

DECIMER – Künstliche Intelligenz für die Erkennung der Strukturen chemischer Moleküle

Autoren: Kohulan Rajan, Achim Zielesny, Christoph Steinbeck

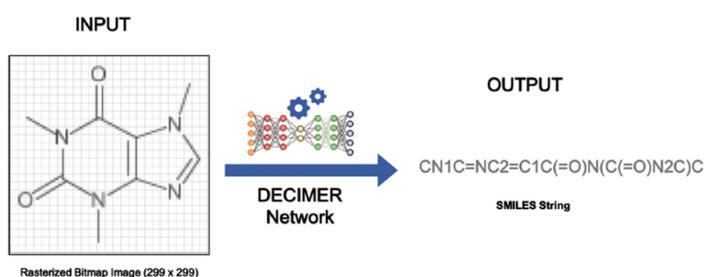


Abb. 1:
Übersetzung des chemischen Formelbildes eines Naturstoffs (hier: Cofein, links) in die zugehörige chemische Information (Atome, chemische Bindungen: SMILES, rechts) mit dem DECIMER-System.

Die vergangenen Jahre sind von erstaunlichen Fortschritten bei der Entwicklung „künstlich intelligenter“ Systeme auf Basis maschinellen Lernens geprägt: Für das komplexe GO-Spiel entwickelt das AlphaGo-Zero-Lernsystem innerhalb von wenigen Trainingsstunden eine die menschlichen Möglichkeiten weit übertreffende Leistungsfähigkeit und schlägt die weltweit besten GO-Spieler, das DeepL-Lernsystem kennzeichnet eine zuvor unerreichte Qualität bei Sprachübersetzungen z.B. vom Deutschen ins Englische. Das AlphaFold-System eröffnet schließlich eine neue Dimension für die erfolgreiche Vorhersage von räumlichen 3D-Proteinstrukturen.

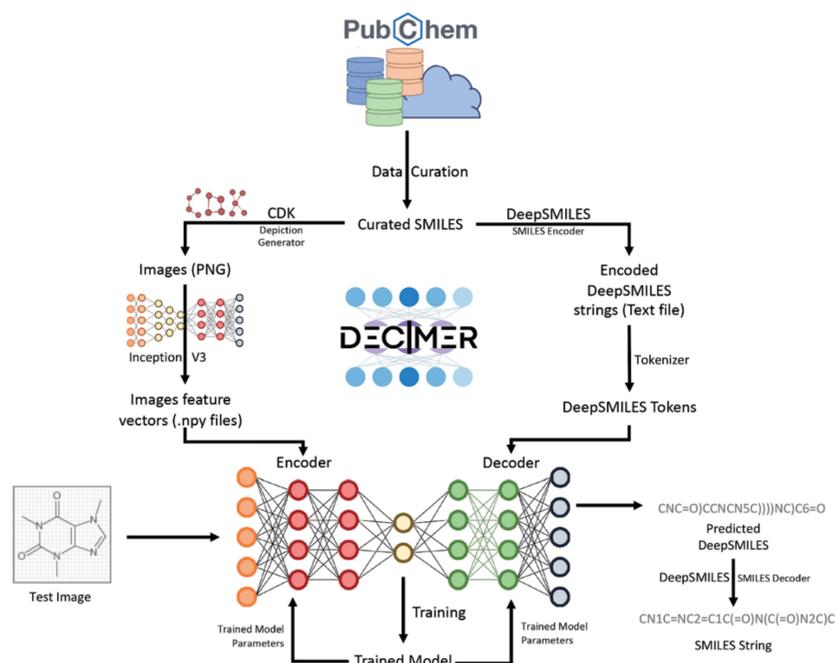


Abb. 2:
Architektur des DECIMER-Lernsystems.

Da die in den Bildern kodierte Formelsprache der Chemie sehr hoch entwickelt ist, ist die Übersetzungsaufgabe „Pixel-Grafik zu SMILES“ äußerst anspruchsvoll (Abbildung 1). Das DECIMER-Lernsystem umfasst dazu modernste Neuronale-Netzwerk-Komponenten wie das Inception-V3-Modell für die primäre Bildverarbeitung und ein erst kürzlich entwickeltes Show-Attend-Tell-Autoencoder-Netzwerk für die anschließende Bildererkennung/Übersetzung (Abbildung 2). Die volle Leistungsfähigkeit des DECIMER-Systems auf Basis großer Datenmengen wird erst mit seiner Cloud-Implementierung erwartet, da die erforderliche Big-Data-Rechenkapazität mit wissenschaftlichen Workstations allein nicht zu erbringen ist [1].

In der Praxis wollen die Projektpartner ihr System zukünftig zur Erkennung chemischer Strukturen von Naturstoffen einsetzen, die in umfangreichen wissenschaftlichen Literaturquellen lediglich als Bild vorliegen [2].

[1] RAJAN, Kohulan, Achim ZIELESNY, Christoph STEINBECK, 2020. DECIMER - Towards Deep Learning for Chemical Image Recognition.

In: Journal of Cheminformatics. 12:65

[2] RAJAN, Kohulan, Henning Otto BRINKHAUS, Achim ZIELESNY, Christoph STEINBECK, 2020. A review of optical chemical structure recognition tools.

In: Journal of Cheminformatics. 12:60.