Before Symptoms Appear AI-Powered LifeClock Forecasts Diseases Years

Researchers describe and validate a novel [artificial intelligence](#) (AI) model trained to use routine electronic health records (EHRs) to determine biological age across the entire human lifespan.

The model, named "[LifeClock](#)," identified two distinct clocks, one for pediatric development and another for adult aging, and could predict the risk of major diseases years before their occurrence. This framework offers a robust, low-cost tool for advancing precision health, accessibility, and personalized medicine.

Study

The present study addressed this knowledge gap by introducing "LifeClock", a biological clock built on a powerful transformer-based artificial intelligence (AI) model called EHRFormer, which uses input-output dual stochastic masking to handle sparse data, adversarial training to eliminate batch effects, and an [autoregressive design](#) for longitudinal prediction.

The model was trained using an extensive dataset from the China Health Aging Investigation (CHAI) project. The dataset comprised 24.6 million longitudinal clinical visits from 9.6 million unique individuals. This longitudinal data, which tracks patients over time, included 184 routine clinical indicators such as laboratory test results and [vital signs](#).

The EHRFormer model was designed to create a "digital representation" of health by analyzing the sequence of clinical visits and their associated routine EHR records. The model's architecture included sophisticated strategies to address common challenges in medical data, such as imputing missing values and eliminating batch effects (variations across hospitals or equipment). The model was trained on data from healthy individuals to establish a baseline for normal development and [aging](#).

Finally, the study evaluated model performance on separate internal and external datasets, including the [UK Biobank](#), thereby ensuring its predictions were robust and generalizable across different populations.

Results

The study's analyses revealed two distinct and separate biological clocks: (1) a "development clock" for individuals under 18, and (2) an "[aging clock](#)" for adults. Training specialized models for each phase significantly improved prediction accuracy, underscoring distinct biological processes during development versus aging.

The biomarkers driving these clocks were almost entirely different. The pediatric clock was strongly influenced by markers related to growth, such as high creatinine and total protein levels. In contrast, the adult clock was found to be driven by indicators of age-related decline, including high urea, low albumin, and high red cell distribution width (RDW).

Encouragingly, both clocks proved highly effective at predicting disease risk. The pediatric clock accurately forecasted the future risk of conditions such as malnutrition, growth, and developmental abnormalities (including growth hormone deficiency).

For example, an analysis of EHR data from children under 12 could predict which individuals in Cluster 14 were at a higher risk for developing pituitary hyperfunction (15.36 times higher risk) and obesity (11.07 times higher risk) later in childhood.

The adult clock similarly proved to accurately predict the risk of major age-related diseases, including diabetes (type 2 diabetes (T2D)), stroke, renal failure, and cardiovascular disease (CVD), with specific clusters showing dramatically elevated risks (e.g., Cluster 20 had 37.7 times higher renal failure risk).

Critically, the model distinctly supported both diagnosing current diseases and predicting future risks: after fine-tuning, it achieved an area under the curve (AUC) of 0.98 for diabetes diagnosis and 0.91 for future diabetes prediction. Furthermore, EHRFormer outperformed traditional models (RNN and XGBoost) in both tasks.

Conclusion

The present study successfully demonstrates that EHRFormer can be leveraged to generate a powerful biological clock, constructed from routine, widely available, and low-cost EHR data.

LifeClock provides a novel framework for understanding the distinct processes of pediatric development and adult aging. The model moves beyond simple chronological age assessment to offer a more dynamic and precise picture of an individual's health.

By identifying at-risk individuals years before symptoms appear, this technology holds the potential to revolutionize preventive medicine and guide personalized interventions.

Future work may potentially involve integrating data from wearable devices and other real-time biometric and health data sources to create an even more adaptive and accurate system for promoting healthy aging.

Source:

https://www.news-medical.net/news/20251028/AI-powered-LifeClock-forecasts-diseases-years-before-symptoms-appear.aspx