

## Paper of the Quarter – QIV/2024 – [STOP-FSGS](#)

### Ecologically sustainable benchmarking of AI models for histopathology

NPJ Digit Med. 2024 Dec 24. [>>Pubmed-Link<<](#)

Yu-Chia Lan, Martin Strauch, Pourya Pilva, Nikolas E J Schmitz, Alireza Vafaei Sadr, Leon Niggemeier, Huong Quynh Nguyen, David L Hölscher, Tri Q Nguyen, Jesper Kers, Roman D Bülow, Peter Boor

---

Die digitale Pathologie steht an der Spitze medizinischer Innovation im Bereich der verbesserten Diagnoseverfahren, insbesondere durch den Einsatz von künstlicher Intelligenz (KI). Allerdings liegt der Fokus bei der Entwicklung dieser leistungsstarken, aber rechenintensiven Modelle meist ausschließlich auf deren diagnostischer Genauigkeit, während ihr Energieverbrauch und ihre Nachhaltigkeit oft vernachlässigt werden. Ein Forschungsteam der Uniklinik RWTH Aachen hat daher den **Environmentally Sustainable Performance (ESPer) Score** entwickelt, der diagnostische Performance und CO<sub>2</sub>-Emissionen kombiniert.

Dieser neue Benchmark integriert erstmals sowohl die diagnostische Leistung als auch die CO<sub>2</sub>-Emissionen von KI-Modellen in einer einzigen Bewertung. In ihrer Studie analysierten die Forscher:innen fünf etablierte KI-Modelle hinsichtlich zweier diagnostischer Aufgaben: die Klassifikation von Nierentransplantationserkrankungen und die Unterscheidung verschiedener Subtypen von Nierenzellkarzinomen. Während alle Modelle eine akzeptable diagnostische Genauigkeit erreichten, offenbarte der ESPer-Score erhebliche Unterschiede in deren CO<sub>2</sub>-Fußabdruck. Das TransMIL-Modell zeigte in beiden Aufgaben eine hohe Präzision bei gleichzeitig vergleichsweise niedrigem Energieverbrauch. Das neuartige Foundation-Modell Prov-GigaPath erzielte eine ähnlich gute diagnostische Leistung, verursachte jedoch eine deutlich höhere Umweltbelastung.

Um die Nachhaltigkeit von KI in der Pathologie weiter zu optimieren, untersuchten die Wissenschaftler:innen zusätzlich verschiedene Strategien zur Reduzierung des CO<sub>2</sub>-Ausstoßes mithilfe des ESPer-Scores. Dazu gehörten unter anderem die Anpassung der verarbeiteten Datenmengen während der Inferenz, die Wahl einer geeigneten Bildauflösung sowie eine gezielte Reduktion der zu analysierenden Bildbereiche. Die Ergebnisse zeigen, dass es durchaus möglich ist, den Energieverbrauch signifikant zu senken, ohne die diagnostische Genauigkeit wesentlich zu beeinträchtigen.

Der ESPer-Score stellt nicht nur ein vielversprechendes Werkzeug für die Forschung dar, sondern könnte auch als neuer Standard für die Industrie und Regulierungsbehörden dienen. Durch seine Anwendung könnten KI-Modelle nicht nur leistungsfähiger, sondern auch umweltfreundlicher gestaltet werden. Nicht zuletzt wurde dies sehr deutlich mit der vor kurzem erfolgten Liveschaltung von Deep Seek. Die Forscher:innen hoffen, mit ihrer Studie einen wichtigen Impuls für nachhaltige Innovationen an der Schnittstelle von Medizintechnologie und Umweltschutz zu geben.

---

**Autor/in:** Yu-Chia Lan, Peter Boor

**Kontakt:** [ylan@ukaachen.de](mailto:ylan@ukaachen.de), [pboor@ukaachen.de](mailto:pboor@ukaachen.de)

## Paper of the Quarter – QIV/2024 – [STOP-FSGS](#)

### Ecologically sustainable benchmarking of AI models for histopathology

NPJ Digit Med. 2024 Dec 24. [>>Pubmed-Link<<](#)

Yu-Chia Lan, Martin Strauch, Pourya Pilva, Nikolas E J Schmitz, Alireza Vafaei Sadr, Leon Niggemeier, Huong Quynh Nguyen, David L Hölscher, Tri Q Nguyen, Jesper Kers, Roman D Bülow, Peter Boor

---

Digital pathology is at the forefront of medical innovation, particularly in enhancing diagnostic procedures through artificial intelligence (AI). However, the development of these powerful yet computationally intensive models typically focus solely on diagnostic accuracy, while their energy consumption and sustainability are often overlooked. To address this, a research team at RWTH Aachen University Hospital has developed the **Environmentally Sustainable Performance (ESPer) Score**, which combines diagnostic performance with CO<sub>2</sub> emissions.

This new benchmark is the first to integrate both diagnostic accuracy and the carbon footprint of AI models into a single evaluation. In their study, the researchers analyzed five widely used AI models for two diagnostic tasks: classifying kidney transplant diseases and distinguishing subtypes of renal cell carcinoma. While all models achieved acceptable diagnostic accuracy, the ESPer Score revealed significant differences in their CO<sub>2</sub> footprint. The TransMIL model demonstrated high precision in both tasks while maintaining relatively low energy consumption. In contrast, the novel foundation model Prov-GigaPath achieved similar diagnostic performance but with a considerably higher environmental impact.

To further enhance AI sustainability in pathology, the researchers explored strategies to reduce CO<sub>2</sub> emissions using the ESPer Score. These included adjusting the amount of data processed during inference, selecting an optimal image resolution, and limiting the analyzed image regions. The findings show that energy consumption can be significantly reduced without substantially compromising diagnostic accuracy.

The ESPer Score is not only a promising tool for research but could also serve as a new standard for industry and regulatory bodies. Its implementation could help make AI models both more efficient and environmentally friendly. This became very apparent with Deep Seek going live recently. The researchers hope their study will inspire further innovations at the intersection of medical technology and sustainability.

---

**Author:** Yu-Chia Lan, Peter Boor

**Contact:** [ylan@ukaachen.de](mailto:ylan@ukaachen.de), [pboor@ukaachen.de](mailto:pboor@ukaachen.de)