expansion:** Enabled diabetic retinopathy screening in areas with no ophthalmologists
- **Early Detection:** Identified referable cases 6-12 months earlier than standard care pathways
- **Cost Effectiveness:** Reduced screening costs by 30-40% compared to traditional methods
- **Time Efficiency:** Results available in under 30 seconds vs. days for specialist review
- **Patient Outcomes:** Increased screening participation rates by 35% in underserved areas
- **Scalability:** Demonstrated feasibility in low-resource settings with minimal infrastructure

Key Takeaway

Google Health's diabetic retinopathy AI demonstrates that success in medical AI requires more than just high accuracy. The combination of rigorous multi-population validation, seamless clinical workflow integration, clear regulatory pathways, and focus on a well-defined clinical need created a sustainable and impactful solution. This case shows that AI tools succeed when they augment rather than replace clinical expertise, particularly in screening applications where early detection can significantly improve patient outcomes.

IBM Watson for Oncology

Lessons from Implementation Challenges in Clinical Decision Support

Implementation Timeline

2013-2015: Development

Partnership with Memorial Sloan Kettering, trained on treatment guidelines and case data

↓

2015-2017: Global Expansion

Rapid deployment to hospitals worldwide, high initial enthusiasm and investment

↓

2017-2018: Challenges Emerge

Reports of unsafe recommendations, poor concordance with local practices

↓

2018-2019: Scaling Back

Multiple hospitals discontinued use, IBM reassessed strategy

Challenge Metrics

$62M Annual License Cost (per hospital)	**34%** Concordance Rate (Some Studies)	**78%** Oncologists Disagreed
230 Hospitals at Peak	**~50** Hospitals Discontinued	**Mixed** Clinical Utility

Key Failure Points

- **Training Data Limitations:** Primarily trained on MSK's protocols and synthetic cases rather than diverse real-world data

Important Lessons Learned

- **Context Matters:** AI systems must be trained and validated on data representative of their deployment environment

- **Geographic Variability:** Recommendations often didn't align with local treatment standards, drug availability, or insurance coverage
- **Lack of Transparency:** "Black box" nature made it difficult for oncologists to understand reasoning behind recommendations
- **Workflow Integration:** Required significant time investment to input patient data with minimal workflow efficiency gains
- **Overpromised Capabilities:** Marketing suggested AI could match expert oncologists, reality showed significant limitations
- **Insufficient Validation:** Limited prospective clinical trials before widespread deployment
- **Cost vs. Value:** High licensing costs not justified by clinical utility or improved outcomes

- **Clinical Expertise Required:** Complex medical decision-making requires nuanced understanding that current AI cannot fully replicate
- **Transparency is Critical:** Clinicians need to understand how AI arrives at recommendations to trust and use them effectively
- **Realistic Expectations:** AI should augment clinical decision-making, not attempt to replace expert judgment
- **Workflow Design:** Technology must integrate seamlessly into existing clinical workflows or provide clear time-saving benefits
- **Continuous Learning:** AI systems need mechanisms for local adaptation and continuous improvement based on feedback
- **Value Proposition:** High costs must be justified by demonstrable improvements in efficiency, accuracy, or patient outcomes

⚠ Key Takeaway

Watson for Oncology's challenges illustrate that technological sophistication alone doesn't guarantee clinical success. The case demonstrates the critical importance of training data diversity, local contextualization, transparency in AI reasoning, realistic capability expectations, and genuine value addition to clinical workflows. Perhaps most importantly, it shows that AI in complex medical decision-making should serve as a decision support tool that augments physician expertise rather than attempting to replace it. The disconnect between marketing promises and clinical reality eroded trust and adoption.

Epic Sepsis Prediction Model

An Evolving Case Study in Real-Time Risk Prediction and Alert Management

Prediction & Alert System

1. Continuous Monitoring

Real-time analysis of vital signs, lab values, and clinical data from EHR

↓

2. Risk Calculation

Machine learning model calculates sepsis risk score every 15 minutes

↓

3. Alert Generation

Automated alert triggered when risk threshold exceeded

↓

4. Clinical Response

Care team evaluates patient and initiates sepsis protocol if appropriate

Performance Profile

63% Positive Predictive Value	**77%** Sensitivity	**67%** Specificity
500+ Hospitals Using	**37%** False Positive Rate	**2017** Initial Release

Ongoing Challenges

- **Alert Fatigue:** High false positive rate (37%) leads to clinician desensitization and alert dismissal

Positive Developments

- **Wide Adoption:** Deployed in over 500 hospitals, indicating perceived clinical value despite challenges

- **Variable Performance:** Prediction accuracy varies significantly across different patient populations and hospital settings
- **Definition Issues:** Sepsis diagnostic criteria evolve (Sepsis-3), requiring model retraining and recalibration
- **Timing Concerns:** Questions about optimal prediction window - too early increases false positives, too late reduces intervention efficacy
- **Local Calibration:** Model requires hospital-specific tuning to account for local patient mix and clinical practices
- **Outcome Measurement:** Difficulty in measuring actual impact on mortality and morbidity due to confounding factors

- **Continuous Improvement:** Epic regularly updates model based on aggregated data and clinical feedback
- **Integration Success:** Seamlessly embedded in Epic EHR workflow, requiring no additional documentation
- **Early Warning Potential:** Can identify at-risk patients 4-6 hours before clinical recognition in some cases
- **Customization Options:** Hospitals can adjust sensitivity thresholds based on their risk tolerance and resources
- **Research Platform:** Provides valuable data for studying sepsis prediction and alert system optimization

⚠ Key Takeaway

The Epic sepsis model represents the reality of many AI clinical decision support systems: promising but imperfect, widely deployed yet controversial. It demonstrates that even with widespread adoption, AI prediction tools face ongoing challenges in balancing sensitivity and specificity, managing alert fatigue, and proving definitive clinical benefit. The case illustrates the importance of continuous model refinement, local customization, and integration with broader quality improvement initiatives. Success may depend less on perfect prediction and more on thoughtful implementation that considers local context, provider workflow, and realistic expectations about AI's capabilities in complex, time-sensitive clinical scenarios.

☑ IDx-DR: First Autonomous AI Diagnostic System

Landmark FDA Approval and Regulatory Pathway for Autonomous Medical AI

Regulatory Journey

2010-2016: Development

Algorithm development and extensive clinical validation studies

↓

2017: FDA Submission

De Novo pathway submission with comprehensive clinical data

↓

April 2018: FDA Approval

First autonomous AI diagnostic system approved - no specialist oversight required

↓

2018-Present: Market Entry

Commercial deployment, reimbursement establishment, ongoing monitoring

Clinical Trial Results

87.4% Sensitivity (Referable DR)	**89.5%** Specificity	**900** Patients in Trial
10 Primary Care Sites	**96%** Image Gradability	**Class II** FDA Classification

Regulatory Success Factors

- **Clear Intended Use:** Precisely defined indication - detection of more than mild diabetic retinopathy in adults with diabetes
- **Prospective Clinical Trial:** Well-designed, multi-center study with pre-specified endpoints comparing to expert graders

Broader Impact

- **Regulatory Precedent:** Established pathway for autonomous AI diagnostics, paving way for future applications
- **Reimbursement Model:** Achieved CPT code and CMS reimbursement, creating economic viability for AI diagnostics

- **Autonomous Function:** Designed to provide screening decision without clinician interpretation of images or results
- **Quality Controls:** Built-in image quality assessment ensures only analyzable images are processed
- **Point-of-Care Design:** Intended for use in primary care settings by non-specialist healthcare providers
- **Locked Algorithm:** Submitted as a fixed algorithm (not continuously learning) to meet regulatory requirements
- **Risk-Benefit Analysis:** Demonstrated that benefits of increased screening access outweigh risks of false positives/negatives

- **Access Improvement:** Enables screening in settings without ophthalmology expertise, particularly benefiting underserved populations
- **Validation Standard:** Set expectations for rigor required in AI clinical validation for regulatory approval
- **Patient Safety Framework:** Demonstrated approach to AI safety monitoring and quality assurance in autonomous systems
- **Commercial Viability:** Proved business model for AI medical devices as standalone diagnostic tools
- **International Recognition:** FDA approval facilitated regulatory approvals in other countries

⚠ Key Takeaway

IDx-DR's FDA approval represents a watershed moment for medical AI, demonstrating that autonomous diagnostic systems can meet regulatory standards when properly validated and positioned. The success stemmed from several factors: a clearly defined and bounded clinical use case, rigorous prospective validation, thoughtful consideration of the point-of-care context, and comprehensive quality control mechanisms. By achieving both regulatory approval and reimbursement, IDx-DR established a viable model for AI medical devices. Most significantly, it showed that autonomous AI can be deployed safely in healthcare settings without specialist oversight when the clinical application is appropriate, the validation is thorough, and proper safeguards are in place. This case provides a blueprint for future autonomous AI diagnostic systems.