

(12) NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG

(19) Weltorganisation für geistiges Eigentum
Internationales Büro



(43) Internationales Veröffentlichungsdatum
1. März 2001 (01.03.2001)

PCT

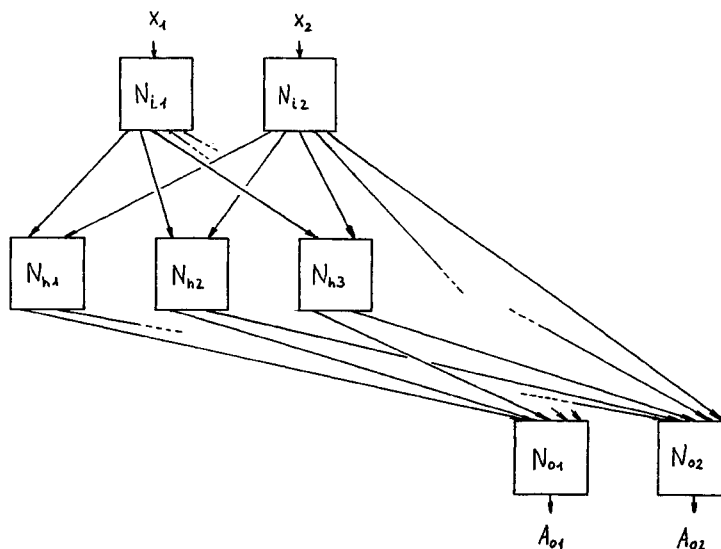
(10) Internationale Veröffentlichungsnummer
WO 01/15078 A2

- (51) Internationale Patentklassifikation⁷: **G06N 3/00** (71) **Anmelder** (für alle Bestimmungsstaaten mit Ausnahme von US): **WILEX BIOTECHNOLOGY GMBH** [DE/DE]; Grillparzer Strasse 10B, 81675 München (DE).
- (21) Internationales Aktenzeichen: PCT/EP00/08280
- (22) Internationales Anmeldedatum: 24. August 2000 (24.08.2000) (72) **Erfinder; und**
(75) **Erfinder/Anmelder** (nur für US): **KATES, Ronald** [US/DE]; Palnkamer Strasse 49, 83624 Otterfing (DE). **HARBECK, Nadia** [DE/DE]; Palnkamer Strasse 49, 83624 Otterfing (DE). **SCHMITT, Manfred** [DE/DE]; Hohenaschauer Strasse, 81669 München (DE).
- (25) Einreichungssprache: Deutsch
- (26) Veröffentlichungssprache: Deutsch
- (30) Angaben zur Priorität: 199 40 577.8 26. August 1999 (26.08.1999) DE (74) **Anwälte: WEICKMANN, H.** usw.; Kopernikusstrasse 9, 81679 München (DE).

[Fortsetzung auf der nächsten Seite]

(54) Title: METHOD FOR TRAINING A NEURAL NETWORK

(54) Bezeichnung: VERFAHREN ZUM TRAINIEREN EINES NEURONALEN NETZES

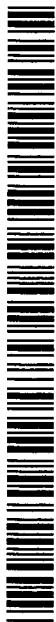


(57) **Abstract:** The object of the inventive method is to train a neural network to determine risk functions in patients in association with a first illness relating to a predetermined illness based on the predetermined training data sets. Said predetermined data sets contain objectifiable information which can be used as a yardstick to measure the pathological condition of the patient. The neural network contains a plurality of neurons arranged in several layers, in addition to synapses connecting said neurons. In the course of the training, the structure of the neural network is simplified by tracking and eliminating the synapses which play no significant role in the evolution of the risk functions. This can be done, for example, by examining a possible correlation between the influences which two sending neurons have on the same receiving neuron, and where possible, eliminating one of the two synapses connecting to the receiving neuron.

(57) **Zusammenfassung:** Das erfindungsgemäße Verfahren dient zum Trainieren eines neuronalen Netzes zur Ermittlung von Risikofunktionen für Patienten im Anschluß an eine Ersterkrankung mit einer vorbestimmten Krankheit auf Grundlage vorgegebener Trainings-Datensätze, welche objektivierbare und

[Fortsetzung auf der nächsten Seite]

WO 01/15078 A2





(81) **Bestimmungsstaaten** (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI-Patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Veröffentlicht:

— *Ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts.*

(84) **Bestimmungsstaaten** (*regional*): ARIPO-Patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI,

Zur Erklärung der Zweibuchstaben-Codes, und der anderen Abkürzungen wird auf die Erklärungen ("Guidance Notes on Codes and Abbreviations") am Anfang jeder regulären Ausgabe der PCT-Gazette verwiesen.

meßtechnisch erfaßte Daten zum Krankheitsbild der Patienten beinhalten. Das neuronale Netz umfaßt eine Mehrzahl von in mehreren Schichten angeordneten Neuronen sowie diese Neuronen verbindende Synapsen. Im Verlaufe des Trainings wird die Struktur des neuronalen Netzes vereinfacht, indem Synapsen aufgespürt und eliminiert werden, welche auf den Verlauf der Risikofunktion keinen wesentlichen Einfluß ausüben. Dies kann beispielsweise dadurch erfolgen, daß man die Einflüsse, die zwei Sende-Neuronen auf ein und dasselbe Empfangs-Neuron ausüben, auf eine mögliche Korrelation untersucht und gegebenenfalls eine der beiden zu dem Empfangs-Neuron führenden Synapsen eliminiert.

Verfahren zum Trainieren eines neuronalen Netzes

5

Beschreibung

1. FACHGEBIET DER ERFINDUNG

10

Die Erfindung betrifft ein Verfahren zum Trainieren eines neuronalen Netzes zur Ermittlung von Risikofunktionen für Patienten im Anschluß an eine Erst-erkrankung mit einer vorbestimmten Krankheit auf Grundlage vorgegebener Trainings-Datensätze, welche objektivierbare und zu einem erheblichen Teil meßtechnisch erfaßte Daten zum Krankheitsbild der Patienten beinhalten, wobei das neuronale Netz eine Eingangsschicht mit einer Mehrzahl von Eingangs-Neuronen und wenigstens eine Zwischenschicht mit einer Mehrzahl von Zwischen-Neuronen umfaßt, sowie ferner eine Ausgangsschicht mit einer Mehrzahl von Ausgangs-Neuronen und eine Vielzahl von Synap-
20 sen, die jeweils zwei Neuronen unterschiedlicher Schichten miteinander verbinden.

2. FACHLICHER HINTERGRUND - STAND DER TECHNIK

25

2.1. Allgemeines

Auf zahlreichen Gebieten sind neuronale Netze zur Auswertung großer Datenmengen zu den bislang üblichen Auswerteverfahren ergänzend hinzu-
30 getreten oder haben diese ersetzt. Es hat sich nämlich gezeigt, daß neuronale Netze besser als die herkömmlichen Verfahren in der Lage sind, in den Datenmengen verborgene, nicht ohne weiteres erkennbare Abhängigkeiten zwischen den einzelnen Eingangsdaten aufzuspüren und aufzuzeigen.

- 2 -

Daher liefern neuronale Netze, welche anhand einer bekannten Datenmenge trainiert worden sind, für neue Eingangsdaten des gleichen Datentyps zuverlässigere Aussagen als die bisherigen Auswerteverfahren.

- 5 Im Bereich medizinischer Anwendungen ist es beispielsweise bekannt, neuronale Netze zur Bestimmung einer Überlebensfunktion für an einer bestimmten Krankheit, etwa Krebs, leidende Patienten einzusetzen. Diese Überlebensfunktion gibt dabei in Abhängigkeit der seit der Ersterkrankung vergangenen Zeit die Wahrscheinlichkeit an, daß bei einem betrachteten
- 10 Patienten ein vorbestimmtes Ereignis eintritt. Dieses vorbestimmte Ereignis muß dabei nicht notwendigerweise der Tod des Patienten sein, wie dies die Bezeichnung "Überlebensfunktion" vermuten läßt, sondern kann ein beliebiges Ereignis sein, beispielsweise eine erneute Erkrankung an Krebs.
- 15 Die Datensätze umfassen eine ganze Reihe von objektivierbaren Angaben, d.h. Daten, auf deren Wert eine das neuronale Netz gegebenenfalls bedienende Person keinen Einfluß hat und deren Wert gewünschtenfalls automatisch erfaßt werden kann. Im Falle von Brustkrebs sind dies Angaben über die Person des Patienten, beispielsweise Alter, Geschlecht und
- 20 dergleichen, Angaben über das Krankheitsbild, beispielsweise Anzahl der von Krebs befallenen Lymph-Knoten, tumor-biologische Faktoren, wie upA (Urokinase-Plasminogen-Aktivator), sein Inhibitor PAI-1 und dergleichen Faktoren, sowie Angaben über die Behandlungsmethode, beispielsweise Art, Dauer und Intensität einer Chemo- oder Strahlen-Therapie. Unnötig zu
- 25 erwähnen, daß eine ganze Reihe der vorstehend erwähnten Angaben, insbesondere der Angaben zum Krankheitsbild, nur unter Einsatz geeigneter Meßapparaturen ermittelt werden können. Darüber hinaus können auch die personen-bezogenen Daten automatisiert von geeigneten Datenträgern abgelesen werden, beispielsweise maschinenlesbaren Personalausweisen
- 30 oder dergleichen. Selbstverständlich können die objektivierbaren Daten dann, wenn sie nicht alle gleichzeitig verfügbar sind, was insbesondere bei den Labor-Meßwerten häufig der Fall sein dürfte, auf einem geeigneten

Speichermedium in einer Datenbank zwischengespeichert werden, bevor sie dem neuronalen Netz als Eingangsdaten zugeführt werden.

5 2.2. Das neuronale Netz als Signalfilter

Gemäß vorstehendem kann man ein neuronales Netz also als eine Art "Signalfilter" auffassen, das aus einem verrauschten und daher an sich noch nicht aussagekräftigen Eingangssignal ein aussagekräftiges Ausgangssignal
10 herausfiltert. Ob bzw. wie gut das Filter seine Funktion erfüllen kann, hängt wie bei jedem Filter davon ab, ob es gelingt, die Intensität des Eigenrauschens des Filters so niedrig zu halten, daß das herauszufilternde Signal nicht in diesem Eigenrauschen untergeht.

15 Die Intensität des "Eigenrauschens" eines neuronalen Netzes ist umso niedriger, je mehr Datensätze einerseits zum Trainieren des neuronalen Netzes zur Verfügung stehen und je einfacher andererseits die Struktur des neuronalen Netzes ist. Zudem steigt die Generalisierungsfähigkeit des Netzes mit zunehmender Vereinfachung der Struktur des neuronalen Netzes
20 an. Bei einer im Stand der Technik üblichen Vorgehensweise befaßt sich daher ein Teil des Trainings neuronaler Netze mit dem Auffinden und Eliminieren von Strukturteilen, die für den Erhalt eines aussagekräftigen Ausgangssignals entbehrlich sind. Bei diesem "Ausdünnen" (in der Fachsprache auch als "pruning" bezeichnet) ist allerdings als weitere
25 Randbedingung zu beachten, daß die Struktur des neuronalen Netzes nicht beliebig "ausgedünnt" werden darf, da mit abnehmender Komplexität des neuronalen Netzes auch dessen Fähigkeit, komplexe Zusammenhänge nachzubilden, und somit dessen Aussagekraft abnimmt.

2.3. Probleme beim medizinischen Einsatz

Nun steht man in der Praxis, insbesondere bei dem eingangs angesprochenen medizinischen Einsatz neuronaler Netze, häufig vor dem Problem, daß
5 lediglich sehr kleine Datenmengen von typischerweise wenigen hundert Datensätzen für das Training des neuronalen Netzes zur Verfügung stehen. Erschwerend kommt hinzu, daß für das Training nicht nur eine Trainings-Datenmenge, sondern ferner auch eine Validierungs-Datenmenge und eine Generalisierungs-Datenmenge vorgesehen werden müssen. Auf die Bedeutung dieser beiden Datenmengen wird weiter unten unter den Punkten 5.5.
10 bzw. 5.7. noch näher eingegangen werden.

Bei derart geringen Datenmengen führte der Einsatz bekannter Pruning-Methoden aber stets zu einer so starken Vereinfachung der Struktur des
15 neuronalen Netzes, daß die Aussagekraft des neuronalen Netzes auf ein nicht zu tolerierendes Maß absank. Um nach Abschluß der Trainingsphase aber dennoch neuronale Netze erhalten zu können, welche aussagekräftige Ausgangssignale liefern, wurden daher im Stand der Technik bei Vorliegen lediglich kleiner Trainingsdatenmengen neuronale Netze mit einer starren,
20 d.h. fest vorgegebenen und unveränderlichen Struktur eingesetzt. Dabei wurde der Komplexitätsgrad bzw. die Einfachheit dieser starren Struktur auf Grundlage von Erfahrungswissen derart gewählt, daß das neuronale Netz einerseits eine hohe Aussagekraft und andererseits ein gerade noch tolerierbares Eigenrauschen aufwies. Bisläng wurde davon ausgegangen,
25 daß die Vorgabe einer unveränderlichen Struktur unumgänglich sei.

Ein weiteres Problem der medizinischen Anwendungen neuronaler Netze ist darin zu sehen, daß zum Training nur "zensierte" Daten vorliegen. Mit dem Begriff "Zensur" wird dabei der Umstand umschrieben, daß für Patienten,
30 die zum Zeitpunkt der Datenerfassung glücklicherweise noch keinen Rückfall erlitten haben, die zukünftige Entwicklung nicht vorauszusehen ist, und

- 5 -

Aussagen über die Überlebensfunktion daher nur bis zum Zeitpunkt der Erhebung der Daten möglich sind.

Es braucht nicht betont zu werden, daß gerade bei medizinischen
5 Anwendungen auf ein wirklich aussagekräftiges Ergebnisses unter gar
keinen Umständen verzichtet werden darf. Es ist nämlich in keinem Fall
akzeptabel, wenn auch nur einem Patient eine Behandlung nur deshalb
vorenthalten würde, weil das neuronale Netz diese nicht für notwendig
erachtet hat. Die Folgen für den Patienten könnten unabsehbar sein.

10 Hinsichtlich der Einzelheiten des vorstehend geschilderten Standes der
Technik sei auf die unter Punkt 6. "Literaturverzeichnis" aufgeführten
Aufsätze verwiesen.

15 3. AUFGABE DER ERFINDUNG

Demgegenüber ist es Aufgabe der Erfindung, ein automatisches Verfahren
zum Trainieren eines neuronalen Netzes zur Ermittlung von Risiko-
20 funktionen für Patienten im Anschluß an eine Ersterkrankung mit einer
vorbestimmten Krankheit bereitzustellen, welches es trotz einer geringen
Anzahl von zur Verfügung stehenden Trainingsdatensätzen erlaubt, ein
neuronales Netz variabler Struktur einzusetzen und dessen Struktur in
wenigstens einem Strukturvereinfachungs-Schritt zu optimieren.

25 4. LÖSUNG DER AUFGABE

Diese Aufgabe wird erfindungsgemäß gelöst durch ein Verfahren zum
30 Trainieren eines neuronalen Netzes zur Ermittlung von Risikofunktionen für
Patienten im Anschluß an eine Ersterkrankung mit einer vorbestimmten
Krankheit auf Grundlage vorgegebener Trainings-Datensätze, welche

- 6 -

objektivierbare und meßtechnisch erfaßte Daten zum Krankheitsbild der Patienten beinhalten, wobei das neuronale Netz umfaßt:

- eine Eingangsschicht mit einer Mehrzahl von Eingangs-Neuronen,
- wenigstens eine Zwischenschicht mit einer Mehrzahl von Zwischen-Neuronen,
- eine Ausgangsschicht mit einer Mehrzahl von Ausgangs-Neuronen, und
- eine Vielzahl von Synapsen, die jeweils zwei Neuronen unterschiedlicher Schichten miteinander verbinden,

wobei das Training des neuronalen Netzes eine Strukturvereinfachungs-Prozedur umfaßt, d.h. das Auffinden und Eliminieren von Synapsen, welche auf den Verlauf der Risikofunktion keinen wesentlichen Einfluß ausüben, indem man entweder

- a1) zwei mit ein und demselben Empfangs-Neuron verbundene Sende-Neuronen auswählt,
- a2) unterstellt, daß die von diesen Sende-Neuronen zum Empfangs-Neuron ausgehenden Signale qualitativ im wesentlichen das gleiche Verhalten zeigen, also miteinander korreliert sind,
- a3) die Synapse eines der beiden Sende-Neuronen zu dem Empfangs-Neuron unterbricht und dafür das Gewicht der Synapse des jeweils anderen Sende-Neurons zu dem Empfangs-Neuron entsprechend anpaßt,
- a4) die Reaktion des gemäß Schritt a3) veränderten neuronalen Netzes mit der Reaktion des unveränderten neuronalen Netzes vergleicht, und
- a5) dann, wenn die Abweichung der Reaktion ein vorbestimmtes Maß nicht überschreitet, entscheidet, die in Schritt a3) vorgenommene Änderung beizubehalten,

oder indem man

- b1) eine Synapse auswählt,
- b2) unterstellt, daß diese Synapse auf den Verlauf der Risikofunktion keinen wesentlichen Einfluß ausübt,

- 7 -

- b3) diese Synapse unterbricht,
- b4) die Reaktion des gemäß Schritt b3) veränderten neuronalen Netzes mit der Reaktion des unveränderten neuronalen Netzes vergleicht, und
- 5 b5) dann, wenn die Abweichung der Reaktion ein vorbestimmtes Maß nicht überschreitet, entscheidet, die in Schritt b3) vorgenommene Änderung beizubehalten.

Ein in der vorstehend beschriebenen Art und Weise trainiertes neuronales
10 Netz unterstützt den behandelnden Arzt beispielsweise bei der Entscheidung, welche Nachbehandlung er bei einem bestimmten frisch operierten Patienten einsetzen soll. Der Arzt kann hierzu dem neuronalen Netz die Patientendaten und die im Labor meßtechnisch erfaßten Daten über das Krankheitsbild der Erstbehandlung eingeben und erhält von dem neuronalen
15 Netz eine Aussage darüber, bei welcher Nachbehandlungsart sich die für den gerade betrachteten Patienten günstigste Überlebensfunktion ergibt. Selbstverständlich kann auch die Aggressivität der einzelnen Nachbehandlungsarten berücksichtigt werden, so daß bei gleich oder annähernd gleich günstiger Überlebensfunktion die für den Patienten schonendste Art der
20 Nachbehandlung ausgewählt werden kann.

5. AUSFÜHRUNGSBEISPIEL

25 Die Erfindung wird im folgenden an einem Ausführungsbeispiel näher erläutert werden.

5.1. Aufbau neuronaler Netze

Fig. 1 zeigt den Aufbau eines neuronalen Netzes, das nach Art eines Vielschicht-Perzeptrons aufgebaut ist. In diesem Fall umfaßt das neuronale Netz:

- eine Eingangsschicht mit einer Mehrzahl von Eingangs-Neuronen N_i (i für "input neuron"),
- wenigstens eine Zwischenschicht mit einer Mehrzahl von Zwischen-Neuronen N_h (h für "hidden neuron"),
- eine Ausgangsschicht mit einer Mehrzahl von Ausgangs-Neuronen N_o (o für "output neuron"), und
- eine Vielzahl von Synapsen, die jeweils zwei Neuronen unterschiedlicher Schichten miteinander verbinden.

In der vereinfachten Ausführungsform gemäß Fig. 1, die aus Anschaulichkeitsgründen der nachfolgenden Diskussion zugrundegelegt werden soll, ist lediglich eine einzige Zwischenschicht vorgesehen, und sind die Neuronen (bzw. Knoten, wie sie häufig auch genannt werden) der Ausgangsschicht sowohl mit jedem Neuron der Eingangsschicht als auch mit jedem Neuron der Zwischenschicht über Synapsen (auch als "Konnektoren" bezeichnet) verbunden.

Die Anzahl der Eingangs-Neuronen wird üblicherweise in Abhängigkeit der Anzahl der zur Verfügung stehenden objektivierbaren Angaben gewählt.

Sollte hierdurch jedoch die zur Ermittlung der Reaktion des neuronalen Netzes benötigte Zeit in einem nicht zu akzeptierenden Maße ansteigen, so kann beispielsweise mit Hilfe von neuronalen Netzen stark vereinfachter Struktur vorab eine Abschätzung der Bedeutung der einzelnen objektivierbaren Angaben auf die Aussagekraft des Gesamtsystems getroffen werden.

Es sei jedoch betont, daß auch diese vorausgehende Abschätzung automatisch und ohne Eingriff der jeweiligen Bedienungsperson erfolgt. Ferner wird die Anzahl der Ausgangs-Neuronen so groß gewählt, daß im Sinne

- 9 -

einer Reihenentwicklung der Überlebensfunktion eine ausreichende Anzahl von Reihenentwicklungs-Termen zur Verfügung steht, um eine aussagekräftige Annäherung an die tatsächliche Überlebensfunktion erzielen zu können. Die Anzahl der Zwischen-Neuronen schließlich wird einerseits so groß gewählt, daß die Ergebnisse des trainierten neuronalen Netzes
5 aussagekräftig sind, und andererseits so klein, daß die zur Ermittlung des Ergebnisses benötigte Zeit akzeptabel ist.

10 5.2. Funktion neuronaler Netze

5.2.1. Allgemeines

Jedes Neuron empfängt ein Stimulationssignal S , verarbeitet dieses gemäß
15 einer vorbestimmten Aktivierungsfunktion $F(S)$ und gibt ein entsprechendes Antwortsignal $A = F(S)$ aus, das allen unter diesem Neuron angeordneten Neuronen zugeführt wird. Das auf das betrachtete Neuron N_y einwirkende Stimulationssignal S_y wird üblicherweise durch Aufsummieren der Antwortsignale A_x der über diesem Neuron N_y angeordneten Neuronen N_x gebildet,
20 wobei die Beiträge der einzelnen Neuronen N_x jeweils mit einem Gewichtungsfaktor w_{xy} in die Summe eingehen, der die Stärke der die beiden Neuronen verbindenden Synapse angibt.

$$\text{Stimulationssignal:} \quad S_y = \sum_x w_{xy} \cdot A_x$$

25

$$\text{Antwortsignal:} \quad A_y = F(S_y)$$

5.2.2. Eingangsschicht

30

Die Stimulationssignale S_i der Eingangsneuronen N_i werden von den einen bestimmten Patienten j betreffenden Eingabedaten $x_{i,j}$ gebildet.

- 10 -

Stimulationssignal: $S_i = x_{i,j}$

Um die Gewichte der Synapsen eines neuronalen Netzes in geeigneter Weise interpretieren zu können, ist es bevorzugt, mit Variablen zu arbeiten, deren Werte in der Größenordnung von 1 liegen. Um dies trotz der üblicherweise recht unterschiedlichen Verteilungen der Eingabedaten erreichen zu können, ist es üblich, die Eingabedaten einer geeigneten Transformation zu unterziehen. Diese Transformation wird von der Aktivierungsfunktion F_i der Eingangsneuronen vorgenommen:

Antwortsignal: $A_i = \tanh[(S_i - S_{i,\text{mittel}})/S_{i,\alpha}]$

Für die Eingabedaten $x_{i,j}$ wird also zum einen der Mittelwert $S_{i,\text{mittel}}$ der der Trainingsdatenmenge angehörenden Patienten j gebildet. Und zum anderen wird ein Skalierungsfaktor $S_{i,\alpha}$ gebildet. Liegt der Wert einer Eingabedatenvariablen $x_{i,j}$ oberhalb des Mittelwerts $S_{i,\text{mittel}}$, so erfolgt die Skalierung gemäß dem 75%-Quartil. Liegt er hingegen unter dem Mittelwert, so erfolgt die Skalierung gemäß dem 25%-Quartil. Durch Einsatz der hyperbolischen Tangens-Funktion als Aktivierungsfunktion F_i schließlich erhält man in einfacher Weise normierte Antwortsignale, deren Werte im Bereich von -1 bis +1 liegen.

Festzuhalten ist, daß man für Eingangsdaten, die bereits die gewünschte Verteilung, kategorische oder binäre Werte aufweisen, die vorstehende Transformation umgehen kann.

5.2.3. Zwischenschicht

Das Stimulationssignal S_n für die Neuronen N_n der Zwischenschicht wird von der gewichteten Summe der Antwortsignale A_i aller Neuronen N_i der Eingangsschicht gebildet:

- 11 -

Stimulationssignal:
$$S_h = \sum_i w_{ih} \cdot A_i$$

Dieses Stimulationssignal S_h wird von den Neuronen N_h gemäß einer vorgegebenen Aktivierungsfunktion F_h , die beispielsweise wiederum die hyperbolische Tangens-Funktion sein kann, in ein Antwortsignal A_h umgesetzt:

Antwortsignal:
$$A_h = F_h(S_h - b_h)$$

Die Parameter b_h werden auf dem Fachgebiet der neuronalen Netze als Bias oder "Vorspannung" des betreffenden Neurons bezeichnet. Wie die Werte der Synapsen-Gewichte w_{xy} werden auch die Werte dieser Vorspannungsparameter b_h im Zuge des Trainings des neuronalen Netzes ermittelt.

5.2.4. Ausgangsschicht

Die Ermittlung des Stimulationssignals S_o und des Antwortsignals A_o für ein Neuron N_o der Ausgangsschicht erfolgt in analoger Weise:

Stimulationssignal:
$$S_o = \sum_i w_{io} \cdot (A_i - c_i) + \sum_h w_{ho} \cdot A_h$$

Antwortsignal:
$$A_o = F_o(S_o - b_o)$$

Die Parameter b_o geben wiederum die "Vorspannung" der Neuronen N_o der Ausgangsschicht an, während die Parameter c_i der Anpassung der Stimulationsbeiträge der Neuronen N_i der Eingangsschicht und N_h der Zwischenschicht dienen. Die Werte sowohl der Parameter b_o als auch der Parameter c_i werden in der Trainingsphase des neuronalen Netzes ermittelt. Hinsichtlich der Vorspannungen b_o kann es dabei vorteilhaft sein, als Randbedingung zu fordern, daß die über die gesamte Trainingsdatenmenge gemittelte Antwort aller Ausgangsneuronen N_o Null ist. Als Aktivierungsfunktion F_o kann für die meisten Anwendungen, insbesondere den hier betrachteten Fall

- 12 -

der Ermittlung der Überlebensfunktion für Krebspatienten, die Identitätsfunktion $F_0(x) = x$ verwendet werden.

Die Antwortsignale A_0 der Ausgangs-Neuronen N_0 geben den jeweiligen
5 Koeffizienten des zugehörigen Terms der Reihenentwicklung der gesuchten Überlebensfunktion an.

5.3. Die Überlebensfunktion

10

Wie vorstehend bereits angesprochen wurde, umfassen die Eingangsdaten Angaben über die Person des Patienten, sowie Angaben über das Krankheitsbild. Alle diese Daten werden zu einem Zeitpunkt $t = 0$ erfaßt, im Falle von Krebspatienten beispielsweise dem Zeitpunkt der Erstoperation. Im
15 Anschluß an die Erstoperation werden die Patienten dann einer bestimmten Nachbehandlung unterzogen, die beispielsweise Chemo- oder/und Strahlentherapien beinhalten kann.

20

Die Überlebensfunktion $S(t)$ gibt nun die Wahrscheinlichkeit an, daß bei einem betrachteten Patienten zum Zeitpunkt t ein bestimmtes Ereignis noch nicht eingetreten ist. Es kann sich bei diesem bestimmten Ereignis beispielsweise um eine erneute Krebserkrankung, im schlimmsten Fall aber auch um den Tod des Patienten handeln. In jedem Fall gilt für die Überlebensfunktion, daß $S(0) = 1$. Darüber hinaus wird üblicherweise $S(\infty) = 1$ angenommen.

25

Gemäß der üblichen Notation kann man auf Grundlage der Überlebensfunktion $S(t)$ eine Ereignisdichte $f(t)$ und eine Risikofunktion $\lambda(t)$ definieren:

$$f(t) = -dS/dt$$

30

$$\lambda(t) = f(t)/S(t)$$

woraus folgt:

$$\lambda(t) = -(d/dt)[\ln S(t)]$$

- 5 Kennt man also den Verlauf der Risikofunktion $\lambda(t)$, so kann man den Verlauf der Überlebensfunktion $S(t)$ durch Integration rekonstruieren.

Es ist die Aufgabe des neuronalen Netzes, den Verlauf der Risikofunktion $\lambda(t)$ nach Art einer Reihenentwicklung zu modellieren:

10

$$\lambda(t) = \lambda_0 \cdot \exp\left[\sum_0 B_0(t) \cdot A_0\right]$$

Dabei bezeichnen die Parameter A_0 gemäß vorstehender Notation die Antwortsignale der Neuronen N_0 der Ausgangsschicht des neuronalen
 15 Netzes. λ_0 ist im Rahmen der vorliegenden Erfindung ein von t unabhängiger Parameter, der als Normierungsfaktor eingesetzt wird. Mit $B_0(t)$ ist ein Satz von Funktionen bezeichnet, die als Basisfunktionen der Reihenentwicklung eine gute Annäherung an den tatsächlichen Verlauf der Risikofunktion ermöglichen. Als Funktionensatz $B_0(t)$ können beispielsweise die fraktalen
 20 Polynome oder auch Funktionen wie t^p (p nicht notwendigerweise ganzzahlig) eingesetzt werden. Im Rahmen der vorliegenden Erfindung wurden $B_{01}(t) = 1$; $B_{02}(t) = \text{const} \cdot t^{1/2}$, ... verwendet.

25 5.4. Training des neuronalen Netzes - Vorbereitungen

5.4.1. Die Optimierungsfunktion

Die Trainingsdatenmenge umfaßt die Datensätze einer Mehrzahl von
 30 Patienten, für die nicht nur die Angaben über die Person und die Angaben über das Krankheitsbild bekannt sind, sondern auch Angaben über die Art der Nachbehandlung und der weitere Krankheitsverlauf. Aus den gesammel-

ten Daten über den weiteren Krankheitsverlauf wird nach den folgenden Regeln eine "tatsächliche Überlebensfunktion" zusammengesetzt: Ist für einen bestimmten Patienten zu einem Zeitpunkt t das vorbestimmte Ereignis, beispielsweise eine erneute Erkrankung oder der Tod des Patienten bereits eingetreten, so wird sein Beitrag δ zur "tatsächlichen Überlebensfunktion" vor dem Zeitpunkt t zu $\delta = 0$ und ab dem Zeitpunkt t zu $\delta = 1$ gesetzt. Patienten, bei denen bis zum Zeitpunkt der Erhebung der Trainingsdatensmenge das vorbestimmte Ereignis noch nicht eingetreten ist ("zensierte" Daten), tragen zu jedem Zeitpunkt zur "tatsächlichen Überlebensfunktion" nur $\delta = 0$ bei. Während der Trainingsphase werden nun die Gewichte w_{xy} der Synapsen und die sonstigen vorstehend unter Punkt 5.2. genannten Optimierungsparameter derart eingestellt, daß die vom neuronalen Netz gelieferte Überlebensfunktion möglichst gut der "tatsächlichen Überlebensfunktion" entspricht.

Dies kann beispielsweise dadurch erzielt werden, daß man eine für diesen Zweck geeignete Optimierungsfunktion O definiert und ein lokales, im günstigsten Fall sogar das globale Minimum dieser Optimierungsfunktion in dem von den Optimierungsparametern aufgespannten Raum sucht. Es ist im Stand der Technik bereits bekannt, zur Definition der Optimierungsfunktion O von einer sogenannten Likelihood-Funktion L auszugehen:

$$O = -\ln L$$

Erfindungsgemäß wird als Darstellung der Likelihood-Funktion

$$L = \prod_j [f_j(t)]^\delta \cdot [S_j(t)]^{1-\delta}$$

gewählt, wobei $f_j(t)$ und $S_j(t)$ gemäß der unter Punkt 5.3. eingeführten Notation die Ereignisdichte und die Überlebensfunktion für den Patienten j der Trainingsmenge bezeichnen. Diese Likelihood-Funktion hat den Vorteil,

- 15 -

daß der Rechenaufwand nur etwa proportional zur Anzahl der in der Trainingsdatenmenge erfaßten Patienten ansteigt.

Eine weitere Möglichkeit zur Darstellung der Likelihood-Funktion ist:

5

$$L = \prod_j \left(\exp[\sum_o B_o(t) \cdot A_{oj}] / \sum_l \exp[\sum_o B_o(t) \cdot A_{ol}] \right)$$

wobei das Produkt über alle Patienten j gebildet wird, bei denen das vorbestimmte Ereignis zum Zeitpunkt t bereits aufgetreten ist, und wobei die
10 erste Summe im Nenner des Quotienten über alle Patienten l gebildet wird, bei denen das vorbestimmte Ereignis zum Zeitpunkt t noch nicht aufgetreten ist.

Der mit dieser Darstellung verbundene Rechenaufwand steigt allerdings
15 etwa proportional zum Quadrat der Anzahl der in der Trainingsdatenmenge erfaßten Patienten an.

5.4.2. Die Initialisierung

20

Zur Initialisierung der Netzoptimierungs-Parameter, beispielsweise der Gewichte der die Neuronen verbindenden Synapsen, können diesen Parametern, wie dies im Stand der Technik an sich bekannt ist, stochastisch kleine Werte zugewiesen werden, die gewissen Normierungsregeln gehorchen. Dabei ist es zusätzlich möglich, in die Normierung Erkenntnisse
25 einfließen zu lassen, die vorab in Testläufen an neuronalen Netzen stark vereinfachter Struktur gewonnen wurden.

5.5. Training des neuronalen Netzes - Simplex-Methode

Die Suche nach einem lokalen bzw. dem globalen Minimum der Optimierungsfunktion wird wie üblich in mehreren Schritten bzw. Zyklen durchgeführt. Erfindungsgemäß wird jedoch zu dieser Suche erstmals die von Nelder und Mead (siehe Punkt 6. "Literaturverzeichnis") vorgeschlagene Simplex-Methode bei einem neuronalen Netz eingesetzt. Unter einem Simplex wird in einem n -dimensionalen Raum ein $(n + 1)$ -dimensionales Gebilde verstanden, das den aktuellen Stützpunkt in dem n -dimensionalen Raum umgibt, also in einem 2-dimensionalen Raum ein Dreieck, in einem 3-dimensionalen Raum eine Dreieckspyramide und so weiter. In welchen Richtungen und mit welchen Abständen vom aktuellen Stützpunkt die $(n + 1)$ Eckpunkte angeordnet werden, wird dabei auf Grundlage der Eigenschaften der Optimierungsfunktion an den Eckpunkten des vorhergehenden Zyklus bestimmt.

Dieses Verfahren führt zu einer streng monoton abfallenden Folge von Stützpunkten. Es kann fortgesetzt werden, bis entweder (innerhalb vorgegebener Genauigkeitsgrenzen) ein lokales oder globales Minimum aufgefunden oder ein weiteres Abbruchkriterium erfüllt worden ist. Im Zusammenhang mit diesem weiteren Abbruchkriterium kommt nun die vorstehend bereits angesprochene Validierungs-Datenmenge ins Spiel:

Der vorstehend angesprochene monotone Abfall der Stützpunkte kann zum einen von tatsächlich objektivierbaren Eigenschaften der für die Trainings-Datenmenge bestimmten Optimierungsfunktion herrühren. Zum anderen ist es jedoch auch möglich, daß der abfallende Verlauf im Bereich einer von stochastischen Fluktuationen hervorgerufenen Senke der Optimierungsfunktion erfolgt. Der letztgenannte Effekt täuscht aber einen Lernerfolg nur vor. Daher werden erfindungsgemäß auch die Eigenschaften der anhand der Validierungs-Datenmenge bestimmten Optimierungsfunktion an den gleichen Stützpunkten untersucht. Stellt man dabei fest, daß auch die Stützpunkte

der "Validierungsdaten-Optimierungsfunktion" monoton abfallen, so kann man davon ausgehen, daß man sich noch in einer "echten" Lernphase des neuronalen Netzes befindet. Stagniert die Folge der Stützpunkte der "Validierungsdaten-Optimierungsfunktion" hingegen oder steigt sie sogar wieder an, so muß davon ausgegangen werden, daß man sich bezüglich der "Trainingsdaten-Optimierungsfunktion" in einer von stochastischen Fluktuationen hervorgerufenen Senke befindet, die einen Lernfortschritt nur vorgaukelt. Die zyklische Durchführung der Simplex-Methode kann daher unterbrochen werden.

Der Vorteil der Simplex-Methode ist hauptsächlich darin zu sehen, daß sie allein anhand der Optimierungsfunktion durchgeführt werden kann und zudem die Schrittweite und Schrittrichtung automatisch festgelegt werden.

5.6. Training des neuronalen Netzes - Strukturvereinfachung ("Pruning")

Nachdem die Suche nach einem lokalen bzw. dem globalen Minimum abgeschlossen ist, wird gemäß einem nächsten Trainingsschritt untersucht, ob sich die Struktur des neuronalen Netzes aufgrund der bislang gewonnenen Erkenntnisse nicht vereinfachen läßt. Bei diesem "Pruning" geht es also um die Untersuchung der Frage, welche der Synapsen so wenig Einfluß auf die Gesamtfunktion des neuronalen Netzes haben, daß auf sie verzichtet werden kann. Dies kann im einfachsten Fall beispielsweise dadurch erfolgen, daß das ihnen zugewiesene Gewicht ein für allemal auf Null gesetzt wird. Grundsätzlich ist es jedoch ebenso denkbar, das Gewicht der betreffenden Synapse auf einen festen Wert "einzufrieren". Vorteilhafterweise sollten sich Simplexoptimierungs-Schritte und Strukturvereinfachungs-Schritte in einem iterativen Verfahren abwechseln. Selbstverständlich wäre es wünschenswert, das neuronale Netz bereits nach dem Ausschluß einer einzigen Synapse einer erneuten Simplexoptimierung zu unterziehen. Dies ist jedoch im Hinblick auf die für das Training insgesamt erforderliche Zeit

nicht vertretbar. In der Praxis hat es sich als günstiger Kompromiß erwiesen, während eines Strukturvereinfachungs-Schritts allenfalls 10% der zu Beginn dieses Schritts noch vorhandenen Synapsen zu entfernen.

5 Zur Strukturvereinfachung kommen erfindungsgemäß die beiden nachfolgend unter den Punkten 5.6.1. und 5.6.2. beschriebenen Verfahren zum Einsatz.

10 5.6.1. Likelihood-Methode

Im Rahmen dieser Methode wird als Referenzwert zunächst der Wert der Likelihood-Funktion auf Grundlage der gesamten Struktur des neuronalen Netzes in seinem gegenwärtigen Trainingszustand berechnet, d.h. unter
15 Verwendung der aktuellen Werte der Gewichte aller Synapsen. Anschließend wird der Einfluß einer vorgegebenen Synapse unterdrückt, d.h. der Wert des Gewichts dieser Synapse wird auf Null gesetzt. Sodann wird der Wert der Likelihood-Funktion für die so vereinfachte Netzstruktur berechnet, und das Verhältnis dieses Werts zum Referenzwert gebildet.

20

Hat man dieses Likelihood-Verhältnis für alle Synapsen berechnet, so beginnt man bei der Durchführung der nachfolgend beschriebenen Schritte bei derjenigen Synapse, für die der Wert des Likelihood-Verhältnisses am nächsten bei Eins liegt:

25

Nimmt man an, daß die Netzstruktur bereits um $(x-1)$ Synapsen vereinfacht worden ist und nunmehr die Bedeutung der x . Synapse untersucht wird, so vergleicht man die folgenden drei Netzstruktur-Varianten: Erstens Gesamtstruktur des neuronalen Netzes in seinem gegenwärtigen Trainingszustand
30 mit allen vor diesem Strukturvereinfachungs-Schritt noch vorhandenen Synapsen, zweitens Netzstruktur unter Ausschluß der $(x-1)$ in diesem Strukturvereinfachungs-Schritt bereits unterdrückten Synapsen, und drittens

- 19 -

Netzstruktur unter Ausschluß nunmehr auch der x . Synapse. Anschließend vergleicht man die dritte Strukturvariante mittels eines Signifikanz-Tests zum einen mit der ersten Strukturvariante (volle Struktur) und zum anderen mit der zweiten Strukturvariante ($(x-1)$ Synapsen unterdrückt). Wenn auch nur
5 einer der beiden Tests eine zu starke Abweichung der dritten Strukturvariante ergibt, so wird die entsprechende Synapse zumindest für den nächstfolgenden Simplexoptimierungs-Schritt beibehalten.

Als Signifikanz-Test kann beispielsweise der an sich bekannte CHI-
10 QUADRAT-Test (vgl. Punkt 6. "Literaturverzeichnis", Dokument)
eingesetzt werden. Alternativ könnte dieser Signifikanz-Test auch mittels der ebenfalls an sich bekannten BOOT-STRAPPING-Methode (vgl. Punkt 6. "Literaturverzeichnis", Dokument)
15 durchgeführt werden. Die Verwendung des CHI-QUADRAT-Tests ist besonders dann günstig, wenn man die Reaktion des neuronalen Netzes auf Grundlage einer Likelihood-Funktion ermittelt. Die BOOT-STRAPPING-Methode eignet sich auch bei anderen Arten von Funktionen zur Darstellung der Reaktion des neuronalen Netzes.

20

5.6.2. Korrelations-Methode

Der Ausschluß bzw. die Unterdrückung von Synapsen nach der Korrelations-Methode beruht auf der Überlegung, daß es möglich sein könnte, daß
25 zwei in ein und derselben Schicht angeordnete Neuronen auf ein Neuron einer darunter angeordneten Schicht qualitativ den gleichen Einfluß ausüben. In diesem Fall sollte sich die Reaktion des neuronalen Netzes, genauer gesagt das Antwortsignal dieses letztgenannten Neurons, im wesentlichen nicht ändern, wenn man dieses Neuron lediglich von einem der
30 beiden darüber angeordneten Neuronen stimulieren läßt und den Einfluß des zweiten Neurons durch eine Stärkung der verbleibenden Synapse berück-

- 20 -

sichtig. Auf die von dem zweiten Neuron zu dem betrachteten Neuron führende Synapse könnte dann verzichtet werden.

a. **Eingangs-Neuronen und Ausgangs-Neuronen verbindende Synapsen**

5

Gemäß Punkt 5.2.4. hat der Beitrag des Antwortsignals zweier Eingangs-Neuronen zum Stimulationssignal eines Ausgangs-Neurons die Form:

$$S_o = w_{1o} \cdot (A_1 - c_1) + w_{2o} \cdot (A_2 - c_2)$$

10

Unterstellt man nun, daß die beiden Antwortsignale A_1 und A_2 gemäß

$$A_2 = m \cdot A_1 + n$$

15 zumindest näherungsweise miteinander korreliert sind und daß das Gewicht w_{1o} größer ist als das Gewicht w_{2o} , so gilt für das Stimulationssignal S_o :

$$S_o = (w_{1o} + w_{2o} \cdot m) \cdot A_1 + (n \cdot w_{2o} - w_{1o} \cdot c_1 - w_{2o} \cdot c_2)$$

20

$$= w_{1o}^* \cdot (A_1 - c_1^*)$$

mit

$$w_{1o}^* = w_{1o} + w_{2o} \cdot m$$

25 und

$$c_1^* = -[(n \cdot w_{2o} - w_{1o} \cdot c_1 - w_{2o} \cdot c_2)] / (w_{1o} + w_{2o} \cdot m)$$

Ist w_{1o}^* nicht klein, so kann man das Verhalten des neuronalen Netzes unter den folgenden Annahmen testen:

30

- 21 -

1. Ersetze das Gewicht w_{1o} durch w_{1o}^* ;
2. Ersetze den Parameter c_1 durch c_1^* ; und
3. Unterdrücke die Synapse vom Eingangs-Neuron N_2 zum Ausgangs-Neuron N_o .

5

Verläuft dieser Test, der beispielsweise wieder als CHI-QUADRAT-Test durchgeführt werden kann, positiv, so kann auf die Synapse vom Eingangs-Neuron N_2 zum Ausgangs-Neuron N_o verzichtet werden.

10

b. Eingangs-Neuronen und Zwischen-Neuronen verbindende Synapsen

In analoger Weise kann man auch den Beitrag des Antwortsignals zweier Eingangs-Neuronen zum Stimulationssignal eines Zwischen-Neurons betrachten, wobei es sich, wie nachstehend sofort klar werden wird, empfiehlt, das Stimulationssignal des Zwischenneurons einschließlich dessen "Vorspannung" zu betrachten:

$$S_h - b_h = w_{1h} \cdot A_1 + w_{2h} \cdot A_2$$

20

Unterstellt man nun wiederum, daß die beiden Antwortsignale A_1 und A_2 gemäß

$$A_2 = m \cdot A_1 + n$$

25

zumindest näherungsweise miteinander korreliert sind und daß das Gewicht w_{1h} größer ist als das Gewicht w_{2h} , so gilt für das Stimulationssignal S_h :

$$S_h - b_h = (w_{1h} + w_{2h} \cdot m) \cdot A_1 + n \cdot w_{2h}$$

30

oder

- 22 -

$$S_h - b_h^* = w_{1h}^* \cdot A_1$$

mit

$$w_{1h}^* = w_{1h} + w_{2h} \cdot m$$

5

und

$$b_h^* = b_h + n \cdot w_{2h}$$

10 Ist w_{1h}^* nicht klein, so kann man das Verhalten des neuronalen Netzes unter den folgenden Annahmen testen:

1. Ersetze das Gewicht w_{1h} durch w_{1h}^* ;
2. Ersetze die Vorspannung b_h durch b_h^* ; und
3. Unterdrücke die Synapse vom Eingangs-Neuron N_2 zum Zwischen-
15 Neuron N_h .

Verläuft dieser Test, der beispielsweise wieder als CHI-QUADRAT-Test durchgeführt werden kann, positiv, so kann auf die Synapse vom Eingangsneuron N_2 zum Zwischenneuron N_h verzichtet werden.

20

c. Zwischen-Neuronen und Ausgangs-Neuronen verbindende Synapsen

25 In analoger Weise können auch von Zwischen-Neuronen zu Ausgangs-Neuronen führende Synapsen behandelt werden. Hinsichtlich der Vorspannungen b_o ist jedoch möglicherweise die unter Punkt 5.2.4. angesprochene weitere Randbedingung zu beachten.

5.6.3. Überprüfung der Topologie

Das vorstehend beschriebene Ausdünnen der Struktur des neuronalen Netzes kann zur Folge haben, daß einzelne Neuronen mit keinem einzigen
5 anderen Neuron mehr in Verbindung stehen. Dies ist beispielsweise der Fall, wenn ein Eingangs-Neuron mit keinem Zwischen-Neuron und auch keinem Ausgangs-Neuron in Verbindung steht, oder wenn ein Ausgangs-Neuron mit keinem Zwischen-Neuron und auch keinem Eingangs-Neuron in Verbindung steht. Es ist daher nur folgerichtig, wenn diese Neuronen, die keinen Einfluß
10 auf die Funktion des neuronalen Netzes mehr haben, vollständig deaktiviert werden.

Einen Spezialfall bilden Zwischen-Neuronen, die zwar noch mit Neuronen der Eingangsschicht in Verbindung stehen, nicht jedoch mit Neuronen der
15 Ausgangsschicht. Diese Zwischen-Neuronen können auf die Funktion des neuronalen Netzes keinerlei Einfluß mehr ausüben. Daher können auch die von der Eingangsschicht zu diesen Zwischen-Neuronen führenden Synapsen unterdrückt werden, d.h. die Gewichte dieser Synapsen auf Null gesetzt werden.

20 Es kann aber auch der umgekehrte Fall auftreten, nämlich daß ein Zwischen-Neuron zwar noch mit der Ausgangsschicht verbunden ist, jedoch keinerlei Verbindung mehr mit der Eingangsschicht hat. Diese Zwischen-Neuronen können allenfalls ein von ihrer "Vorspannung" abhängendes Antwortsignal
25 an die Ausgangs-Neuronen abgeben. Ein derartiges Signal hat aber keinerlei Informationsgehalt, der für die Funktion des neuronalen Netzes von Bedeutung wäre. Somit können auch die restlichen Synapsen dieser Zwischen-Neuronen unterdrückt werden.

5.7. Generalisierung

Nach Abschluß der Trainingsphase ist es erforderlich, die Leistungsfähigkeit des trainierten neuronalen Netzes zu überprüfen, um ein Maß dafür zu bekommen, welche Aussagekraft die von diesem neuronalen Netz gelieferten Überlebensfunktionen tatsächlich haben. Hierzu wird die vorstehend bereits angesprochene Generalisierungs-Datenmenge eingesetzt, die auf das Training des neuronalen Netzes keinerlei Einfluß genommen hat und somit eine objektive Aussage ermöglicht.

5.8. Abschließende Bemerkungen

Abschließend sei erwähnt, daß neben den vorstehend explizit angesprochenen tumor-spezifischen Faktoren upA und PAI-1, die Aussagen über die Invasion zulassen, auch weitere derartige Faktoren berücksichtigt werden können. Es sind dies unter anderem Faktoren für die Proliferation, beispielsweise die S-Phase und Ki-67, sowie für weitere das Tumorwachstum beeinflussende Prozesse.

6. LITERATURVERZEICHNIS

Zum Stand der Technik wird auf die folgenden Druckschriften verwiesen:

1. Barnhill S, Zhang Z, US 5,769,074, 'Computer assisted methods for diagnosing diseases'.
2. Bellotti M, et al. (1997), 'Neural networks as a prognostic tool for patients with non-small cell carcinoma of the lung', Mod Pathol. Dec, 10(12), 1221.

- 25 -

3. Biganzoli, E. Boracchi, P. Mariani, L., Marubini, E. (1998), 'Feed Forward Neural Networks for the Analysis of Censored Survival Data: A Partial Logistic Regression Approach', *Statistics in Medicine*, 17, 1169.
- 5
4. Bostwick DG (1998), 'Practical clinical application of predictive factors in prostate cancer: A review with an emphasis on quantitative methods in tissue specimens', *Anal Quant Cytol Histol.* Oct, 20(5), 323-42. Review.
- 10
5. Bottaci L, et al. (1997), 'Artificial neural networks applied to outcome prediction for colorectal cancer patients in separate institutions', *Lancet.* Aug 16, 350, 469.
- 15
6. Bryce TJ, et al(1998), 'Artificial neural network model of survival in patients treated with irradiation with and without concurrent chemotherapy for advanced carcinoma of the head and neck', *Int J Radiat Oncol Biol Phys.* May 1, 41(2), 339.
- 20
7. Burke HB, et al. (1997), 'Artificial neural networks improve the accuracy of cancer survival prediction', *Cancer.* Feb 15, 79(4), 857.
8. Cox, D. (1972) 'Regression Models And Life Tables', *J R Stat Soc [B]* 34, 187.
- 25
9. Comanor L, Minor J, US 5,860,917, 'Method and apparatus for predicting therapeutic outcomes'.
- 30
10. De Laurentiis, M. & Ravdin, P. (1994), 'Survival Analysis of Censored Data, Neural Network Analysis Detection of Complex Interactions Between Variables', *Breast Cancer Res Tr* 32, 113.

- 26 -

11. Ebell, M. (1993), 'Artificial Neural Network Model for Predicting Failure to Survive Following In-Hospital Cardiopulmonary Resuscitation', *The Journal of Family Practice*, 36, 297.
- 5 12. Faraggi, D. & Simon, R. (1995), 'A Neural Network Model for Survival Data', *Statistics in Medicine*, 14, 73.
13. Faraggi, D. Simon, R., Yaskil, E., Kramar A., (1997), 'Bayesian Neural Network Models for Censored Data', *Biometrical Journal*, 39,
10 519.
14. Fogel, D., Wasson, E., & Boughton, E. (1995), 'Evolving Neural Networks for Detecting Breast Cancer', *Cancer Letters*, 96, 49.
- 15 15. Gray, R. (1992) 'Flexible Methods for Analyzing Survival Data Using Splines, with Applications to Breast Cancer Prognosis', *J. American Statistical Association*, 87, 942.
16. Hamamoto I, et al. (1995), 'Prediction of the early prognosis of the
20 hepatectomized patient with hepatocellular carcinoma with a neural network', *Comput Biol Med.* Jan, 25(1), 49.
17. Hastie, T., Sleeper, L. & Tibshirani, R. (1992) 'Flexible Covariate Effects in the Proportional Hazards Model', *Breast Cancer Res Tr* 22,
25 241.
18. Hilsenbeck, S., & Clark, G. (1996), 'Practical p-Value Adjustment for Optimally Selected Cutpoints', *Statistics in Medicine*, 15, 103.
- 30 19. Hilsenbeck, S., Clark, G., & McGuire, W. (1992), 'Why Do So Many Prognostic Factors Fail to Pan Out?', *Breast Cancer Res Tr* 22, 197.

- 27 -

20. Jefferson MF, et al. (1997), 'Comparison of a genetic algorithm neural network with logistic regression for predicting outcome after surgery for patients with nonsmall cell lung carcinoma', *Cancer*. Apr 1, 79(7), 1338.
- 5
21. Kalbfleisch, J. & Prentice, R. (1980), *The Statistical Analysis of Failure Time Data*, Wiley.
22. Kappen HJ, et al. (1993), 'Advanced ovarian cancer: Neural network analysis to predict treatment outcome', *Ann Oncol.*, 4 Suppl 4, 31.
- 10
23. Kharchenko EP. (1996), '[The use of a neural network model for cancer prognosis on the basis of immunological indices]', *Biull Eksp Biol Med*. Aug, 122(8), 206. Russian.
- 15
24. Knorr, K., Hilsenbeck, S., Wenger, C. et al., (1992) 'Making the Most of Your Prognostic Factors, Presenting a More Accurate Survival Model for Breast Cancer Patients', *Breast Cancer Res Tr* 22, 251.
- 20
25. Koutsoukos AD, et al. (1994), 'Discrimination techniques applied to the NCI in vitro anti-tumour drug screen: predicting biochemical mechanism of action', *Stat Med*. Mar 15-Apr 15, 13(5-7), 719.
26. Le Cun, Y., Denker, J., Solla, S. (1990), 'Optimal Brain Damage', in *Advances in Neural Information Processing Systems 2*, (Hrsg. D. Touretzky), San Mateo
- 25
27. Liestol, K., Anderson, P. & Anderson, U. (1994), 'Survival Analysis and Neural Nets', *Statistics in Medicine*, 13, 1189.
- 30
28. Mariani L, et al. (1997), 'Prognostic factors for metachronous contralateral breast cancer, a comparison of the linear Cox regression

- 28 -

model and its artificial neural network extension', *Breast Cancer Res Treat.* Jun, 44(2), 167.

- 5 29. Marsh JW, et al. (1998), 'Liver transplantation in the treatment of hepatocellular carcinoma', *J Hepatobiliary Pancreat Surg.*, 5(1), 24.
30. Marsh JW, et al. (1997), 'The prediction of risk of recurrence and time to recurrence of hepatocellular carcinoma after orthotopic liver transplantation, a pilot study. *Hepatology.* Aug, 26(2), 444.
- 10 31. McGuire WL, et al. (1992), 'Treatment decisions in axillary node-negative breast cancer patients', *J Natl Cancer Inst Monogr.*, (11), 173.
- 15 32. McGuire, W., Tandon, A., Allred, D. Chamness, G. & Clark, G. (1990), 'How To Use Prognostic Factors in Axillary Node-Negative Breast Cancer Patients ', *J. Natl Canc Inst* 82, 1006.
- 20 33. Naguib RN, et al. (1998), 'Neural network analysis of combined conventional and experimental prognostic markers in prostate cancer: a pilot study' *Br J Cancer.* Jul, 78(2), 246.
- 25 34. Naguib RN, et al. (1997), 'Prediction of nodal metastasis and prognosis in breast cancer: a neural model', *Anticancer Res.* Jul-Aug, 17(4A), 2735.
35. Nelder, J.A., & Mead, R. (1965), ' A simplex method for function minimization', *Computer Journal* 7, 308.
- 30 36. Nowlan, S. J. (1991), *Soft Competitive Adaption, Neural Network Learning Algorithms based on Fitting Statistical Mixtures*, PhD Thesis, School of Comp. Sc., Carnegie Mellon, Pittsburgh.

37. Press, W., Teukolsky, S., Flannery, B. & Vetterling, W. (1992), *Numerical Recipes, The Art of Scientific Computing*, Cambridge.
38. Ravdin, P. & Clark, G. (1992) 'A Practical Application of Neural Network Analysis for Predicting Outcome of Individual Breast Cancer Patients' , *Breast Cancer Res Tr* 22, 285.
39. Ravdin, P. Clark, G. Hilsenbeck, S. , et al. (1992), 'A Demonstration that Breast Cancer Recurrence can be Predicted by Neural Network Analysis', *Breast Cancer Res Tr* 21,47.
40. Ravdin, P. McGuire, W., Clark, G., US 5,862,304, 'Method for predicting the future occurrence of clinically occult or non-existent medical conditions'.
41. Rehkugler, H. & Zimmerman, H. (1994) *Neuronale Netze in der Ökonomie*, Vahlen Verlag.
42. Ripley, B. (1993) 'Statistical Aspects of Neural Networks', in *Networks and Chaos - Statistical and Probabilistic Aspects*, Hrsg. Barndorff-Nielsen, O. Jensen, J. & Kendall, W., Chapman & Hill.
43. Ritter, H., Martinetz, T. & Schulten, K. (1990), *Neuronale Netze, Eine Einführung in die Neuroinformatik Selbstorganisierender Netzwerke*, Addison-Wesley.
44. Royston, P. & Altman. D. (1994), 'Regression using fractional polynomials of continuous covariates, parsimonious parametric modelling', *App. Stat.* 43, 429.

- 30 -

45. Schumacher, M., Rossner, R. Vach, W., (1996), 'Neural Networks and logistic regression', *Computational Statistics & Data Analysis*, 21, 661 (Part I).
- 5 46. Snow PB, et al. (1994), 'Artificial neural networks in the diagnosis and prognosis of prostate cancer, a pilot study. *J Urol*. Nov, 152 (5 Pt 2), 1923.
47. Speight PM, et al. (1995), 'The use of artificial intelligence to identify
10 people at risk of oral cancer and precancer', *Br Dent J*. Nov 25, 179(10), 382.
48. Tibshirani, R. (1996), 'A comparison of some error estimates for neural network procedures', *Neural Computation* 8, 152.
- 15 49. Tracey KJ, et al. (1992), 'Tumor necrosis factor in metabolism of disease, hormonal actions versus local tissue effects', *Nouv Rev Fr Hematol.* , 34 Suppl, S37.
- 20 50. Tu, J. (1996) 'Advantages and Disadvantages of Using Artificial Neural Networks versus Logistic Regression for Predicting Medical Outcomes', *J. Clin. Epidemiol.* 49, 1225.
51. Warner, B. & Misra, M, (1996), 'Understanding Neural networks as
25 statistical tools,' *The American Statistician* 50, 284.
52. Willoughby TR, et al. (1996), 'Evaluation and scoring of radiotherapy treatment plans using an artificial neural network', *Int J Radiat Oncol Biol Phys*. Mar 1, 34(4), 923.

Ansprüche

1. Verfahren zum Trainieren eines neuronalen Netzes zur Ermittlung
5 von Risikofunktionen für Patienten im Anschluß an eine Ersterkrankung mit einer vorbestimmten Krankheit auf Grundlage vorgegebener Trainings-Datensätze, welche objektivierbare und meßtechnisch erfaßte Daten zum Krankheitsbild der Patienten beinhalten, wobei das neuronale Netz umfaßt:

- 10 - eine Eingangsschicht mit einer Mehrzahl von Eingangs-Neuronen,
- wenigstens eine Zwischenschicht mit einer Mehrzahl von Zwischen-Neuronen,
- eine Ausgangsschicht mit einer Mehrzahl von Ausgangs-Neuronen, und
- 15 - eine Vielzahl von Synapsen, die jeweils zwei Neuronen unterschiedlicher Schichten miteinander verbinden,

dadurch gekennzeichnet,

daß das Training des neuronalen Netzes eine Strukturvereinfachungs-
20 Prozedur umfaßt, d.h. das Auffinden und Eliminieren von Synapsen, welche auf den Verlauf der Risikofunktion keinen wesentlichen Einfluß ausüben, indem

a1) man zwei mit ein und demselben Empfangs-Neuron verbundene Sende-Neuronen auswählt,

25 a2) man unterstellt, daß die von diesen Sende-Neuronen zum Empfangs-Neuron ausgehenden Signale qualitativ im wesentlichen das gleiche Verhalten zeigen, also miteinander korreliert sind,

30 a3) man die Synapse eines der beiden Sende-Neuronen zu dem Empfangs-Neuron unterbricht und dafür das Gewicht der Synapse des jeweils anderen Sende-Neurons zu dem Empfangs-Neuron entsprechend anpaßt,

- 32 -

a4) man die Reaktion des gemäß Schritt a3) veränderten neuronalen Netzes mit der Reaktion des unveränderten neuronalen Netzes vergleicht, und

5

a5) man dann, wenn die Abweichung der Reaktion ein vorbestimmtes Maß nicht überschreitet, entscheidet, die in Schritt a3) vorgenommene Änderung beizubehalten.

2. Verfahren nach Anspruch 1,

10

dadurch gekennzeichnet, daß die beiden Sende-Neuronen in ein und derselben Schicht angeordnet sind.

3. Verfahren nach Anspruch 1 oder 2,

dadurch gekennzeichnet, daß man in Schritt a3) ferner den Wert der Vorspannung des Empfangs-Neurons anpaßt.

15

4. Verfahren zum Trainieren eines neuronalen Netzes gemäß dem Oberbegriff des Anspruchs 1 und gewünschtenfalls dem Kennzeichnen eines der Ansprüche 1 bis 3,

20

dadurch gekennzeichnet,

daß das Training des neuronalen Netzes eine Strukturvereinfachungs-Prozedur umfaßt, d.h. das Auffinden und Eliminieren von Synapsen, welche auf den Verlauf der Risikofunktion keinen wesentlichen Einfluß ausüben, indem

25

b1) man eine Synapse auswählt,

b2) man unterstellt, daß diese Synapse auf den Verlauf der Risikofunktion keinen wesentlichen Einfluß ausübt,

b3) man diese Synapse unterbricht,

b4) man die Reaktion des gemäß Schritt b3) veränderten neuronalen Netzes mit der Reaktion des unveränderten neuronalen Netzes vergleicht, und

30

- 33 -

b5) man dann, wenn die Abweichung der Reaktion ein vorbestimmtes Maß nicht überschreitet, entscheidet, die in Schritt b3) vorgenommene Änderung beizubehalten.

- 5 5. Verfahren nach Anspruch 4,
 dadurch gekennzeichnet, daß man dann, wenn im Rahmen der
 Strukturvereinfachungs-Prozedur bereits n-1 Synapsen eliminiert
 worden sind und man die Stärke des Einflusses einer n-ten Synapse
 überprüft, die Reaktion des um n Synapsen verminderten neuronalen
10 Netzes nicht nur mit der Reaktion eines um lediglich n-1 Synapsen
 verminderten Netzes vergleicht, sondern auch mit der Reaktion des
 neuronalen Netz mit seiner zu Beginn dieser Strukturvereinfachungs-
 Prozedur vorliegenden vollen Struktur, und daß man die Eliminierung
 der n-ten Synapse nur dann beibehält, wenn die Abweichung der
15 Reaktion bei beiden Vergleichen ein vorbestimmtes Maß nicht
 überschreitet.
6. Verfahren nach einem der Ansprüche 1 bis 5,
 dadurch gekennzeichnet, daß man zur Darstellung der Reaktion des
20 neuronalen Netzes den Wert einer Likelihood-Funktion für das
 neuronale Netz ermittelt.
7. Verfahren nach einem der Ansprüche 1 bis 6,
 dadurch gekennzeichnet, daß man den Vergleich von Struktur-
25 varianten des neuronalen Netzes mittels eines Signifikanz-Tests
 durchführt.
8. Verfahren nach Anspruch 7,
 dadurch gekennzeichnet, daß man den Vergleich von Struktur-
30 varianten des neuronalen Netzes mittels des an sich bekannten CHI-
 QUADRAT-Tests durchführt.

- 34 -

9. Verfahren nach Anspruch 7,
dadurch gekennzeichnet, daß man den Vergleich von Strukturvarianten des neuronalen Netzes mittels der an sich bekannten BOOT-STRAPPING-Methode durchführt.

5

10. Verfahren nach einem der Ansprüche 1 bis 8,
dadurch gekennzeichnet, daß man zum Vergleich zweier Strukturvarianten des neuronalen Netzes das Verhältnis der Werte der Likelihood-Funktionen für diese beiden Strukturvarianten ermittelt.

10

11. Verfahren zum Trainieren eines neuronalen Netzes gemäß dem Oberbegriff des Anspruchs 1 und gewünschtenfalls dem Kennzeichen eines der Ansprüche 1 bis 10,

dadurch gekennzeichnet,

15

daß das Training des neuronalen Netzes eine Optimierungs-Prozedur umfaßt, in welcher die Stärken der einzelnen Synapsen, d.h. die Stärken der Verbindungen zwischen den Neuronen, optimiert werden, und daß zu dieser Optimierung die an sich bekannte Simplex-Methode eingesetzt wird.

20

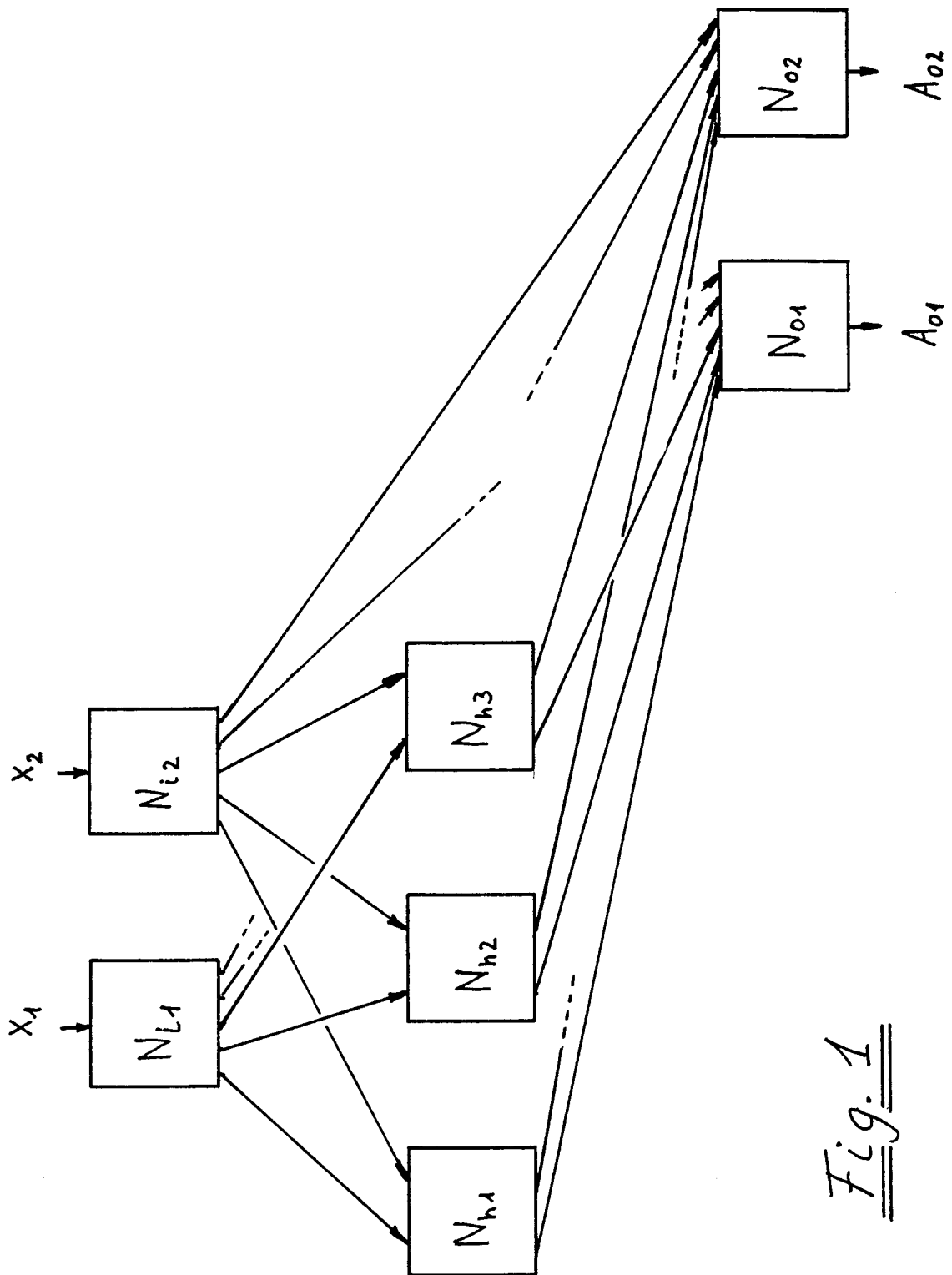


Fig. 1

专利名称(译)	训练神经网络的方法		
公开(公告)号	EP1232478A2	公开(公告)日	2002-08-21
申请号	EP2000962377	申请日	2000-08-24
[标]申请(专利权)人(译)	威丽克斯股份公		
申请(专利权)人(译)	WILEX AG		
当前申请(专利权)人(译)	WILEX AG		
[标]发明人	KATES RONALD HARBECK NADIA SCHMITT MANFRED		
发明人	KATES, RONALD HARBECK, NADIA SCHMITT, MANFRED		
IPC分类号	G06N3/08 A61B5/00 G06F19/00 G06N3/00		
CPC分类号	A61B5/00 A61B5/415 A61B5/418 A61B5/7267 G06N3/08 G16H50/70		
优先权	19940577 1999-08-26 DE		
其他公开文献	EP1232478B1		
外部链接	Espacenet		

摘要(译)

本发明方法的目的是训练神经网络，以基于预定的训练数据集确定与预定疾病相关的第一疾病相关的患者的风险功能。所述预定数据集包含可客观化的信息，其可用作衡量患者病理状况的尺度。除了连接所述神经元的突触之外，神经网络还包含排列成若干层的多个神经元。在训练过程中，通过跟踪和消除在风险函数的演化中不起重要作用的突触来简化神经网络的结构。例如，这可以通过检查两个发送神经元对相同接收神经元的影响之间的可能相关性来进行，并且在可能的情况下，消除连接到接收神经元的两个突触中的一个。