

Visuelle Analyse medizinischer Daten

Bernhard Preim · Steffen Oeltze
Klaus Tönnies

Einleitung

In der Medizin fällt eine enorme Vielfalt an Daten an, die sich auf Diagnosen und die Behandlung von Erkrankungen, Therapien und Nachsorge bezieht. Zu diesen Daten zählen einfache Zahlen, z. B. Ergebnisse von Laboruntersuchungen, daraus abgeleitete Daten, wie z. B. die standardisierte Einschätzung des Schweregrades einer Erkrankung, qualitative Informationen und komplexe medizinische Bilddaten, wie z. B. Computertomografie-Datensätze (CT-Datensätze). Derartige Bilddaten sind standardmäßig mit einer Vielzahl von Informationen verknüpft, die den Patienten betreffen, bzw. mit Informationen, die die konkreten Einstellungen des bildgebenden Geräts und Merkmale der Daten, wie die räumliche und zeitliche Auflösung, betreffen. Hinzu kommen oft Ergebnisse einer feingeweblichen Diagnostik, bei der Gewebeproben mikroskopisch untersucht werden, um auf zellulärer Ebene Krankheiten zu analysieren und auf diese Weise z. B. einen möglichen Krebsverdacht zu klären. Eine Vielzahl dieser Daten liegt mittlerweile digital vor, und die in der Medizin verbreiteten Informationssysteme (Krankenhaus-, Labor-, Radiologieinformationssysteme) ermöglichen eine relativ starke Integration der Daten. Besonders weit fortgeschritten ist die Digitalisierung in der Radiologie. Dagegen steht die digitale Befundung mikroskopischer Bilddaten in der Pathologie noch am Anfang. Dies ist darauf zurückzuführen, dass die bei der Digitalisierung entstehenden Datenmengen extrem groß sind (mehrere Gigabyte pro Bild) und dass Standards für die Betrachtung erst im Entstehen sind.

Zeitlicher und räumlicher Bezug. Ein wesentliches Merkmal der Patientendaten ist ihre Zeitabhängigkeit. Diagnostische und therapeutische Entscheidungen basieren oft auf einer Abschätzung, wie schnell sich bestimmte Parameter ändern. So wäre der Krebsverdacht bei Beurteilung eines Tumorherdes wesentlich geringer, wenn sich der Herd im Vergleich zu Aufnahmen aus dem Vorjahr praktisch nicht verändert hat, als wenn ein schnelles Wachstum beobachtet wird. Ein räumlicher Bezug ist bei den individuellen Patientendaten meist nicht von Bedeutung, spielt aber in der Epidemiologie bzw. im öffentlichen Gesundheitswesen eine wichtige Rolle.

Bildanalyse und visuelle Exploration. In den meisten Anwendungsszenarien wird die statistische Datenanalyse mit der visuellen Exploration kombiniert, um Daten auszuwerten. Bei der Auswertung medizinischer Bilddaten, wie z. B. Kernspintomografie (auch Magnetresonanztomografie bzw. MRT) und CT, kommen häufig die Bildverarbeitung und -analyse hinzu. Die Qualität der Daten ist meist nicht optimal, weil Kompromisse zwischen Untersuchungszeit, Strahlenbelastung und Bildqualität gefunden werden müssen. Die Bilddaten werden durch spezielle Filter so verbessert, dass das aufnahmebedingte Rauschen reduziert wird. Relevante Strukturen werden identifiziert und abgezogen

DOI 10.1007/s00287-010-0491-x
© Springer-Verlag 2010

Bernhard Preim · Steffen Oeltze · Klaus Tönnies
Otto-von-Guericke-Universität Magdeburg, FIN/ISG,
Magdeburg
E-Mail: Bernhard.preim@ovgu.de

Zusammenfassung

In der Medizin werden große Mengen an Daten generiert, die sich auf diagnostische Prozeduren, Behandlungsentscheidungen und Ergebnisse der Behandlung beziehen. Medizinische Bilddaten, z. B. Computertomografie (CT) und Kernspintomografiedaten (MRT), werden häufig akquiriert. Diese Daten müssen effizient analysiert werden, um klinische Entscheidungen adäquat zu unterstützen. Insbesondere müssen Bildanalysetechniken, wie die Segmentierung und Quantifizierung anatomischer Strukturen und die visuelle Exploration der Daten, integriert werden. Neben den Anforderungen der individuellen Behandlung ergeben sich weitere Herausforderungen für die Datenauswertung aus den Bedürfnissen der klinischen Forschung, der öffentlichen Gesundheitsvorsorge und der Epidemiologie. Die Rolle des Benutzers ist hier die eines Forschers, der Daten untersucht und dabei z. B. potenzielle Korrelationen zwischen Risikofaktoren und der Entstehung von Erkrankungen analysiert. Die visuelle Exploration, bei der oft mehrere koordinierte Ansichten genutzt werden, und statistische Analysen müssen dazu geeignet integriert werden. Oft sind dabei die räumliche (geografische) Verteilung der Patienten und die zeitliche Entwicklung von Erkrankungsfällen wesentlich. Daher müssen die medizinischen Daten in ihrem räumlichen und zeitlichen Bezug repräsentiert werden, sodass eine enge Verbindung zwischen geografischen Informationssystemen und der Datenvisualisierung entsteht.

(Segmentierung) und Datensätze, die miteinander verglichen werden sollen, werden so transformiert, dass die Bildinhalte optimal korrespondieren (Registrierung).

Workflowunterstützung. Die im praktischen Einsatz befindlichen Systeme zeichnen sich durch eine nahtlose Integration und eine starke Unterstützung klinischer Workflows aus. Dabei kommen verstärkt Konzepte und Systeme aus dem Bereich der betrieblichen Informationssysteme zum Einsatz, wie beispielsweise SAP-Anpassungen. In diesem Artikel geht es allerdings um die Methoden der Auswer-

tung und nicht um deren Integration in umfassende klinische Prozesse.

Visuelle Analyse in der Epidemiologie und in der öffentlichen Gesundheitsvorsorge

Über die patientenindividuellen Daten hinaus werden in der Epidemiologie und der medizinischen Forschung die Daten großer Patientengruppen im Zusammenhang analysiert, um z. B. Risikofaktoren für eine Erkrankung zu entdecken oder vorherzusagen, wie sich eine Krankheit ausbreiten wird. In diesen Bereichen ist die Datenmenge nicht nur wesentlich größer, sondern es ergeben sich grundsätzlich neue Anforderungen an die Auswertung, insbesondere in Bezug auf statistische Analysen. So werden beispielsweise Krebserkrankungen und ihre Behandlung umfassend dokumentiert und diese Dokumentationen werden in Krebsregistern verwaltet. Für viele Fragestellungen ist es wichtig, Daten nicht nur zeitlich, sondern auch räumlich referenziert zu speichern. Bekannte Beispiele solcher Fragestellungen mit klarem Raumbezug sind die Fragen nach einer erhöhten Krebsgefahr in der Nähe von Kernkraftwerken oder Mobilfunkmasten. In der klinischen Forschung werden die Wirksamkeit bzw. das Nebenwirkungsspektrum neuer Therapien (Medikamente, Operationsverfahren, physiotherapeutische Behandlungen) in großen Studien untersucht. Da es dabei auch um die Erfassung relativ seltener Nebenwirkungen geht, entstehen in diesem Zusammenhang oft große Datensätze mit mehreren Tausend Patienten, meist aus verschiedenen Krankenhäusern. In den frühen Phasen der klinischen Forschung, speziell in der pharmazeutischen Forschung, sucht man nach Wirkstoffkandidaten, also nach Substanzen, die eine bestimmte Zielstruktur, z. B. Tumorzellen oder entzündliche Ablagerungen in Blutgefäßen, gezielt attackieren [12]. In dieser Phase geht es nicht darum, Daten in großen Mengen möglichst standardisiert auszuwerten, sondern um die Generierung von Hypothesen. Die epidemiologische und klinische Forschung beeinflusst die Patientenversorgung in starkem Maße. Empfehlungen und Leitlinien, wie Patienten zu untersuchen und zu behandeln sind, basieren auf derartigen Forschungsergebnissen.

In der Krankenversorgung in Arztpraxen, Krankenhäusern und Rehabilitationseinrichtungen spielen Methoden der visuellen Analyse bisher

Abstract

In medicine, large amounts of data are generated relating to diagnostic procedures, treatment decisions and treatment outcome. Medical image data, as from CT and MRI, are frequently used to support the treatment process. These data need to be efficiently analyzed to support clinical decisions. In particular image analysis, such as segmentation and quantification of anatomic structures, and visual exploration are integrated to support this process. Over and above, individual treatment, clinical research, public health and epidemiology pose significant challenges for an efficient analysis of massive amounts of data. The user's role is that of a researcher who investigates data to identify, e. g. potential correlations between risk factors and disease development. Visual exploration, often using multiple coordinated views, and statistical analysis needs to be carefully integrated to support such tasks. Often, the spatial distribution of patients and the development of diseases over time are essential. Thus, medical data has to be represented along with spatio-temporal information leading to a close connection between geographic information systems and data visualization.

praktisch keine Rolle. Dies hängt zum einen damit zusammen, dass der Bedarf nach umfassender (statistischer) Auswertung relativ gering ist, zum anderen damit, dass die Krankenversorgung unter hohem Zeitdruck steht und Varianten der Datenauswertung, die den zeitlichen Bedarf in der Versorgung deutlich erhöhen würden, kaum anwendbar sind. Zu den wenigen Ausnahmen könnten Darstellungen in der Intensivmedizin gehören, bei denen Anästhesisten anhand einer Vielzahl von Parametern den Zustand des Patienten überwachen und kritische Zustände geeignet hervorgehoben werden müssen [1]. Ganz anders ist die Situation in der klinischen und epidemiologischen Forschung. Einerseits zwingt die Datenmenge ohnehin dazu, ausgefeilte Methoden der statistischen Analyse, einschließlich Hypothesengenerierung und Exploration, einzusetzen. Andererseits besteht das Hauptinteresse darin, neue Erkenntnisse zu formulieren und zu verifizieren, sodass auch ein hoher Aufwand gerechtfertigt ist.

Visuelle Analyse in der Epidemiologie

In der Epidemiologie wird untersucht, wovon die Häufigkeit von Erkrankungen (*Prävalenz*) und die Sterblichkeit (*Mortalität*) abhängen. Regionale Unterschiede, Abhängigkeiten vom Lebensstil (Gewicht, Rauchen, Schwangerschaften, ...), sozio-ökonomische Faktoren (Einkommen, Ausbildung, Daten zur gesundheitlichen Versorgung) und Zusammenhänge zwischen verschiedenen Erkrankungen sind dabei von Interesse. Weit verbreitet sind einfache visuelle Analysewerkzeuge, wie Scatterplots und Choroplethen-Karten. Scatterplots stellen Abhängigkeiten eines (möglichen) Risikofaktors von Prävalenz oder Mortalität dar und werden oft mit der Darstellung einer Regressionslinie (bestmögliche lineare Anpassung) verbunden (Abb. 1). Zusätzlich wird der Korrelationskoeffizient berechnet. Choroplethen-Karten sind thematische (Land-) Karten, bei denen die Darstellung der Regionen durch Farben bzw. Muster an einen Wert, z. B. die Mortalitätsrate, angepasst ist. Um mit Choroplethen-Karten auch Zusammenhänge zwischen einem möglichen Risikofaktor und einer epidemiologischen Variable (Prävalenz, Mortalität) darstellen zu können, werden die Werte kategorisiert, z. B. indem man jeweils niedrige, normale und hohe Werte unterscheidet, woraus sich neun Kombinationen ergeben – bei geeigneter Wahl von neun Farben ist eine gute Unterscheidbarkeit gegeben [2].

Scatterplot-Matrizen und parallele Koordinaten.

Für die komplexen Fragen der Epidemiologie sind einfache Visualisierungs- und Analysetechniken aber unzureichend. Oft ist das komplexe Zusammenspiel mehrerer Faktoren zielführend. Scatterplot-Matrizen, die paarweise Abhängigkeiten zwischen mehreren Faktoren darstellen können und parallele Koordinaten, die zeitliche Veränderungen mehrerer Faktoren, z. B. Veränderungen in der medizinischen Versorgung, darstellen können, sind hilfreich [2]. Eine Scatterplot-Matrix ist eine quadratische Matrix, wobei die Elemente (i, j) der Matrix Scatterplots enthalten, die die Abhängigkeit des Parameters i vom Parameter j darstellen. Die Elemente (i, i) , die auf der Hauptdiagonale liegen, werden genutzt, um die Häufigkeitsverteilung des Parameters i darzustellen (Abb. 1). Parallele Koordinaten sind eine Technik der mehrdimensionalen Visualisierung, bei der die Dimensionen auf parallele vertikale Achsen mit gleichem Abstand

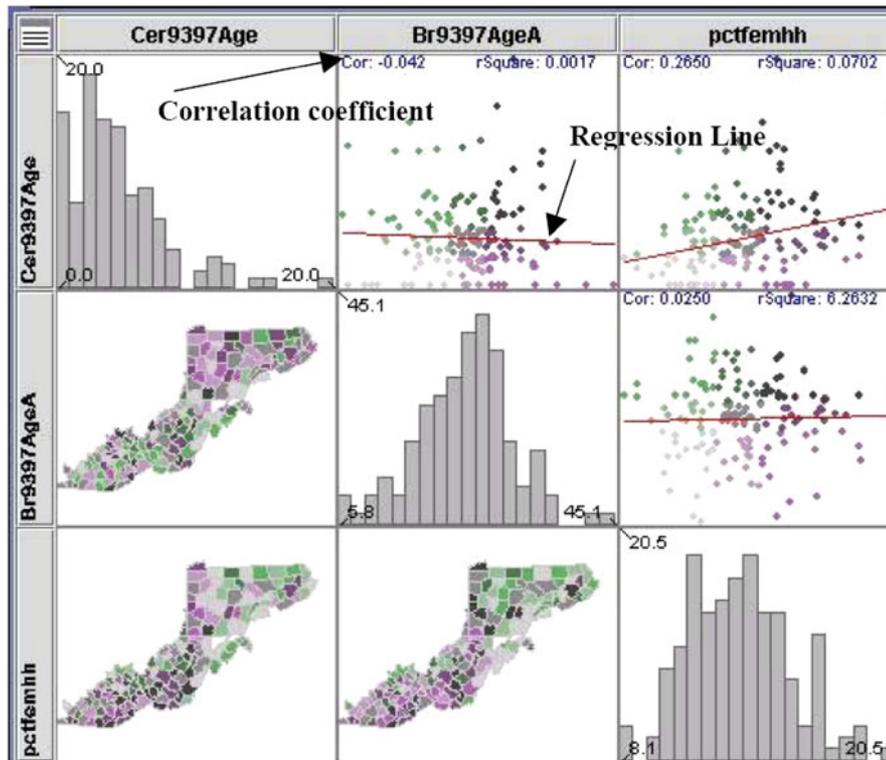


Abb. 1 Eine Scatterplot-Matrix dient dazu, die Häufigkeit von Brust- und Zervixkarzionerkrankungen in Abhängigkeit vom Einkommen (und sozioökonomischen Status) zu untersuchen. In der Hauptdiagonale ist die Häufigkeitsverteilung des Parameters (Häufigkeit von Zervix- und Brustkarzinomen bzw. Einkommen in unterschiedlichen Regionen) dargestellt. Die überlagerte Regressionslinie und der Korrelationskoeffizient vermitteln Informationen für die Aussagekraft. In den Feldern unterhalb der Hauptdiagonale wird die regionale Verteilung der jeweiligen Korrelation in Choroplethen-Karten mit bivariaten Farbskalen dargestellt

abgebildet werden und die einzelnen Datensätze auf Streckenzüge abgebildet werden, wobei jeder Streckenzug alle Achsen entsprechend der Attributausprägungen für die einzelnen Dimensionen schneidet (Abb. 1). Formal wird also ein Punkt P mit den Koordinaten (p_1, p_2, \dots, p_n) des n -dimensionalen Merkmalsraums auf einen Streckenzug abgebildet, der die Punkte P_1, P_2, \dots, P_n auf den Achsen Y_1, Y_2, \dots, Y_n verbindet. Die grundlegende Technik stößt bei wachsenden Datenmengen schnell an Grenzen. Viele Speziallösungen sind entwickelt worden, um die Skalierbarkeit zu verbessern. Vor allem Farbe und Transparenz sind dazu modifiziert worden, um Streckenzüge besser unterscheidbar zu machen und die interaktive Auswahl von Teilbereichen ermöglicht eine Hervorhebung relevanter Teilmengen (Abb. 2).

Die meisten dieser Visualisierungstechniken sind darauf beschränkt, skalare Werte, wie Einkommen und Arztdichte, darzustellen. Teilweise

sind aber auch kategorisierte Daten wesentlich, also Daten, die nominal bzw. ordinal sind. Ähnlich wie in anderen Anwendungsgebieten reicht die visuelle Darstellung für eine tiefgründige Analyse nicht aus. Interaktionstechniken, mit denen entweder im Attributraum (Brushing, Abb. 1) oder im geografischen Kontext interessierende Teilmengen selektiert werden können, kommen häufig zum Einsatz [2].

Bei Langzeitstudien in der Community Medicine (zum Beispiel die SHIP-Studie an der Universität Greifswald, www.medizin.uni-greifswald/cm/fv/ship.html) ergeben sich gänzlich neue Fragestellungen an die Bildanalyse. Hier werden sehr unterschiedliche (Bild-)Daten von Bevölkerungsgruppen über mehrere Jahrzehnte erhoben und unter strengen Bedingungen in Hinblick auf vielfältige Fragestellungen ausgewertet. Geeignete Verfahren müssen die visuelle Analyse auch auf die Analysewerkzeuge selbst beziehen können.

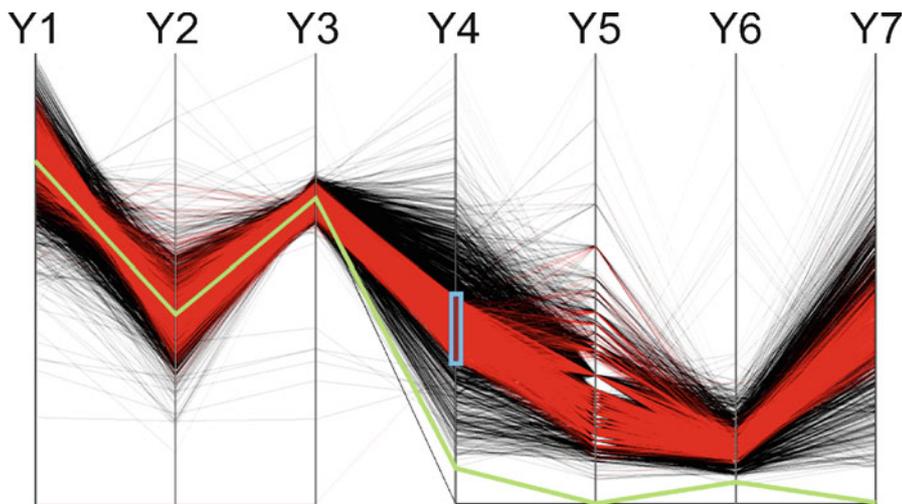


Abb. 2 Parallele Koordinatendarstellung, wobei die Datensätze jeweils sieben Attributausprägungen aufweisen. Der Benutzer interessiert sich besonders für Datensätze, deren Werte für Attribut Y4 in einem bestimmten Intervall liegen (vertikales Rechteck). Diese Datensätze sind rot hervorgehoben, sodass Korrelationen zu Ausprägungen anderer Attribute deutlich werden. Datensätze, die das Kriterium nicht erfüllen, werden abgeschwächt dargestellt, wobei das Ausmaß der Abschwächung mit dem Abstand zum Intervall korreliert. Die Auswahl eines interessierenden Bereichs kann auch genutzt werden, um diese Datensätze in anderen Sichten darzustellen und wird dann als Brushing bezeichnet

Die Adaption an neue Analyseprobleme, eine eventuelle notwendige Kalibrierung bei Wechsel des Bildgebungssystem oder die Integration von Nutzererfahrung sollte durch Feedback zwischen Verhalten eines Analysewerkzeugs und der Analyseanfrage durch geeignete Visualisierungs- und Analyseverfahren unterstützt werden. Analysewerkzeuge selbst sollten nach wenigen, aber methodisch gehaltenen Strategien entwickelt werden, um Synergien aus der Kombination einer visuellen Analyse mit automatischen Methoden zu erzeugen.

Visuelle Analyse in der öffentlichen Gesundheitsvorsorge

Die Gesundheitsämter sammeln Daten, mit denen sie die Ausbreitung von Infektionskrankheiten bewerten und Vorhersagen für die weitere Ausbreitung treffen können. Auch die Abschätzung von möglichen Auswirkungen von Schadstoffbelastungen bzw. die Identifikation hygienischer Probleme zählen zu den Aufgaben der öffentlichen Gesundheitsvorsorge („public health“). Entscheidungsunterstützung für die Gesundheitspolitik, aber auch für die Heimaufsicht bzw. für Schulverwaltungsämter (Schulschließungen bei starker Verbreitung von Infektionskrankheiten), ist die wichtigste Aufgabe. Noch wichtiger als bei epidemiologischen Fragen ist dabei der räumliche

Bezug. Dies betrifft die individuelle Mobilität von Patienten und die Ausbreitung einer Erkrankung. Geografische Karten, graphbasierte Darstellungen, Clusteranalysetechniken und die Integration der Clusteranalyseergebnisse mit geografischen Daten sind dabei bewährte Techniken [5].

Oft ist es wichtig, mehrere Erkrankungen im Blick zu haben – so tritt eine Lungenentzündung häufig als Komplikation einer Grippeerkrankung auf. Vorhersagen der weiteren Entwicklung sind vor allem wichtig, um den lokalen Bedarf an Impfstoffen und anderen Medikamenten abzuschätzen. Bewährt haben sich für diese Zwecke Multiparametervisualisierungen, wie Farbicons, die die Häufigkeit mehrerer Erkrankungen im Zusammenhang darstellen. Diese iconhaften Darstellungen lassen sich exakt über einer Landkarte positionieren. Für eine leichte Interpretierbarkeit ist es wichtig, dass sich die Icons nicht überlappen, aber auch andere Information wie Gebietsgrenzen und wichtige Flüsse möglichst wenig verdeckt werden, damit die geografische Zuordnung unterstützt wird [13]. Ausgefeilte Methoden der visuellen Analyse wurden von [8] genutzt, um Zusammenhänge zwischen Atemwegserkrankungen bei Hunden und Menschen zu erkennen und damit evtl. früher auf eine nahende Epidemie hinweisen zu können. Insbesondere wurden dynamische geografische Visualisierungen

genutzt – also Visualisierungen, die sich über die Zeit ändern und dadurch die regional unterschiedliche Krankheitsausbreitung veranschaulichen. Die visuell gefundenen Auffälligkeiten wurden mit statistischen Methoden gezielt untersucht. So konnte der Einfluss von industriellen Abfällen auf Erkrankungen von Haustieren effizient untersucht werden.

Visuelle Analyse in der klinischen Forschung

In Krankenhäusern fallen viele Daten an, die für die Forschung hochinteressant sind, selbst wenn sie teilweise dafür nicht erhoben wurden. So entstehen in pathologischen Instituten teilweise große Sammlungen von Gewebeproben, deren Digitalisierung und Annotation ein großes Potenzial für die Forschung bietet, aber auch eine große Herausforderung bzgl. der Auswertung darstellt [9]. Als Beispiel für den Einsatz von Methoden der visuellen Analyse in der klinischen Forschung werden im Folgenden drei Beispiele vorgestellt:

- die Auswertung von zeitveränderlichen Bilddaten (Perfusions-Daten),
- die Auswertung von Simulationsdaten, mit denen z. B. die Verteilung der Luft in den Atemwegen untersucht wird und
- die Auswertung von Genexpressionsdaten.

Während im ersten Beispiel Daten aus der klinischen Routine genutzt werden, werden in den anderen beiden Beispielen Daten ausgewertet, die bisher vorrangig im Rahmen der medizinischen Forschung entstehen.

Diese Anwendungen erfordern verschiedene Sichten auf die Daten und profitieren davon, dass diese Sichten geeignet synchronisiert bzw. koordiniert werden. In allen Anwendungen ist der anatomische bzw. biologische Kontext wichtig, so dass abgeleitete Daten in Relation zu bekannten anatomischen Regionen analysiert und interpretiert werden können.

Visuelle Analyse von Perfusions-Daten

Bei Perfusions-Daten handelt es sich um eine besondere Form von medizinischen Bilddaten (meist MRT, aber auch CT), die darstellen, wie sich ein Kontrastmittel, das über die Blutgefäße eingespritzt wird, im Körper über die Zeit verteilt. Dieses zeit-

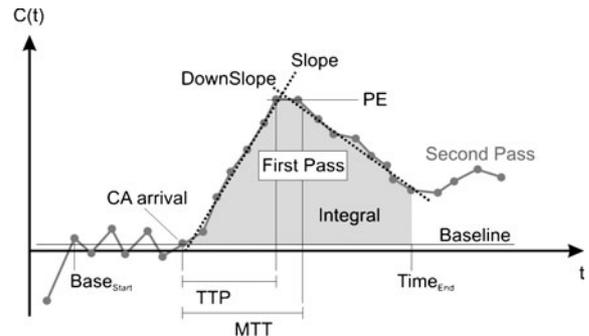


Abb. 3 Für selektierte Bildpunkte oder Regionen werden Zeit-Intensitätskurven abgeleitet. Die Intensität schwankt zunächst stochastisch, ehe das Kontrastmittel ankommt (CA arrival), steigt dann, wobei der Anstieg (Slope) relevant ist. PE bezeichnet das Maximum (Peak Enhancement) und TTP, die Zeit bis zum Erreichen dieses Maximums (Time To Peak). Das Integral charakterisiert, wieviel Kontrastmittel insgesamt im ersten Durchgang angereichert wird. Der Schwerpunkt unter der Kurve ist die Mean Transit Time (MTT)

liche Verhalten ist regional stark unterschiedlich; in den Blutgefäßen selbst bzw. in gut durchbluteten Bereichen ist mit einer schnellen und starken Kontrastmittelanreicherung zu rechnen. Dagegen ist in schlecht durchbluteten Bereichen oder solchen, die gar nicht (mehr) durchblutet werden, mit einer verringerten und verzögerten Anreicherung zu rechnen. Zeit-Intensitätskurven pro Bildpunkt oder gemittelt über eine bestimmte Region charakterisieren dieses Verhalten. Prinzipiell ist diese Information hochinteressant für die Diagnose wichtiger Erkrankungen. So deutet eine starke Kontrastmittelanreicherung eines fraglichen Tumors, gefolgt von einem Signalabfall, stark auf dessen Bösartigkeit hin. Grund ist, dass ein bösartiger Tumor für sein schnelles Wachstum eine besonders gute Gefäßversorgung benötigt. Diese vielen Gefäße sind aber oft unreif, löchrig und halten daher das Kontrastmittel nicht lange. Ein anderes Beispiel betrifft die Schlaganfalldiagnostik, bei dem die Mediziner die Lage und Ausdehnung eines nicht mehr mit Blut versorgten Schlaganfallkerngebietes aber auch das evtl. noch zu rettende minderdurchblutete Gehirngewebe in der Umgebung analysieren. Um die Diagnostik zu erleichtern und zu präzisieren, werden sogenannte Perfusions-Parameter genutzt, die die Kurve charakterisieren. Dies sind deskriptive Parameter, wie z. B. das Maximum der Kontrastmittelanreicherung und der Zeitpunkt, zu dem dieses Maximum erreicht wird (Abb. 3).

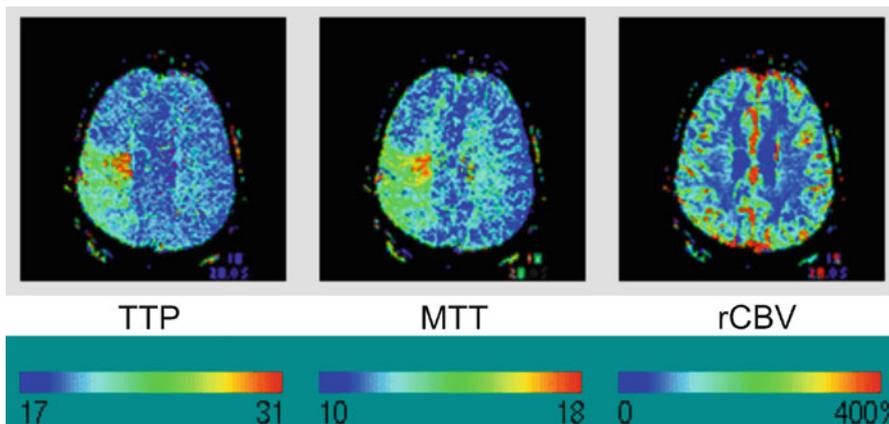


Abb. 4 Die Perfusions-Parameter werden für die ausgewählte Schicht einer Computertomografie des Schädels farbkodiert dargestellt. Neben den in Abb. 3 erklärten Parametern kommt hier das „relative Cerebral Blood Volume“ (rCBV) hinzu (rechtes Bild). Die Farbskala ist an die tatsächlich vorkommenden minimalen und maximalen Werte für den jeweiligen Parameter angepasst. Der Arzt muss Auffälligkeiten in diesen Darstellungen mental überlagern. Im linken und mittleren Bild fällt schon mit bloßem Auge die Asymmetrie auf; die Blutversorgung ist in einer Gehirnhälfte stark verzögert. Das rechte Bild (rCBV) charakterisiert die Gesamtmenge des aufgenommenen Kontrastmittels und zeigt kaum Auffälligkeiten; die von der verzögerten Versorgung betroffenen Gehirnabschnitte werden demnach noch gut versorgt

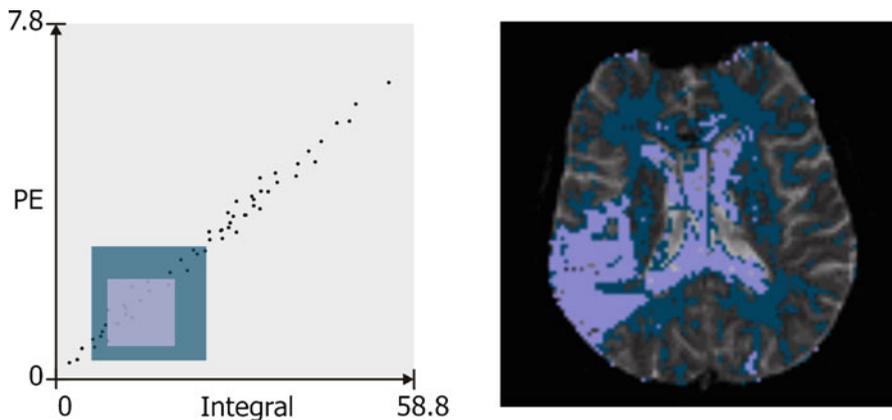


Abb. 5 Ein Scatterplot, in dem die Merkmalsausprägungen von zwei Perfusions-Parametern dargestellt sind, wird zur Auswahl interessierender Regionen genutzt. Die beiden rechteckigen Selektionen repräsentieren Intervalle für die beiden Perfusions-Parameter und führen dazu, dass die Bildpunkte, die diesen Spezifikationen entsprechen, im rechten Bild hervorgehoben werden (Brushing)

Die Routinediagnostik, besonders bei Schlaganfallverdacht, wo wichtige Therapieentscheidungen in sehr kurzer Zeit fallen müssen, basiert auf einer qualitativen Betrachtung der Bilddaten und einiger abgeleiteter Zeit-Intensitätskurven, wobei jeweils nur wenige Parameterkarten betrachtet werden (Abb. 4, 5).

Forschungsfragen in der Perfusions-Diagnostik sind nun darauf gerichtet, die Aussagekraft einzelner Perfusions-Parameter bzw. deren Kombination zu verstehen. Diese Frage ist insofern schwierig, als dass die Antwort in subtiler Weise von Parametern

der Bildgebung abhängt. Welches Kontrastmittel wird in welcher Dosis verabreicht? Wie schnell wird es injiziert? Welche Parameter des Computertomografen bzw. des MRT-Gerätes werden eingestellt? Hinzu kommen die Abhängigkeiten vom Patienten: Alter, Gewicht, Geschlecht, bei Frauen sogar die Frage, in welcher Phase des Menstruationszyklus sie sich befindet, beeinflussen die Ergebnisse einer Perfusions-Bildgebung.

Um diese Fragen zu beantworten, wurde an der Universität Magdeburg ein Framework entwickelt, in dem für einen Patientendatensatz untersucht werden

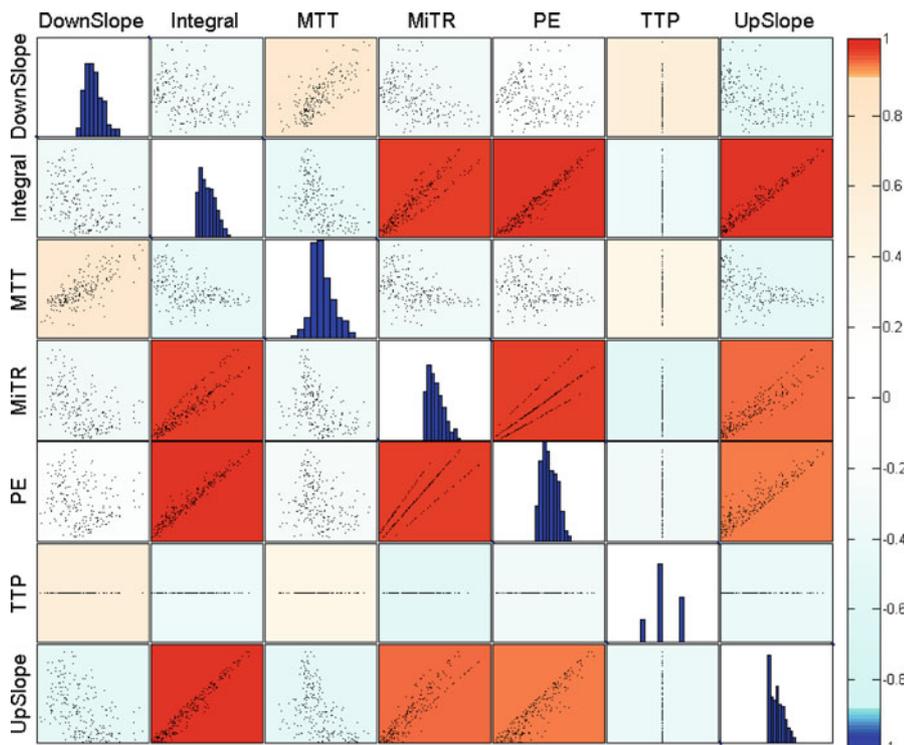


Abb. 6 Eine Scatterplot-Matrix veranschaulicht die Korrelationen von Perfusions-Parametern. Starke positive und negative Korrelationen werden durch die gewählte Farbskala hervorgehoben

kann, wie stark Perfusions-Parameter korrelieren (Abb. 6). So kann die Auswertung auf unabhängige Parameter konzentriert werden [11]. Die verbleibenden Parameter wurden dann mittels Principal Component Analysis (PCA) weiter reduziert, wobei sich über verschiedene Datensätze und sogar Anwendungsgebiete relativ einheitliche Trends zeigten. Die PCA transformiert einen n -dimensionalen Datensatz in ein neues orthogonales Koordinatensystem. Dabei wird eine Eigenwertanalyse der Kovarianzmatrix vorgenommen, wobei die Eigenvektoren Linearkombinationen der ursprünglichen Dimensionen sind und das neue Koordinatensystem bilden. Die zugehörigen Eigenwerte bestimmen die Bedeutung der Eigenvektoren, wobei häufig sehr wenige Eigenvektoren eine hohe Bedeutung haben und andere somit vernachlässigt werden können – die Dimension der Daten wird reduziert. Die PCA (in einigen Communities auch als Karhunen-Loeve-Transformation bekannt) ist eine von mehreren üblichen Techniken der Dimensionsreduktion [14].

Die Darstellung der Werte in den neuen Dimensionen wurde dann für ein Brushing genutzt, also für die Auswahl relevanter Regionen in den Bilddaten. Es ließ sich zeigen, dass klinisch relevante Gebiete,

wie das Kerngebiet eines Schlaganfalls oder das gefährdete Gewebe in der Umgebung dieses Kerns, dadurch sicher und schnell identifiziert werden konnten. Die Arbeiten wurden weiterentwickelt, mit dem Ziel, direkt anhand der Zeit-Intensitätskurven interessierende Gebiete auszuwählen [10]. Dafür war es nötig, übersichtliche Darstellungen einer sehr großen Anzahl an Zeit-Intensitätskurven zu generieren (Abb. 7).

Bildanalyse in der Auswertung von Perfusions-Daten. Die alleinige Nutzung von Datenanalyse- und Visualisierungstechniken ist insofern unzureichend, als dass die manuelle und subjektive Selektion von Regionen eine wichtige Rolle spielt. Regionen sollten einerseits nicht zu klein sein, sodass die Mittelwertbildung in einer Region zu möglichst zuverlässigen Ergebnissen führt. Andererseits setzt eine regionenbasierte Auswertung voraus, dass die Bildelemente in der Region in Bezug auf die Perfusions-Parameter homogen sind. Methoden der Regionenanalyse und Klassifikation können genutzt werden, um zu einer zuverlässigeren Regionenauswahl zu kommen [4]. Zudem sind Teilregionen auch nach der Korrektur von Bewegungsartefakten von Registrierungsfehlern behaftet. Das führt vor al-

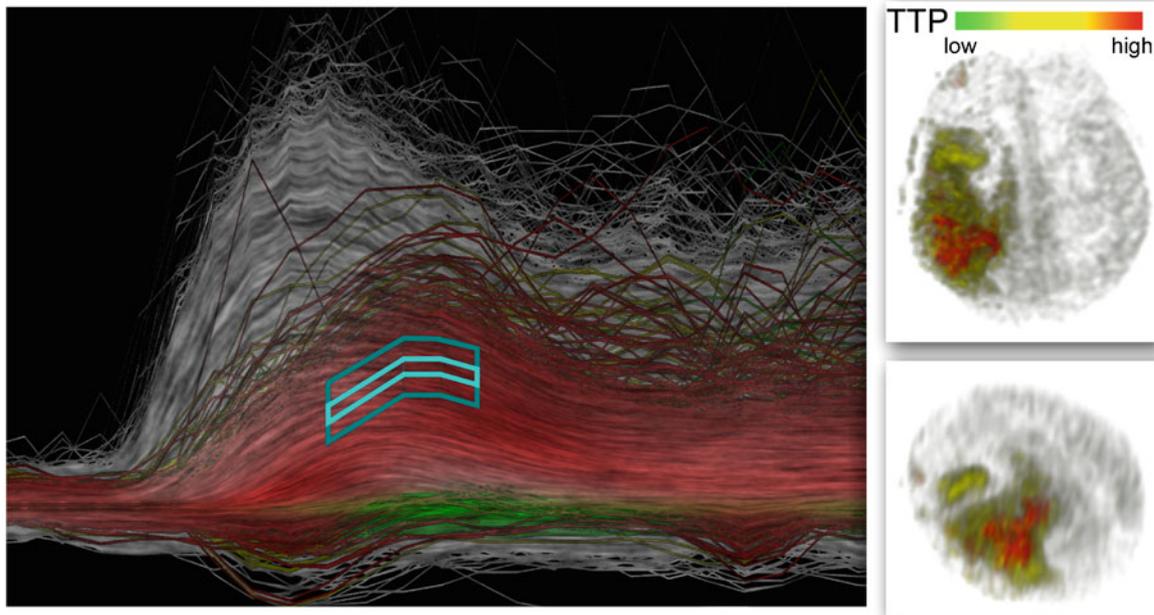


Abb. 7 Unter tausenden von Zeit-Intensitätskurven werden diejenigen hervorgehoben, die einer Referenzkurve ähnlich sind. Die Bildpunkte, die zu Kurven gehören, die dieses Kriterium erfüllen, werden hervorgehoben. Sowohl der Kern eines Schlaganfalls (rot) als auch seine gefährdete Umgebung (grün) können so mittels Brushing hervorgehoben werden. (Aus [10])

lem bei kleinen Regionen zu einer Verfälschung von Perfusions-Parametern.

Visuelle Analyse von Simulationsdaten

Simulationen spielen in der klinischen Forschung eine wichtige Rolle. Biomechanische Simulationen dienen z. B. in der Orthopädie dazu, die Belastung von Implantaten und Knochen vor Becken-, Knie- und Schulteroperationen vorherzusagen und so die Auswahl eines passenden Implantats zu unterstützen. In der HNO-Chirurgie versucht man, durch biophysikalische Simulationen vorherzusagen, wie sich das frequenzabhängige Hörvermögen nach einer Mittelohroperation verändern wird. Blutflussimulationen werden genutzt, um zu verstehen, wie krankhaft veränderte Gefäßabschnitte durchströmt werden und wie groß davon ausgehend die Blutungsgefahr ist. Die Simulationen führen jeweils zu einer großen Menge an zeitveränderlichen Daten. Die Auswertung dieser Simulationsergebnisse im anatomischen Kontext ist äußerst schwierig. Bisher dominieren in diesem Bereich Simulationswerkzeuge und deren Möglichkeiten der Ergebnisdarstellung, z. B. einfache Stromlinien, um den Blutfluss darzustellen. Hilfreich wären angepasste Methoden und eine Verknüpfung mit

Interaktionsmöglichkeiten, die die Darstellung auf relevante Teilbereiche einschränken. Ein aktuelles Beispiel ist in Zusammenarbeit des Zuse-Instituts Berlin mit dem VRVis in Wien entstanden. Das in Wien entstandene „SimVis“-Tool wurde in dieser Kooperation so angepasst, dass der simulierte Luftfluss durch die Atemwege sowie die abgeleiteten Informationen dargestellt werden können [15]. Abbildung 8 zeigt ein Beispiel, wobei vier der insgesamt sechs koordinierten Bildschirmfenster dargestellt sind.

Auch bei der Behandlung krankhaft veränderter Gefäßabschnitte werden Implantate (Drahtgitter, die Stents genannt werden) zur Behandlung eingesetzt. Wiederum ist die anatomisch passende Platzierung (Durchmesser und Krümmung des Drahtgitters) notwendige, aber nicht hinreichende Voraussetzung für ein befriedigendes Operationsergebnis. Man erhofft sich durch Simulationen des Blutflusses vor und nach der Stentimplantation vor allem Fortschritte in der Diagnostik und Behandlung von potenziell lebensgefährlichen Gefäßerweiterungen (Aneurysmen). Abbildung 9 zeigt dafür ein Beispiel. Die hier beschriebene visuelle Analyse von Simulationsdaten ist allerdings noch im Forschungsstadium – der klinische Wert ist nicht abschließend

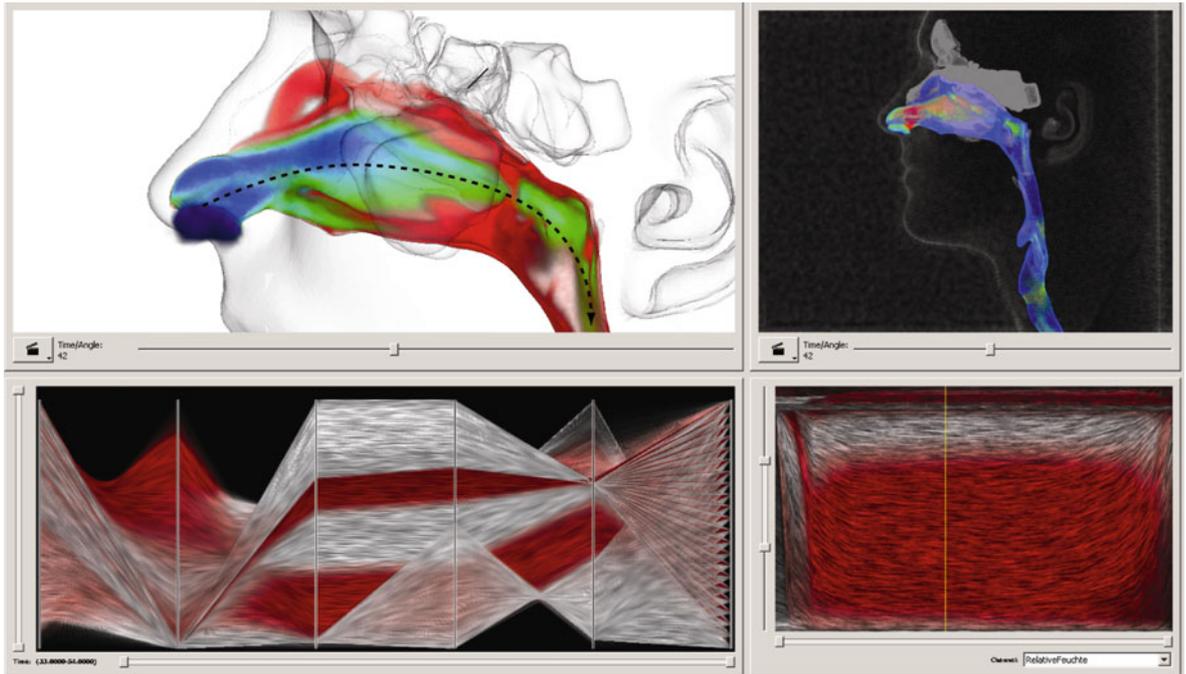


Abb. 8 Visuelle Analyse von Druck-, Temperatur- und Geschwindigkeitsverteilung beim Ein- und Ausatmen. Die oberen Ansichten stellen die Anatomie und ausgewählte Aspekte der Simulation (links die Temperatur, rechts die Geschwindigkeit) dar. Die unteren Ansichten repräsentieren Attributansichten und können genutzt werden, um Regionen auszuwählen, die bestimmte Merkmale aufweisen. Links unten ist eine parallele Koordinatenansicht, die u. a. Druck, Temperatur und Luftfeuchtigkeit als Dimensionen einhält. Rechts unten ist die Veränderung der Luftfeuchtigkeit über die Zeit dargestellt. (Aus [15])

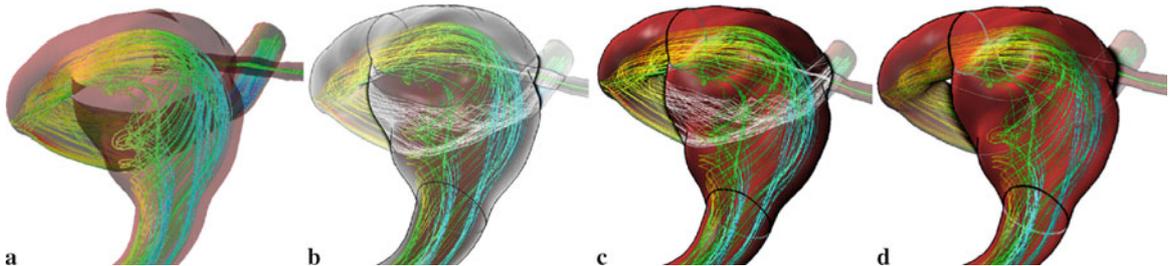


Abb. 9 Visuelle Analyse von simuliertem Blutfluss in einem zerebralen Aneurysma. Die Stromlinien stellen den Blutfluss dar, wobei die Farbe die Flussgeschwindigkeit repräsentiert. Gegenüber dem naiv erzeugten Bild (a) ist die Transparenz des umgebenden Gefäßes so angepasst worden, dass die Gefäßränder gut erkennbar sind und die Blutflussinformation im Inneren verbessert wird (b). Durch eine bessere Farbskala (c) lässt sich der räumliche Eindruck weiter verbessern (c). In (d) sind schließlich die verdeckten Stromlinien entfernt worden. (Aus [3])

geklärt und der hohe Aufwand für die Simulationen steht einem Routineeinsatz bisher im Wege.

Visuelle Analyse von Genexpressionsdaten

Medizinische Daten fallen auch in der biomedizinischen Grundlagenforschung an, insbesondere in den sehr aktiven Gebieten der Genetik bzw. der molekularen Zellbiologie. Dabei geht es z. B. darum, Genexpressionsdaten zu analysieren, um zu verstehen, wie stark bestimmte Gene in bestimmten

Zellen oder Zellgruppen exprimiert werden (Genexpressionsdaten) und wie die Genexpression normale Prozesse wie Wachstum, Zellteilung und Anpassung an veränderte Umweltbedingungen steuert bzw. sich bei krankhaften Veränderungen anpasst. Diese Daten werden in der pharmazeutischen Forschung oft mit Daten über biologische und chemische Prozesse in Proteinen (Proteomics) und toxikologischen Daten verknüpft. So wird z. B. untersucht, welche Muster der Genexpression bei Leukämiepatienten

aufzutreten, welche Veränderungen der Proteine in den betroffenen Zellen auftreten und wo Medikamente ansetzen müssten, um die Verbreitung der krankhaften Zellen zu hemmen [12].

Die meisten Gene sind nur in wenigen Zellen überhaupt aktiv. Genexpressionsdaten werden in standardisierten Experimenten in großer Zahl gewonnen und in DNA-Microarrays repräsentiert. Ein einzelnes Experiment enthält bis zu 30.000 Werte, wobei meist die Ergebnisse mehrerer Experimente im Zusammenhang ausgewertet werden [12]. Ein umfassendes und besonders sichtbares Projekt zur Analyse derartiger Daten wird derzeit an der TU Graz durchgeführt, wobei ebenfalls die zuvor besprochenen Techniken der interaktiven visuellen Analyse zum Einsatz kommen [6, 7], vor allem parallele Koordinatendarstellungen sind hier bedeutsam. Das Grazer Projekt ist besonders ambitioniert, da die Genexpressionsdaten im Zusammenhang mit „biological pathways“ betrachtet werden, um Zusammenhänge zu krankhaften Veränderungen zu identifizieren. Diese pathways sind generische Modelle von biologischen Aktivitäten, wie Stoffwechsel oder Immunreaktionen und werden als Graphen dargestellt.

Zusammenfassung

Das Potenzial von visuellen Analysemethoden in der Epidemiologie und in der öffentlichen Gesundheitsvorsorge ist erheblich und unbestritten, wobei die regionale Verteilung (geografischer Bezug) und zeitliche Veränderungen wesentliche Aspekte darstellen. Auch bei der Auswertung klinischer Studien und bei ausgewählten Fragen der patientenorientierten Forschung besteht ein hoher Bedarf an speziellen Methoden der Auswertung. In der Praxis dominiert die statistische Auswertung, die Mediziner in ihrer Ausbildung erlernen. Die grafische Auswertung ist in der Praxis bisher ebenfalls stark auf einfache Methoden beschränkt – die Verfügbarkeit in verbreiteten Statistikpaketen spielt dabei eine wesentliche Rolle. Eine bessere Unterstützung ist mit hoher Wahrscheinlichkeit möglich; bedarf aber einer engen Abstimmung mit den Anwendern,

eines präzisen Verständnisses der typischen Probleme und einer sorgfältigen Evaluierung. Einige der in der Literatur bisher dargestellten Techniken nutzen zwar medizinische Daten, dienen aber eher dazu, eine Visualisierungstechnik zu verfeinern als die Anwendungsprobleme konkret zu lösen. Systeme und Algorithmen der geografischen Visualisierung haben ein besonderes Potenzial, um an die Anwendungsprobleme angepasst zu werden. Die Anpassung müsste die Terminologie, die verwendeten grafischen Symbole und die Workflows betreffen, mit denen Daten in den jeweiligen medizinischen Fachdisziplinen ausgewertet werden.

Literatur

1. Bade R, Schlechtweg S, Miksch S (2004) Connecting time-oriented data and information to a coherent interactive visualization. Proc of ACM SIGCHI, pp 105–112
2. Dai X, Gahegan M (2005) Visualization based approach for exploration of health data and risk factors. In: Proceedings of the 8th International Conference on Geo-Computation. University of Michigan, USA, 31 July–3 August 2005
3. Gasteiger R, Neugebauer M, Kubisch C, Preim B (2010) Adapted surface visualization of cerebral aneurysms with embedded blood flow information. Eurographics Workshop on Visual Computing for Biology and Medicine (EG VCBM), Leipzig, 1–2 June 2010, pp 25–32
4. Glaßer S, Schäfer S, Oeltze S, Preim U, Tönnies K-D, Preim B (2009) A visual analytics approach to diagnosis of breast DCE-MRI data. Proc of Vision, Modeling, and Visualization (VMV), pp 351–362
5. Guo D (2007) Visual analytics of spatial interaction patterns for pandemic decision support. Int J Geogr Inf Sci 21(8):859–877
6. Lex A, Streit M, Kruijff E, Schmalstieg D (2010a) Caleydo: Design and evaluation of a visual analysis framework for gene expression data in its biological context. In: Proceedings of the IEEE Pacific Visualization Symposium, Taipei, Taiwan, 2–5 March 2010, pp 57–64
7. Lex A, Streit M, Partl C, Kashofer K, Schmalstieg D (2010b) Comparative analysis of multidimensional, quantitative data. IEEE Transactions on Visualization and Computer Graphics, Bd 16(6)
8. Maciejewski R, Tyner B, Jang Y et al (2007) LAHVA: linked animal-human health visual analytics. Proc of the IEEE Symposium on Visual Analytics Science and Technology, pp 27–34
9. Müller H, Reihls R, Sauer S, Zatloukal K, Streit M, Lex A, Schlegl B, Schmalstieg D (2009) Connecting genes with diseases. In: Proceedings of the 2009 13th International Conference on Information Visualization 2009, 13th International Conference on Information Visualization, 14–17 July 2009, Barcelona, Spain, pp 323–330
10. Muigg P, Kehrer J, Oeltze S, Piringer H, Doleisch H, Preim B, Hauser H (2008) A four-level focus + context approach to interactive visual – analysis of temporal features in large scientific data. Comput Graph Forum (EuroVis) 27(3):775–782
11. Oeltze S, Muigg P, Doleisch H, Hauser H, Preim B (2007) Interactive visual analysis of perfusion data. IEEE Trans Vis Graph 13(6):1392–1399
12. Saffer JD, Burnett VL, Chen G, van der Spek P (2004) Visual analytics in the pharmaceutical industry. IEEE Comput Graph Appl 24(5):10–15
13. Schulze-Wollgast P, Schumann H, Tominski C (2003) Visual analysis of human health data. IRMA International Conference, Philadelphia, USA
14. Ward M, Grinstein GG, Keim D (2010) Interactive Data Visualization: Foundations, Techniques, and Applications. AK Peters
15. Zachow S, Muigg P, Hildebrandt T, Doleisch H, Hege H-C (2009) Visual exploration of nasal airflow. IEEE Trans Vis Graph 15(6):1407–1414