Docket #: S24-269

Methods for Tokenizing Deep Learning Models from Raw Sequencing Data

Stanford researchers have developed a novel method to process raw genomic sequencing data into deep learning models directly, eliminating the dependency on genome assemblies with applications in drug discovery and diagnostics.

Current deep learning models for genomic data are constrained by their reliance on genome assemblies as inputs and cannot process raw genomic sequencing data directly. This constraint significantly reduces the scope of training data available, thereby limiting the predictive power and applicability of these models. Consequently, this hinders the ability of AI to deliver its full potential in areas such as drug discovery, diagnostics, and biological research.

To address this need, Stanford researchers have developed a novel method to process raw genomic sequencing data into deep learning models directly, eliminating the dependency on genome assemblies. By leveraging raw sequencing data, this approach produces models with superior zero-shot prediction accuracy, enabling their application across a broader range of genomic data. This innovation will empower researchers and companies to tackle complex challenges in drug discovery and diagnostics, affording new opportunities for precision medicine.

Stage of Development:

Prototype. The next steps include scaling up the model size and fine-tuning it for specific clinical applications, such as disease diagnostics and therapeutic development.

Applications

- Drug discovery
- Patient diagnostics
- Experimental design in biological research

Advantages

- Uniquely capable of processing raw sequencing data to generate AI models, unlike existing models that rely on genome assemblies.
- Expanded the training data scope, empowering more robust and generalizable AI models
- Faster analysis of genomic data
- Scalable impact from diagnostics to biological discovery

Patents

- Published Application: 20260018248

Innovators

- Julia Salzman

Licensing Contact

Seth Rodgers

Licensing Manager, Life Sciences

Email