

Vernetzte Daten für eine Diabetesforschung der Zukunft



Die Menge an Daten zum Thema Diabetes wächst kontinuierlich. Die Volkskrankheit

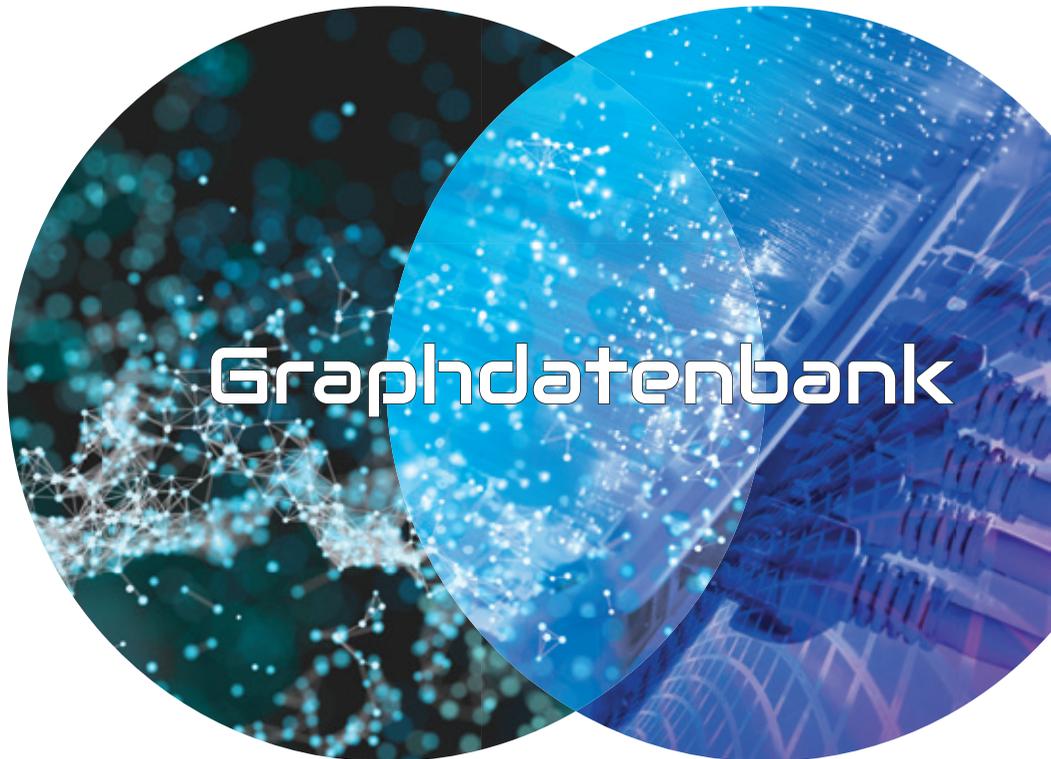
Diabetes lässt sich in einer vollkommen neuen Dimension erforschen - wenn es gelingt, die vielfältigen und bisher meist in separaten Datenbanken gespeicherten

Informationen miteinander zu vernetzen und mit modernen Informations- und Kommunikationstechnologien wie der künstlichen Intelligenz zu analysieren.

***Dr. Alexander Jarasch, Dr. Astrid Glaser,
Prof. Dr. Martin Hrabě de Angelis, München,
Deutsches Zentrum für Diabetesforschung***

Bei Hausärzten, in Kliniken, in Universitäten, Forschungseinrichtungen, Labors, bei den Krankenkassen, in Pharmaunternehmen: Die Menge an medizinischen Daten wächst kontinuierlich und exponentiell. Dabei reicht die Datenvielfalt von Blutwerten, Biomarkern, Befunden, Ultraschallaufnahmen sowie Röntgen- und CT-Bildern bis in das menschliche Genom. Hinzu kommen unzählige Daten aus Studien und Publikationen sowie Patientendaten, die heute nicht mehr nur beim Arzt gesammelt werden, sondern auch rund um die Uhr von medizinischen Geräten, Gesundheits- oder Sport-Apps. Allein in den USA umfasst die Menge der Gesundheitsdaten bereits mehr als 150 Exabyte - das sind 150 Mio. Terabyte [Raghupathi 2014].

In diesen Datenbergen schlummern ungeheure Schätze. Doch „Big Data“ bietet nicht per se einen Mehrwert. Um die verborgenen Schätze bergen zu können, müssen die Daten auch zugänglich, interoperabel und durchsuchbar sein. Erst dann lässt sich durch den Einsatz moderner IT-Lösungen wie Data Mining und künstli-



Graphdatenbank

cher Intelligenz (siehe „Definitionen“) „Big Data“ in „Smart Data“ verwandeln.

In der Diabetesforschung gibt es derzeit vielfältige separate Daten

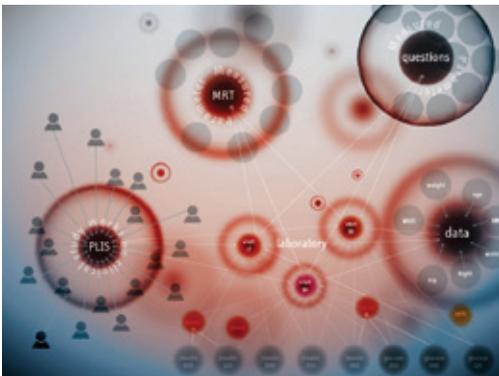
Das gilt insbesondere auch für die Volkskrankheit Diabetes. Diabetes ist eine äußerst komplexe Erkrankung, die durch ein vielschichtiges Zusammenspiel von Genen, Lebensstil und Umweltfaktoren entsteht. Daher erforschen zahlreiche Arbeitsgruppen und Experten verschiedener Disziplinen (u. a. aus Grundlagenforschung, Epidemiologie, Versorgungsforschung und Klinik) die Stoffwechselerkrankung. Jede dieser Gruppen sammelt jeweils wichtige Daten zu einzelnen Aspekten der Erkrankung. In der Diabetesforschung stammen die Daten u. a. aus Kohorten, klinischen Studien, Bioproben, präklinischen Modellen, Geno- und Phäntypisierung, Omics-Analysen sowie aus Untersuchungen zu epigenetischen Veränderungen des genetischen Codes. Die Daten und Ergebnisse der Studien und Untersuchungen sind jedoch sehr heterogen, oft unstrukturiert und wer-

den an verschiedenen Standorten in separaten Datenbanken (Datensilos) gespeichert. Zudem sind die Informationen meist gar nicht oder nur sehr schwer durchsuchbar. Doch um die Entstehung der Krankheit zu verstehen und neue präzise Therapieansätze (Precision Medicine) entwickeln zu können, müssen Informationen aus unterschiedlichen Bereichen miteinander verbunden und ausgewertet werden. Denn mehr als jede einzelne Datenquelle kann die Vernetzung aller verfügbaren Datenbanken das Wissen zur Diabetesentstehung, -prävention und -behandlung erweitern.

Forschungsdaten miteinander vernetzen

Das Deutsche Zentrum für Diabetesforschung (DZD) – mit fünf Partnern und fünf assoziierten Partnern – entschloss sich daher, DZDconnect als standortübergreifendes Daten- und Knowledge-Management aufzubauen. Die Grundlage dafür bildet eine Graphdatenbank. Sie soll die bereits vorhandenen Daten aus unterschiedlichen Quellen im DZD zugänglich, wiederverwertbar und übergreifend nutz-

bar machen – gemäß dem FAIR-Prinzip (Findable, Accessible, Interoperable und Re-usable). Damit will das DZD in Zukunft seine Forschungsdaten aus den heterogenen Quellen zusammenführen, strukturieren und mithilfe innovativer IT-Technologien – wie Artificial Intelligence – analysieren, um so bisher unerkannte Querverbindungen und Muster sichtbar zu machen [Möller 2019, Jarasch 2018]. Im Sinne eines interdisziplinären Forschungsansatzes lassen sich so bislang verborgene Zusammenhänge der Krankheit aufdecken. Warum erkranken manche Menschen an Diabetes, andere nicht? Welche Rolle spielen dabei Gene, Ernährung, Bewegung und Umweltfaktoren? Lässt sich schon in jungen Jahren das Diabetesrisiko bestimmen? Und wie kann man der Erkrankung entsprechend vorbeugen? Die Verknüpfung unterschiedlicher Datensätze kann



Mithilfe der Graphtechnologie können Daten standortübergreifend verbunden und abgefragt werden.

hier grundsätzlich neue Antworten liefern. So könnten z. B. neue Marker für Prädiabetes oder andere Subtypen des Diabetes entdeckt werden. Zudem könnte das System über Natural Language Processing (NLP) in Zukunft Texte einlesen, analysieren und relevante Informationen selbstständig in die Datenbank integrieren,

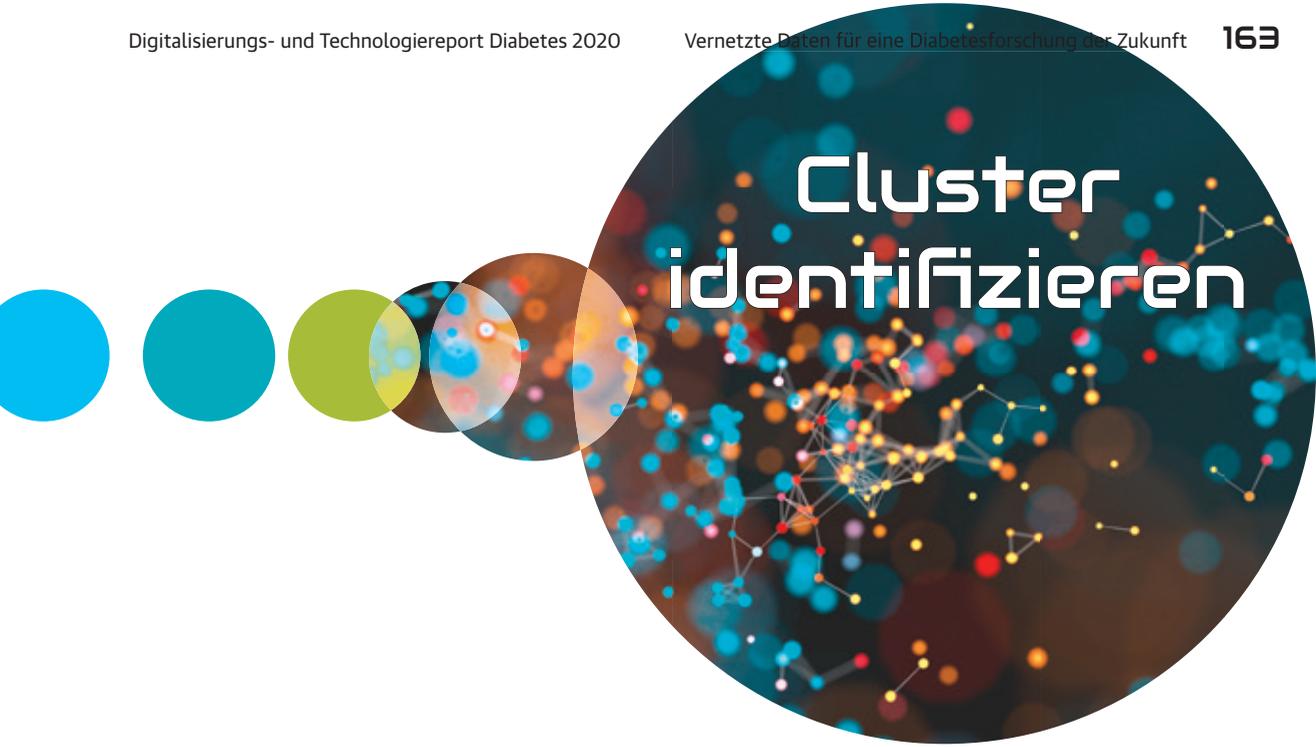
sobald zentrale Begriffe signifikant häufig oder im richtigen semantischen Umfeld auftreten.

Subtypen des Diabetes entdeckt

Diabetes ist eine Erkrankung, die sich sehr heterogen manifestiert. Studien aus Skandinavien [Ahlqvist 2018] zeigten, dass man Diabetiker entsprechend der Schwere des Verlaufs des Diabetes in verschiedene Cluster einteilen kann. Das DZD konnte diese Befunde an 1105 Patienten aus der Deutschen Diabetes-Studie bestätigen. Mithilfe der in Schweden erprobten Analysemethode (Clustering-Algorithmus) suchten die Wissenschaftler in dem umfangreichen Datensatz nach auffälligen Mustern und Gesetzmäßigkeiten und entdeckten fünf Cluster, die eine Aufteilung des Diabetes in Subtypen ermöglichen. Zwei dieser neuen Subtypen – der milde adipositasbedingte Diabetes und der milde altersbedingte Diabetes – haben weniger schwere Krankheitsverläufe. Die drei übrigen Subtypen gehen mit einem hohen Risiko von Diabetesfolgen einher: Patienten mit schwerem insulinresistentem Diabetes haben vermutlich ein höheres Risiko für Erkrankungen der Leber und Nieren, während jene mit einem schweren insulindefizitären Diabetes eher an Netzhautschäden und diabetischer Neuropathie erkranken dürfen. Der dritte Subtyp mit häufigen Komplikationen ist der schwere autoimmune



Stratifizierung in Diabetes-Subtypen



Cluster identifizieren

Diabetes, der dem klassischen Typ-1-Diabetes entspricht [Zaharia 2019].

Unterschiedliche Arten des Prädiabetes

Unterschiedliche Subtypen gibt es bereits bei der Vorstufe des Diabetes (Prädiabetes). Das zeigen Auswertungen der DZD-Multicenterstudie „Prädiabetes-Lebensstil-Interventions-Studie“ (PLIS). Nicht jeder Prädiabetiker hat das gleich hohe Risiko, später auch einen Diabetes zu entwickeln. Es gibt vielmehr eine Hochrisikogruppe. Bei Probanden, die an einer Fettleber mit Insulinresistenz oder einer Insulinsekretionsstörung leiden, kommt es mit einer sehr hohen Wahrscheinlichkeit zu einer manifesten Diabeteserkrankung. Zudem ist das Risiko erhöht, später auch Folgeerkrankungen zu entwickeln. Untersuchungen deuten darauf hin, dass eine intensive Lebensstilintervention mit viel Bewegung und einer nachhaltig begleitenden Beratung hier helfen kann, den Ausbruch der Stoffwechselerkrankung hinauszuzögern oder gar zu verhindern [Fritsche 2019].

Digitales Diabetes Präventions-Zentrum (DDPC)

Wir müssen aber davon ausgehen, dass noch mehr als diese schon beschriebenen Subtypen des Dia-

betes bzw. Prädiabetes existieren. Daher gilt es, weitere Cluster - respektive Patientengruppen - zu identifizieren und herauszufinden, welche davon jeweils welche Eigenschaften aufweisen und wie die verschiedenen Subtypen am besten erkannt und behandelt werden könnten. Doch um diese Untergruppen entdecken zu können, bedarf es sehr viel größerer Datenmengen. Daher wollen wir ein Digitales Diabetes Präventions-Zentrum (DDPC) aufbauen. Der Datenpool für das DDPC soll sich nicht nur aus Forschungsdaten des DZD speisen, sondern auch aus weiteren vorhandenen medizinischen Daten sowie Gesundheitsdaten. Informationen, die bisher bei nationalen Forschungseinrichtungen sowie Krankenkassen, Krankenhäusern und niedergelassenen Ärzten vorliegen, können wertvolle Hinweise liefern. Zusätzlich werden Daten gesunder Personen aus der allgemeinen Bevölkerung gebraucht, um die Entstehung von Typ-2-Diabetes besser verstehen und die dringend benötigten frühen Prädiktoren für unterschiedliche Subtypen aufspüren zu können.

Mögliche Quellen für diese Informationen könnten die heute schon weitverbreiteten Tracking-Armbänder, smarten Uhren, Wearables oder Fitness- und Gesundheits-Apps sein. Bis-

lang greifen vor allem große internationale IT- und Pharmaunternehmen auf selbst erhobene Gesundheitsdaten von Privatpersonen zu. Um diesen Datenschatz nicht vorrangig der Industrie zur Nutzung zu überlassen, ist es wichtig, dass auch die Forschung Zugang zu diesen Daten bekommt. Dem Wunsch, diese vielfältigen Datenquellen nutzen zu können, stehen derzeit strenge Datenschutzbestimmungen entgegen. Gemeinsam mit der Politik, den Fachgesellschaften und Patientenverbänden müssen wir deshalb nach Lösungen suchen, die den Missbrauch digitaler Daten ausschließen und zugleich deren Nutzung für die medizinische Forschung nicht behindern. Der von der Deutschen Diabetes Gesellschaft (DDG) entwickelte „Code of Conduct Digital Health“ bietet hierzu eine gute Grundlage [Deutsche Diabetes Gesellschaft 2018].

Deutliche Mehrheit der Deutschen bereit zur Datenspende

Die Bereitschaft der Menschen, ihre Gesundheitsdaten in Form einer Datenspende der Wissenschaft zur Verfügung zu stellen, ist sehr groß. Fast vier von fünf Deutschen (79%) würden ihre Gesundheitsdaten anonym und unentgeltlich digital für die medizinische Forschung zur Verfügung stellen. Das zeigt eine Forsa-Umfrage im Auftrag der Technologie- und Methodenplattform für die vernetzte medizinische Forschung [TMF 2019]. Ein ähnliches Ergebnis ergab sich auch in der in diesem D.U.T. veröffentlichten Untersuchung, in der 69,7% aller befragten Patienten ihre Daten anerkannten Institutionen zu wissenschaftlichen Zwecken überlassen würden.

DDPC: Politik überzeugen

Das DDPC ist als nationale Initiative mit verschiedenen Standorten in Deutschland angedacht. Die mathematische und informationstechni-

sche Fachkompetenz bringt das Helmholtz Zentrum München ein, die klinische Fachkompetenz kommt vom DZD und weiteren nationalen Partnern an Kliniken und Forschungseinrichtungen. Mit neuen IT-Technologien aus dem Bereich Artificial Intelligence ist es möglich, die oben beschriebenen unterschiedlichen Datenquellen miteinander zu verbinden und auszuwerten. Wichtig dabei ist die enge Zusammenarbeit von Datenexperten mit all jenen Experten aus Forschung und Klinik. Im DDPC sollen die Diabetes-Data-Institute von Klinischen Zentren des DZD flankiert werden, die auch „Proof of principle“-Studien durchführen. Aus dem so generierten Wissen lassen sich präzise Präventionsstrategien ableiten und individuelle Empfehlungen für (noch) gesunde Bürgerinnen und Bürger abgeben. Noch ist das DDPC eine Zukunftsvision. Es gilt, die Politik zu überzeugen, in ein solches Zentrum zu investieren.

Künstliche Intelligenz in der Diabetesforschung

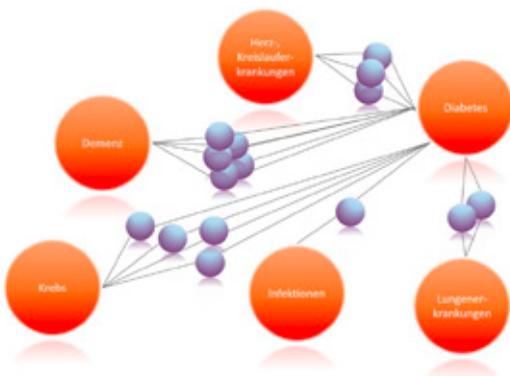
Weltweit nutzen Diabetesforscher die neuen Möglichkeiten moderner IT-Technologien. An welchen unterschiedlichen Fragestellungen die Wissenschaftler arbeiten, zeigen ein paar Beispiele: Wissenschaftlern aus München und Dresden ist es gelungen, u. a. mithilfe von Deep-Learning-Methoden, den Risk Score für Typ-1-Diabetes deutlich zu verbessern. Der neue Risikotest bezieht bis zu 41 Typ-1-Diabetes-Risikogen-Regionen ein [Bonifacio 2018]. IBM forscht an einem KI-gestützten Screening-Tool, um Typ-1-Diabetes frühzeitig diagnostizieren zu können [Hu 2019].

Derzeit arbeitet das Pharmaunternehmen Roche mit IBM Watson Health an einem neuen datenbasierten Vorhersagemodell (mit Real World Data) für diabetesbedingte chronische Nierenerkrankungen [Ravizza 2019]. In einem Pilotprojekt aus

München gelang es Computational-Biology-Experten, mit Techniken von Deep-Learning Retinopathien – eine häufige Folgeerscheinung des Diabetes – besser zu klassifizieren. Sie trainierten den Computer darauf, kranke und gesunde Augen zu erkennen und die Patienten entlang der Schwere und des Fortschreitens der Erkrankung einzuordnen [Eulenberg 2017]. Diese Strategie wird nun weiterverfolgt, um neue Therapien und individuelle Vorsorgekonzepte zu entwickeln. Auch andere Gruppen arbeiten an Systemen zum Netzhautscreening, um diabetische Veränderungen an der Netzhaut zu erkennen. Bereits 2018 hat die amerikanische Gesundheitsbehörde (Food and Drug Administration, FDA) ein erstes System zur Anwendung künstlicher Intelligenz zur Retinopathiediagnostik zugelassen [U.S. Food and Drug Administration 2018]. Als erste Diabetesklinik in Deutschland setzt das Diabetes Zentrum Mergentheim seit Ende 2018 künstliche Intelligenz für Untersuchungen auf diabetische Retinopathie ein.

Verbindungen zwischen verschiedenen Volkskrankheiten erkennen

Diabetes ist mit vielen anderen großen Volkskrankheiten wie Demenz, Krebs, Lungen-, Infektions-, Herz- und Kreislauferkrankungen assoziiert. Um



Verbindungen in Daten zu unterschiedlichen Volkskrankheiten entdecken

© DZD

Zusammenhänge zwischen diesen Erkrankungen und mögliche gemeinsame frühe Prädiktoren zu erkennen, arbeiten die sechs Deutschen Zentren der Gesundheitsforschung (DZG) zusammen. Hier gilt es, Synergien zu nutzen und datenschutzkonforme Datenverknüpfungen aufzubauen.

Nationale Forschungsdateninfrastruktur

Um Daten sicher und datenschutzkonform forschungseinrichtungsübergreifend nutzen zu können, bedarf es entsprechender IT-Infrastrukturen. Ende 2018 haben Bund und Länder den Aufbau einer Nationalen Forschungsdateninfrastruktur (NFDI) beschlossen. In der NFDI sollen Datenbestände in einem aus der Wissenschaft getriebenen Prozess systematisch erschlossen, langfristig gesichert und entlang der FAIR-Prinzipien über Disziplinen- und Ländergrenzen hinaus zugänglich gemacht werden. Um den Aufbau dieser Infrastruktur konnten sich bis zum Herbst 2019 NFDI-Konsortien aller Fachrichtungen bewerben. Die DZG haben gemeinsam mit den vier Konsortien der Medizininformatik-Initiative (MII) unter der Federführung des DZD und der Technologie- und Methodenplattform für die vernetzte medizinische Forschung (TMF) einen Antrag für den Aufbau einer interoperablen, nutzerfreundlichen und nachhaltigen Infrastruktur für die Integration und den Austausch von Daten aus verschiedenen Bereichen der biomedizinischen Forschung und klinischen Versorgung gestellt. Ziel dieses neuen „NFDI for Future Medicine Konsortium (NFDI4MED)“ ist es, ein qualitätsorientiertes Datenmanagement als integralen Bestandteil des Forschungsdatenzyklus im Bereich der patientenorientierten biomedizinischen Forschung zu ermöglichen. Die Struktur des Forschungsdatenmanagements von NFDI4MED soll eine dezentrale nationale Infrastruktur umfassen, die vorhandene Datenbanken und digitale Ressourcen aus den verschiedenen For-

schungsdisziplinen und -organisationen miteinander vernetzt. Im NFDI4Medicine-Konsortium sind dies mehr als 100 Universitäts- und Universitätsklinikstandorte, Forschungseinrichtungen und -organisationen sowie Bundesinstitute. Durch Maßnahmen wie diese wird ein wichtiger Grundstein gelegt, um die Menge an Daten, die in der biomedizinischen Forschung und klinischen Versorgung entstehen, für die Gesundheit der Menschen und neue Präventions- und Therapiemaßnahmen sinnvoll einzusetzen.

Quellen:

- Ahlqvist E, Storm P, Käräjämäki A, Martinell M, Dorkhan M, Carlsson A, Vikman P, Prasad RB, Aly DM, Almgren P, Wessman Y, Shaat N, Spégel P, Mulder H, Lindholm E, Melander O, Hansson O, Malmqvist U, Lernmark Å, Lahti K, Forsén T, Tuomi T, Rosengren AH, Groop L: Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. *Lancet Diabetes Endocrinol* 2018; 6: 361–369
- Bonifacio E, Beyerlein A, Hippich M, Winkler C, Vehik K, Weedon MN, Laimighofer M, Hattersley AT, Krumsiek J, Frohnert B, Steck AK, Hagopian WA, Krischer JP, Lernmark Å, Rewers MJ, She J, Toppari, Akolkar, Oram RA, Rich S, Ziegler A; TEDDY Study Group: Genetic scores to stratify risk of developing multiple islet autoantibodies and type 1 diabetes: a prospective study in children. *PLoS Med* 2018; 15: e1002548
- Deutsche Diabetes Gesellschaft: Rahmenpapier für einen Code of Conduct Digital Health der Deutschen Diabetes Gesellschaft (DDG) zur digitalen Transformation. https://www.deutsche-diabetes-gesellschaft.de/fileadmin/Redakteur/Ueber_uns/Code_of_Conduct_der_DDG_Digital_Health_19092017.pdf. 21. August 2018 (Zugriff: 23.10.2019)
- Eulenberg P, Köhler N, Blasi T, Filby A, Carpenter AE, Rees P, Theis FJ, Wolf FA: Reconstructing cell cycle and disease progression using deep learning. *Nat Commun* 2017; 8: 463
- Fritsche A: Diagnose von Risikogruppen im Prädiabetes. *Diabetes Kongress*, 31.05.2019
- Hu J: AI offers hope for earlier screening for type 1 diabetes. *IBM Research Blog*, June 10, 2019. <https://www.ibm.com/blogs/research/2019/06/ai-diabetes/> (Zugriff: 24.10.2019)
- Jarasch A, Glaser A, Häring H, Roden M, Schürmann A, Solimena M, Theiss F, Tschöp M, Wess G, Hrabec de Angelis M: Mit Big Data zur personalisierten Diabetesprävention. *Diabetologie* 2018; 14: 486–492
- Möller D, Ottleben I: Vernetzt gegen Diabetes: Wie Big Data harmonisieren? *Laborpraxis*, 21.01.2019. <https://www.laborpraxis.vogel.de/vernetzt-gegen-diabetes-wie-big-data-harmonisieren-a-784562/> (Zugriff: 25.10.2019)
- Raghupathi W, Raghupathi V: Big data analytics in healthcare: promise and potential. *Health Inf Sci Syst* 2014; 2: 3
- Ravizza S, Huschto T, Adamov A, Böhm L, Büsser A, Flöther FF, Hinzmann R, König H, McAhren SM, Robertson DH, Schleyer T, Schneidinger B, Petrich W: Predicting the early risk of chronic kidney disease in patients with diabetes using real-world data. *Nat Med* 2019; 25: 57–59
- TMF – Technologie- und Methodenplattform für die vernetzte medizinische Forschung: Deutliche Mehrheit der Deutschen bereit zur Datenspende für die medizinische Forschung. *Presseinformation* 27.08.2019. <http://www.tmf-ev.de/News/articleType/ArticleView/articleId/4456.aspx> (Zugriff: 29.11.2019)
- U.S. Food and Drug Administration: FDA News Release: FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems. April 11, 2018. <https://www.fda.gov/news-events/press-announcements/fda-permits-marketing-artificial-intelligence-based-device-detect-certain-diabetes-related-eye> (Zugriff: 23.10.2019)
- Zaharia OP, Strassburger K, Strom A, Bönhof GJ, Karusheva Y, Antoniou S, Bódís K, Markgraf DF, Burkart V, Müssig K, Hwang JH, Asplund O, Groop L, Ahlqvist E, Seissler J, Nawroth P, Kopf S, Schmid SM, Stumvoll M, Pfeiffer AFH, Kabisch S, Tselmin S, Häring HU, Ziegler D, Kuss O, Szendroedi J, Roden M: Risk of diabetes-associated diseases in subgroups of patients with recent-onset diabetes: a 5-year follow-up study. *Lancet Diabetes Endocrinol* 2019; 7: 684–694

Definitionen

Data Mining

Data Mining ist die systematische Anwendung computergestützter Methoden, um in vorhandenen Datenbeständen Muster, Trends oder Zusammenhänge zu finden.

Künstliche Intelligenz

Künstliche Intelligenz (KI), auch artifizielle Intelligenz (AI) genannt, ermöglicht Maschinen, u. a. aus Erfahrung zu lernen, sich auf neu eingehende Informationen einzustellen und Aufgaben zu bewältigen, die „menschenähnliches“ Denkvermögen erfordern. In den meisten Fällen ist mit „KI“ eine „schwache KI“ gemeint, die konkrete Anwendungsprobleme zu lösen versucht. Dazu werden Methoden wie das Maschinelle Lernen (Machine Learning) und das Verarbeiten natürlicher Sprache (NLP: Natural Language Processing) genutzt.

Machine Learning

Maschinelles Lernen beschreibt den Erwerb von Wissen durch ein künstliches System. Ähnlich wie ein Mensch generiert der Computer Wissen aus Erfahrung und kann eigenständig Lösungen für neue und unbekannte Probleme finden. Dazu wird das System zunächst mit Beispieldaten trainiert, in denen es Muster und Gesetzmäßigkeiten erkennt. Die aus den Daten gewonnenen Erkenntnisse lassen sich verallgemeinern und für neue Problemlösungen oder für die Analyse bisher unbekannter Daten verwenden.

Deep Learning

Deep Learning (frei übersetzt: tiefgehendes Lernen) ist ein Teilbereich des maschinellen Lernens (Machine Learning). Beim Deep Learning werden Lernalgorithmen sowie ständig wachsende Datenmengen in große künstliche neuronale Netzwerke eingespeist. Die Lernmethoden orientieren sich an der Arbeitsweise des menschlichen Gehirns, das ebenfalls aus miteinander verschalteten Neuronen besteht. Die Fähigkeit der Netzwerke, zu „denken“ und zu „lernen“, verbessert sich kontinuierlich, indem immer mehr Daten verarbeitet werden.

Natural Language Processing (NLP)

Natural Language Processing versucht, natürliche Sprache bzw. Texte zu erfassen und mithilfe von Regeln und Algorithmen computerbasiert zu verarbeiten. Dafür verwendet NLP verschiedene Methoden und Ergebnisse aus den Sprachwissenschaften und kombiniert sie mit moderner Informatik und künstlicher Intelligenz. Ziel ist es, gesprochene bzw. geschriebene Sprache zu erkennen, zu analysieren und den Sinn zur weiteren Verarbeitung zu extrahieren. Dafür muss das System nicht nur einzelne Wörter und Sätze verstehen, sondern auch Textzusammenhänge erfassen.
