

1 Einleitung

Der Begriff ‚künstliche neuronale Netze‘ erfährt seit Jahren durch die Medien viel Aufmerksamkeit, ohne daß die Bedeutung immer scharf definiert worden wäre. Oft wurde der Begriff mißverständlich gebraucht. Der häufig mit künstlichen neuronalen Netzen in Verbindung gebrachte Traum vom Universalsystem zum Lösen aller Aufgaben (general problem solver) blieb bisher unerfüllt [13]. Es gibt eine Vielzahl Definitionen für künstliche neuronale Netze. Im folgenden sind zwei davon aufgeführt.

“Neuronale Netze (NN), oft auch als künstliche neuronale Netze (KNN) oder artificial neural networks (ANN) bezeichnet, sind informationsverarbeitende Systeme, die aus einer großen Anzahl einfacher Einheiten (Zellen, Neuronen) bestehen, die sich Information in Form der Aktivierung der Zellen über gerichtete Verbindungen (connections, links) zusenden.” ([175], S. 23).

Künstliche neuronale Netze sind abstrakte Modelle, deren Funktion und Struktur sich an Nervensystemen lebender Organismen orientiert. Die ursprüngliche Idee bei der Entwicklung künstlicher neuronaler Netze war die Simulation der Funktionen natürlicher neuronaler Netze als informationsverarbeitende Systeme [137].

Wesentliche Eigenschaften natürlicher Nervenzellen und Nervensysteme wurden unter dem Gesichtspunkt der Informationsverarbeitung analysiert und mit abstrakten Modellen und in erheblich geringerer Komplexität nachgebildet. Beim menschlichen Gehirn wird von 100 Milliarden Nervenzellen ausgegangen, die jeweils mit ca. 10000 anderen Nervenzellen verbunden sind. Diese Komplexität ist mit der heute verfügbaren Technik nur mit unrealistischem Aufwand machbar.

Analog zu den in Nervensystemen vorhandenen Neuronen bestehen künstliche neuronale Netze aus parallel arbeitenden Prozessoren. Die Prozessoren sind je nach Netzwerkstruktur in mehreren Schichten angeordnet und miteinander verbunden. Jeder Verbindung zwischen zwei Prozessoren ist ein Gewicht zugeordnet. Beim ‘Lernen’ der Daten wird die Stärke der Verbindung zwischen den Prozessoren optimiert, so daß das künstliche neuronale Netz nach der Lernphase in der Lage ist, für neue Daten das Ergebnis zu prognostizieren.

Die Verwendung neuronaler Netze in der Informatik hat gezeigt, daß sich je nach Anwendungszusammenhang unterschiedliche Netzwerkstrukturen für die verschiedenen Aufgabenstellungen eignen. So haben z.B. künstliche neuronale Netze, die für Bildverarbeitungsaufgaben optimiert wurden, sehr viele Eingabeneuronen [29, 96, 156], künstliche neuronale Netze für Spracherkennungsaufgaben dagegen vergleichsweise wenige [3, 85, 116, 143].

Der Vorteil künstlicher neuronaler Netze liegt in der Möglichkeit innerhalb kurzer Zeit ein Modell zu finden. Bei Vorhandensein komplexer Strukturen, bei denen lineare oder einfache funktionelle Ansätze für eine Erklärung der Zusammenhänge nicht ausreichen, ist mit neuronalen Netzen häufig eine Verbesserung der Prognosequalität zu erreichen.

Künstliche neuronale Netze werden in vielen Bereichen der Wissenschaft untersucht [6, 8, 30, 31, 37, 62, 63, 64, 71, 86, 99, 110, 112, 172, 173]. Für Biologen ist die Analogie zu natürlichen neuronalen Netzen interessant. Mit künstlichen neuronalen Netzen wird versucht, Erklärungen für Eigenschaften und Verhalten der natürlichen Netze zu finden [108, 117, 144, 163]. In der Informatik stehen meistens andere Dinge im Vordergrund der Forschung: die Lernfähigkeit und die Effizienz künstlicher neuronaler Netze [10, 11, 40, 74, 89, 90, 101].

Moderne Computerprozessoren arbeiten derzeit bereits mit Schaltgeschwindigkeiten von über 800 MHz ($8 \times 10^8/s$) und auch die Gigahertzgrenze ist vor einiger Zeit überschritten worden. Trotzdem erreichen diese

modernen Maschinen bei weitem nicht die Leistungsfähigkeiten von menschlichen Gehirnen, die mit einer Schaltgeschwindigkeit von nur ca. 1 kHz arbeiten ($10^3/s$) [175]. Wichtigster Grund dafür ist die massive Parallelität bei der Signalverarbeitung des Gehirns. Während bei den üblichen Computern nur der Prozessor aktiv ist und die an festen Speicherstellen abgelegten Informationen seriell verarbeitet werden, sind beim Gehirn die Neuronen zu fast jedem Zeitpunkt aktiv und erlauben ein gleichzeitiges Verarbeiten der assoziativ gespeicherten Informationen. Um diesen Unterschied auszugleichen, wurden massiv parallele Rechnersysteme geschaffen, die aus mehr als Tausend Prozessoren bestehen. Doch auch mit diesen Maschinen kann die Leistungsfähigkeit des menschlichen Gehirns nicht erreicht werden. Diese Maschinen sind hochspezialisiert und nur für wenige Aufgaben, wie z.B. aufwendige Berechnungen geeignet. Wenn versucht wird, mit Computern allgemeine Bildverarbeitungsaufgaben zu lösen, erweisen sich die Maschinen dafür als ungeeignet. Nach Messungen von Psychologen benötigen Menschen ca. 0,1 s um einen bekannten Gegenstand oder eine bekannte Person sicher zu erkennen. Die heute verwendeten Systeme zur Identifikation von Personen brauchen erheblich mehr Zeit und funktionieren deutlich schlechter als das menschliche Gehirn [146].

Bei aktuellen Hard- und Softwareentwicklungen wird versucht, parallele Signalverarbeitungsprozesse zu etablieren [9, 28, 57, 60, 73, 151]. Jedoch fehlen geeignete Algorithmen, um alle Aufgaben effizient auf mehrere Prozessoren automatisch zu verteilen. Außerdem führt der Kommunikationsaufwand zwischen den Prozessoren bei verteilter Aufgabenlösung zu erheblichen Geschwindigkeitsverlusten [16, 53, 100, 155].

Aus statistischer Sicht und speziell für Prognoseschätzungen sind vor allen Dingen die Lernfähigkeit und die Flexibilität künstlicher neuronaler Netze interessant [65, 132].

“Statistik ist die Kunst, Daten zu gewinnen, darzustellen, zu analysieren und zu interpretieren, um zu neuem Wissen zu gelangen” [142]. Dieser Definition folgend, darf die Verwendung künstlicher neuronaler Netze für Datenanalyse nicht als Statistik bezeichnet werden. Im Unterschied zu klassischen statistischen Methoden stellt die Interpretation der Strukturen künstlicher neuronaler Netze ein bedeutendes Problem dar.

Hinter den zu analysierenden Daten stehen häufig Muster, die sich nur mit dafür geeigneten Methoden erkennen lassen. Falls nur ein Faktor das Ergebnis beeinflusst und dieser Faktor nur einen linearen Effekt auf das Ergebnis hat, ist die Aufdeckung des Zusammenhangs mit statistischen Methoden relativ einfach zu bewältigen. Schwieriger ist es, wenn mehrere Faktoren gleichzeitig das Ergebnis beeinflussen, die Faktoren nicht linear wirken oder sich gegenseitig beeinflussen. Deshalb wurden statistische Verfahren immer weiter verfeinert.

Die Prognoseschätzung hat für die Analyse von Daten in der Medizin eine große Relevanz. Aus diesem Grund sind in den letzten Jahren wesentliche Impulse für die Entwicklung von Methoden zur Prognoseschätzung aus der Biostatistik und Bioinformatik gekommen [18, 25, 44, 52, 130, 134, 135, 153, 158, 169].

Für eine möglichst genaue Prognoseschätzung gibt es verschiedene Gründe. Prognoseschätzungen sind sinnvoll, um zum Beispiel Patienten über einen möglichen Verlauf ihrer Erkrankung zu informieren. Informationen zur Prognose können für Entscheidungsfindungen herangezogen werden, z.B. ob bei einem Patienten eine oder andere Untersuchung zusätzlich durchgeführt werden soll. Außerdem können Prognoseangaben sehr gut für die Bewertung neuer diagnostischer Tests in der Medizin genutzt werden.

Im Bereich der Sozialforschung können sie beispielsweise angewendet werden, um bei Arbeitssuchenden die Chancen auf dem Arbeitsmarkt zu erklären. Im Sinne einer Entscheidungsfindung können sie eingesetzt werden, um herauszufinden, welche zusätzlichen Ausbildungen die Arbeitsmarktchancen bei Arbeitslosen verbessern können. Auch die Bewertung der Wirksamkeit von Steuerungsmechanismen für den Arbeitsmarkt ist ein denkbare Einsatzgebiet.

Unabhängig vom Einsatzgebiet sind Prognosemodelle sehr gut geeignet, den Einfluß eines einzelnen Faktors auf das Ergebnis zu bewerten, wenn gleichzeitig auch andere Faktoren das Ergebnis beeinflussen und diese Faktoren gleichfalls bekannt sind und gemessen werden können.

Die Informationen von Prognosemodellen können für die Auswahl eines geeigneten Studiendesigns bedeutsam sein. So kann z.B. in Abhängigkeit vom Prognosemodell über die Notwendigkeit eines stratifizierten Designs bei kontrollierten Studien entschieden werden.

Wegen all dieser Verwendungsmöglichkeiten muß an Prognosemodelle die Forderung nach möglichst genauer Beschreibung eines zugrundeliegenden Musters in den Daten gestellt werden. Außerdem werden bei solchen Modellen möglichst gute Verallgemeinerungseigenschaften benötigt. Ein Prognosemodell sollte immer auch für vergleichbare Daten, die eventuell auch unter anderen Bedingungen erhoben wurden oder aus anderen Einrichtungen oder Regionen stammen, zutreffen.

1.1 Einführung in die Grundbegriffe neuronaler Netze

Künstliche neuronale Netze (KNN), oft auch vereinfachend nur neuronale Netze genannt, sind Informationssysteme, die in Anlehnung an wichtige Aspekte der Funktion von Nervenzellen entwickelt wurden. Sie erlauben die Modellierung komplexer mathematischer Zusammenhänge.

Zur Lösung eines Problems werden Netzparameter durch eine definierte Lernstrategie verändert. Nach einer ausreichenden Lernphase kann das künstliche neuronale Netz für unbekannte Daten eine Prognose treffen. Die genaue Arbeitsweise von künstlichen neuronalen Netzen ist in den folgenden Abschnitten dargestellt.

Künstliche neuronale Netze basieren auf Berechnungen, die auf ein Netzwerk von einfachen Prozessoren verteilt werden. Diese Prozessoren können als künstliche Neuronen verstanden werden. Die Neuronen und die Verbindungen zwischen ihnen werden mit Computerprogrammen simuliert. Die Beschreibung der Funktion und wesentlicher Eigenschaften künstlicher neuronaler Netze erfolgt in Anlehnung an die Arbeiten von Rumelhart [141].

1.1.1 Aufbau künstlicher neuronaler Netze

Künstliche neuronale Netze bestehen aus *Neuronen*¹, dem *Verbindungsnetzwerk* der Neuronen, einer *Propagierungsfunktion* und einer *Lernregel*. In Abbildung 1 ist ein vorwärtsgerichtetes Netz mit zwei trainierbaren Stufen schematisch dargestellt. Von den Neuronen der Eingabeschicht werden eingegebene Daten in das Netz weitergeleitet. Von der Ausgabeschicht werden die Informationen für den Benutzer aus dem Netz ausgegeben. Die verdeckte Schicht ist üblicherweise für Anwender neuronaler Netze nicht sichtbar.

¹ in der Literatur auch als Zellen, Elemente, Perzeptrons oder Units bezeichnet

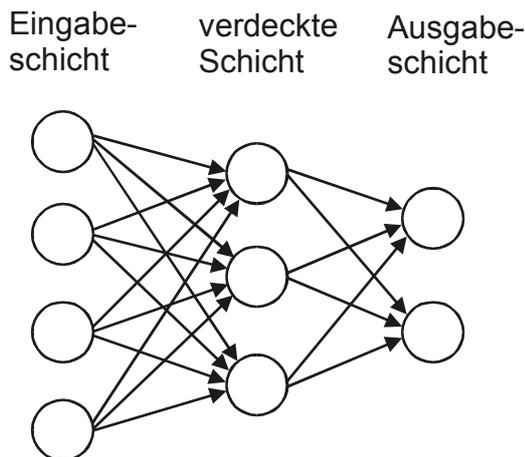


Abbildung 1: Vorwärtsgerichtetes zweistufiges neuronales Netz

Häufig verfügen die Netze über mehr als eine verdeckte Schicht. Bei n verdeckten Schichten wird ein künstliches neuronales Netz als (n+1)-stufig bezeichnet.

1.1.2 Neuronen als kleinste Einheit künstlicher neuronaler Netze

Das Neuron besitzt einen Aktivierungszustand $a_j(t)$. Mit der Aktivierungsfunktion f_{act} wird der neue Aktivierungszustand $a_j(t+1)$ berechnet. Für das Neuron j ergibt sich aus der alten Aktivierung a_j , der Netzeingabe net_j , einem Schwellwert θ_j unter Verwendung der Aktivierungsfunktion der neue Aktivierungszustand mit

$$a_j(t+1) = f_{act}(a_j(t), net_j(t), \theta_j) .$$

Die Ausgabe o des Neurons j wird mit der Ausgabefunktion f_{out} bestimmt:

$$o_j = f_{out}(a_j) .$$

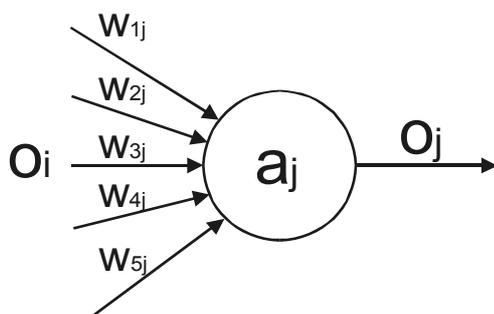


Abbildung 2: Neuron

Ein künstliches neuronales Netz kann als gewichteter, gerichteter Graph angesehen werden. Die gewichteten Verbindungen zwischen den Neuronen entsprechen den Kanten. Das Gewicht der Verbindung zwischen Neuron i und j wird als w_{ij} bezeichnet.

Mit der Propagierungsfunktion wird aus den Eingabewerten in das Neuron o_j (in der Regel Ausgabe anderer Neuronen) der Wert berechnet, den Neuron j in das Netz eingibt. Dabei wird um das Gewicht w korrigiert:

$$net_j(t) = \sum_i o_i(t) w_{ij} .$$

Die Lernregel ist ein Verfahren, nach dem das Netz darauf trainiert wird, aus gegebenen Werten die gewünschte Ausgabe zu generieren. Lernen in künstlichen neuronalen Netzen wird meist durch die Modifikation von Gewichten realisiert. Durch den Einsatz einer Minimierungsfunktion wird versucht, den Fehler zwischen erwarteter und berechneter Ausgabe des Netzes möglichst klein zu halten.

1.1.3 Beispiel eines neuronalen Netzes

Zur Demonstration der Arbeitsweise künstlicher neuronaler Netze wird gern die Funktion für die exklusive Oder-Verknüpfung (XOR) binärer Werte verwendet. Die Funktion liefert den Wert eins nur dann, wenn entweder Eingabe 1 oder Eingabe 2 den Wert eins besitzen. Falls Eingabe 1 und Eingabe 2 den gleichen Werte besitzen, liefert die Funktion den Wert Null. Die möglichen Ausgaben dieser Funktion sind in Tabelle 1 dargestellt.

Tabelle 1 : Ausgaben der XOR-Funktion

Eingabe 1	Eingabe 2	Ausgabe
0	0	0
0	1	1
1	0	1
1	1	0

Bei einem künstlichen neuronalen Netzwerk für die XOR-Funktion werden nur die Binärwerte 0 und 1 für die Aktivierung der Neuronen benötigt. Es genügt ein Netz mit zwei Neuronen in der Eingabeeinheit, einem Neuron in der verdeckten Einheit und einem Neuron in der Ausgabeneinheit (Abbildung 3).

Die Eingabeneuronen (N1 und N2, Abbildung 3) leiten die Eingabewerte unverändert weiter, sie besitzen als Aktivierungs- und Ausgabefunktion die Identitätsfunktion und einen Schwellenwert 0. Neuron N3 besitzt einen Schwellenwert von 1,5. Nur wenn von den Neuronen N1 und N2 der Wert 1 ankommt, wird der Schwellenwert 1,5 überschritten. Neuron N4 erhält dann von Neuron N3 die Eingabe -2 . Dadurch werden die direkten Eingaben von den Neuronen N1 und N2 (jeweils 1) kompensiert. Neuron N4 liefert für diesen Fall die Ausgabe 0. Wenn jedoch nur Neuron N1 die Eingabe N1 erhält, wird der Schwellenwert von Neuron N3 nicht erreicht. Neuron N3 liefert an Neuron N4 den Wert 0, von Neuron N1 kommt der Wert 1 und von Neuron N2 der Wert 0. Zusammen ergibt das den Wert 1. Der Schwellenwert 0,5 wird überschritten, Neuron N4 liefert als Ausgabe 1.

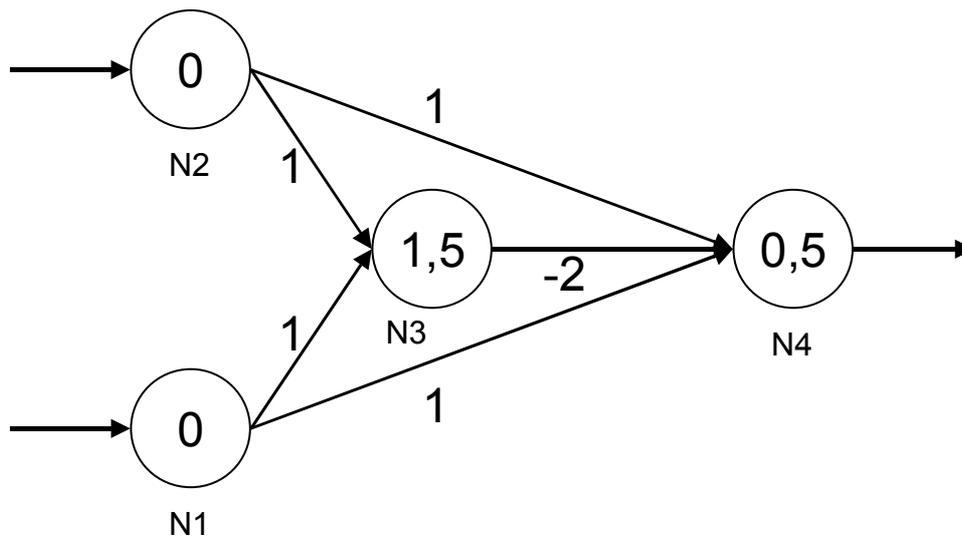


Abbildung 3: Beispiel für ein neuronales Netz zur Präsentation der XOR-Funktion

Symbolisch lassen sich die Gewichte W und die Schwellenwerte dieses Netzes in der Form

$$W = \begin{bmatrix} w_{11} & w_{12} & w_{13} & w_{14} \\ w_{21} & w_{22} & w_{23} & w_{24} \\ w_{31} & w_{32} & w_{33} & w_{34} \\ w_{41} & w_{42} & w_{43} & w_{44} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & -2 \\ 0 & 0 & 0 & 0 \end{bmatrix} \text{ und } \theta = \begin{bmatrix} 0 \\ 0 \\ 1,5 \\ 0,5 \end{bmatrix}$$

darstellen.

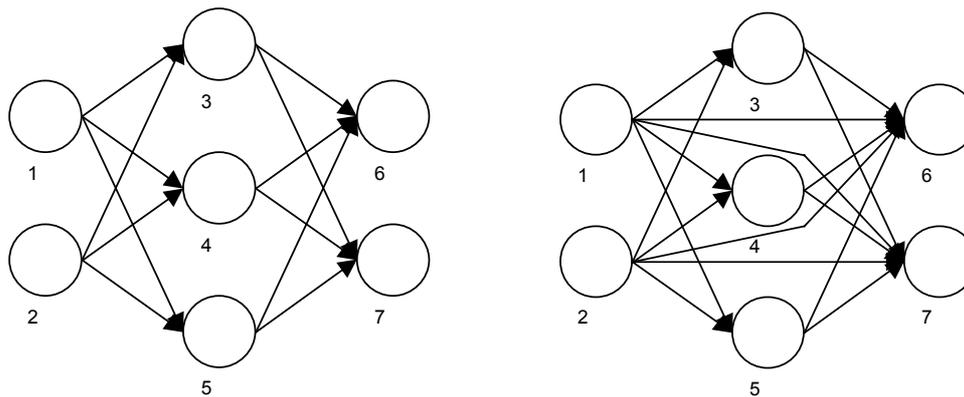
Bei diesem Beispiel läßt sich leicht nachrechnen, daß das Netz für alle vier verschiedenen Möglichkeiten der Dateneingabe jeweils das richtige Ergebnis liefert. Üblicherweise sind neuronale Netze sehr viel komplexer und eine Simulation ist nur am Computer möglich. Die Bedeutung der Gewichte und Schwellenwerte ist bei komplexeren Netzen ohne Hilfsmittel nicht überschaubar.

1.1.4 Topologien neuronaler Netze

1.1.4.1 Netze ohne Rückkopplung (feedforward-Netze)

Bei dieser Art von Netzen gibt es keine Verbindungen, die direkt oder über ein zwischengeschaltetes Neuron wieder zum Ausgangsneuron zurückführen. Aus mathematischer Sicht handelt es sich bei diesen Netzen um azyklische Graphen.

Bei den Netzen ohne Rückkopplung wird zwischen ebenenweise verbundenen Netzen und Netzen, bei denen Verbindungen zwischen allen Ebenen existieren (*skip layer connections*), unterschieden (Abbildung 4).



a) Ebenenweise verbundenes Netz

b) Vollständig verbundenes Feedforward-Netz

Abbildung 4: Netze ohne Rückkopplung

Die Netze ohne Rückkopplung werden üblicherweise für die statistische Datenanalyse verwendet. Eine größere Anzahl von Verbindungen bedeutet aus statistischer Sicht auch immer eine Erhöhung der Zahl der Freiheitsgrade und birgt die Gefahr einer Überanpassung des Modells (overfitting). Andererseits bringt eine Erhöhung der Freiheitsgrade auch immer einen Zugewinn an Flexibilität.

1.1.4.2 Netze mit Rückkopplung (feedback-Netze)

Rückkopplung bedeutet, daß die Ausgaben eines Neurons wieder in das gleiche Neuron oder einer vorhergehende Schicht eingehen.

Bei den Netzen mit Rückkopplung wird nach Zell [175](S. 78) zwischen Netzen mit direkten Rückkopplungen, Netzen mit indirekten Rückkopplungen, Netzen mit Rückkopplungen innerhalb einer Schicht und vollständig verbundenen Netzen unterschieden.

Obwohl die Netze mit Rückkopplung eine noch höhere Flexibilität versprechen als die vorwärts gerichteten Netze, wird in Anlehnung an die statistische Literatur in dieser Arbeit darauf nicht weiter eingegangen. Netze mit Rückkopplungen wurden im Rahmen dieser Arbeit nicht verwendet. Schon bei den vorwärts gerichteten Netzen existieren viele Möglichkeiten zur Optimierung der Netze. Die Einführung weiterer Verbindungen zwischen den Neuronen bedeutet zusätzlich auch eine Erhöhung der Freiheitsgrade des Netzes.

1.1.5 Lernregeln

Die Bezeichnung "Lernen" ist die optimistische Umschreibung des Vorgangs der Funktionsanpassung. Ziel des Lernvorgangs ist eine möglichst gute Übereinstimmung zwischen der erwünschten und der tatsächlichen Ausgabe des neuronalen Netzes. Da bei der Entwicklung der künstlichen neuronalen Netze versucht wurde, die Eigenschaften des Gehirns nachzubilden, wurde in diesem Zusammenhang der Begriff "Lernen" eingeführt.

Das Lernen neuronaler Netze stellt unter statistischen Gesichtspunkten einen sehr interessanten Aspekt dar. Über das Lernen wird letztendlich die Qualität des resultierenden neuronalen Netzes festgelegt. Obwohl es über Lernverfahren und deren Verbesserung eine Vielzahl Publikationen gibt [45, 61, 79, 82, 94, 141, 168, 175], konnte bisher kein optimales Lernverfahren gefunden werden. Vielmehr ist die Auswahl des Lernverfahrens wesentlich durch den Anwendungszusammenhang, also durch die Netzwerkstruktur und durch die vorhandenen Daten bestimmt. Für das Lernen sind verschiedene Anpassungsformen denkbar:

- Aufbau neuer Verbindungen
- Löschen von Verbindungen
- Veränderung des Gewichts von Verbindungen
- Veränderung des Schwellenwertes von Neuronen
- Modifikation der Aktivierungsfunktion
- Modifikation der Propagierungsfunktion
- Modifikation der Ausgabefunktion
- Hinzufügen von Neuronen
- Entfernen von Neuronen

Von diesen vielen Möglichkeiten, die einzeln oder in Kombination angewendet werden können, wird häufig die Modifikation der Gewichte in den Vordergrund gestellt. Die bekanntesten Lernregeln sind die Hebbsche Lernregel, die Delta-Regel für Netze mit einer trainierbaren Ebene (auch Widrow-Hoff-Regel) und die Backpropagation-Regel.

Die Hebbsche Lernregel lautet

$$\Delta w_{ij} = \eta h(o_i, w_{ij}) g(a_j, t_j),$$

wobei Δw_{ij} die Gewichtsänderung, η die Lernrate (eine Konstante), o_i die Aktivierung des Vorgängerneurons i , w_{ij} das Gewicht zwischen dem Vorgängerneuron i zum Neuron j , a_j die Aktivierung des Nachfolgeneurons j , und t_j die erwartete Aktivierung darstellen. Die Gewichtsänderung ist das Produkt aus den näher zu spezifizierenden Funktionen h und g und der Lernrate.

Die Lernregel für Backpropagation lautet

$$\Delta w_{ij} = \eta o_i \delta_j, \text{ mit}$$

$$\delta_j = \begin{cases} f_j'(net_j)(t_j - o_j) & \text{falls } j \text{ ein Ausgabeneuron ist} \\ f_j'(net_j) \sum_k (\delta_k w_{jk}) & \text{falls } j \text{ ein verdecktes Neuron ist} \end{cases}$$

Üblicherweise wird als Propagierungsfunktion die Funktion

$$net_j(t) = \sum o_i(t) w_{ij}$$

verwendet. Mit der üblicherweise verwendeten logistischen Aktivierungsfunktion

$$f(x) = \frac{1}{1 + e^{-x}}$$

und der Identität als Ausgabefunktion ergibt sich für die Backpropagation

$$\delta_j = \left\{ \begin{array}{ll} o_j(1 - o_j)(t_j - o_j) & \text{falls } j \text{ ein Ausgabeneuron ist} \\ o_j(1 - o_j) \sum_k (\delta_k w_{jk}) & \text{falls } j \text{ ein verdecktes Neuron ist} \end{array} \right\}.$$

Häufig wird zwischen überwachtem und nicht überwachtem Lernen unterschieden. Beim überwachten Lernen wird von einem Trainer zu jedem Eingabemuster der Trainingsmenge eine erwartete Ausgabe vorgegeben. Aufgabe des Lernverfahrens ist die Veränderung der Netzgewichte, so daß das Netz nach wiederholter Präsentation des Trainingsmusters eine korrekte Ausgabe liefert.

Das unüberwachte Lernen wird z.B. bei Kohonen's selbstorganisierenden Karten verwendet [92]. Dieses Verfahren dient dazu, ähnliche Eingabemuster in möglichst gleiche Kategorien einzuteilen und zuzuordnen. Dabei werden nur Eingabemuster präsentiert. Die Anwendung künstlicher neuronaler Netze in dieser Arbeit konzentriert sich auf überwachtes Lernen.

1.2 Eigenschaften neuronaler Netze

Wenn neuronale Netze als Methode zur Modellierung statistischer Zusammenhänge betrachtet werden, lassen sich einige nützliche, aber auch nachteilige Eigenschaften gegenüber den klassischen statistischen Methoden feststellen:

Lernfähigkeit	Neuronale Netze werden durch automatische Verfahren an die Daten angepaßt. In der Terminologie der Informatik wird dieser Prozess als Lernen oder als Training bezeichnet. Dabei erfolgt das Training mit einer Teilmenge der vorhandenen Daten. Das trainierte Netz wird später an weiteren Daten validiert und getestet.
Parallelität	Neuronale Netze sind ein Ansatz, der sich sehr gut auf Parallelcomputern umsetzen läßt. Besonders für die Analyse von großen Datenmengen mit komplexen Zusammenhängen sind daher Vorteile zu erwarten.
Robustheit	Neuronalen Netzen wird eine gewisse Robustheit gegen Rauschen in den Daten nachgesagt, vorausgesetzt, die Netze wurden richtig trainiert.
Wissenserwerb	Es ist kompliziert, Netze mit existierendem Vorwissen zu initialisieren, d.h. bekanntes Wissen über die Daten kann nur sehr begrenzt in das neuronale Netz eingebracht werden.
Wissensanalyse	Es ist sehr schwierig, das Wissen eines Netzes zu analysieren und zu verstehen. Zwar gibt es mittlerweile einige Vorschläge und Verfahren dazu, die aber nur in begrenztem Umfang anwendbar sind.

1.3 Angaben zur Geschichte der künstlichen neuronalen Netze

Bereits 1943 wurden von McCulloch und Pitts die ersten künstlichen neuronalen Netze entworfen [107]. Obwohl eine praktische Anwendung dieser ersten Ideen nicht belegt ist, hatten sie einen wichtigen Einfluß auf die später folgenden Arbeiten [175].

1949 wurde von Hebb ein einfaches, aber universell einsetzbares Lernverfahren vorgestellt. Dieses Verfahren wurde später als Hebb'sche Lernregel bekannt [77].

Der erste erfolgreich einsetzbare Neurocomputer (Mark I Perceptron) wurde von Frank Rosenblatt, Charles Whitman und Mitarbeitern am MIT entwickelt. Diese Maschine war ein Analogrechner und funktionierte noch mit motorgetriebenen Potentiometern [139].

1969 erschien das Buch "Perceptrons" [109]. In diesem Werk von Marvin Minsky und Seymour Papert wurde eine genaue mathematische Analyse des Perzeptrons vorgenommen. U.a. wurde darin gezeigt, daß mit dem Perzeptron bestimmte mathematische Probleme nicht gelöst werden können, wie z.B. das XOR-Problem (siehe Beispiel, S. 10).

Nach der Publikation des Buches „Perceptrons“ gab es bis zum Beginn der achtziger Jahre auf dem Gebiet der künstlichen neuronalen Netze nur wenige Arbeiten.

Einen wichtigen Einfluß hatte die in *Nature* erschienene Publikation "Learning internal representations by back-propagating errors" [141]. Darin wurde das Back-Propagation-Lernverfahren mit genauer mathematischer Beschreibung als Gradientenabstiegsverfahren des Netzwerkfehlers erneut vorgestellt. Im Gegensatz zur Erstpublikation in der Dissertation von Paul Werbos [170], wurde dieses Mal das Verfahren von vielen Wissenschaftlern aufgegriffen.

In den folgenden Jahren hat sich der Forschungsbereich zu einem wichtigen Zweig der Informatik entwickelt. Zahlreiche Anwendungen sind aus den Forschungslabors hervorgegangen.

Die Verwendung neuronaler Netze konzentriert sich bisher hauptsächlich auf die Bereiche der automatischen Bilderkennung, der Spracherkennung und der Robotik. Die Gemeinsamkeit dieser Einsatzgebiete ist, daß aufgrund weniger empirischer Daten eine Generierungsfunktion erzeugt werden kann.

1.4 Neuronale Netze und Statistik

In den letzten Jahren ist die Verwendung künstlicher neuronaler Netze für die Datenanalyse populär geworden. Als neue Entwicklung auf diesem Gebiet zeigt sich, daß die Hersteller von großen Datenbankservern interessiert sind, für die Benutzer Werkzeuge zum *data mining*, also zur automatischen Analyse großer Datenbestände mitzuliefern. Hier sollen auch künstliche neuronale Netze eingesetzt werden [97, 102, 129]. Die neuronalen Netze werden als interessante Ergänzung der statistischen Methoden angesehen, da von ihnen häufig bessere Ergebnisse bei der automatischen Datenanalyse erwartet werden als von statistischen Verfahren [39, 84, 136, 171, 174].

Mittlerweile gibt es zahlreiche Beispiele für die erfolgreiche Verwendung neuronaler Netze, besonders bei der Lösung von Klassifikationsproblemen zum Beispiel in [4, 24, 33, 48]. Auch für die Analyse von Verlaufsdaten wurden verschiedene Ansätze vorgestellt und an Beispielen demonstriert [14, 22, 23, 32, 42, 43, 44, 49, 98, 118, 120, 121]. Die grundsätzliche Idee besteht im Ersetzen von einfachen funktionellen (meist linearen) statistischen Modellen durch sehr flexible neuronale Netze.

Bisher konnte an einzelnen Beispielen gezeigt werden, daß neuronale Netze den klassischen statistischen Verfahren bei bestimmten Problemstellungen zumindest gleichwertig sind [119, 122, 133].

Der Hauptvorteil der neuronalen Netze liegt darin, daß sie eine Methode darstellen, die unabhängig von Verteilung und Skalierung der zu analysierenden Daten verwendet werden kann. Durch die automatische Modellsuche können theoretisch sehr komplexe Modelle innerhalb kürzester Zeit gefunden werden. Ein entscheidender Vorteil ist, daß alle für die Erklärung der interessierenden Variablen relevanten Informationen erhalten bleiben.

Es ist aber eine Illusion, anzunehmen, daß neuronale Netze das genaue Wissen über ein System ersetzen können. Algorithmus und Netze können nur zuverlässige Prognosen treffen, wenn das Netz die notwendigen Informationen enthält und mindestens ebensoviele Freiheitsgrade wie das System besitzt [137]. Für neuronale Netze gilt, wie auch für Statistik im Allgemeinen, daß die besten Methoden unzureichende oder inkohärente Daten nicht in zuverlässige Ergebnisse verwandeln können.

1.5 Neuronale Netze und statistische Regression

Wie bereits dargestellt, werden Backpropagation-Netze dann verwendet, wenn bestimmte Eingabewerte auf Ausgabewerte abgebildet werden sollen. Insofern kann ein Backpropagation-Netz als Funktion aufgefaßt werden. Das statistische Äquivalent dazu stellt die Regression dar. Das Problem optimaler Gewichte für eine Trainingsmenge T von Ein- und Ausgabepaaren im Fall eines linearen Zusammenhangs wird in der Statistik als multiple lineare Regression bezeichnet.

Wenn in einem neuronalen Netz die Neuronen eine lineare Aktivierungsfunktion verwenden, entspricht das im wesentlichen einer linearen Regression. Bei den üblicherweise verwendeten sigmoiden Aktivierungsfunktionen werden jedoch komplexe, nichtlineare Regressionsprobleme gelöst.

1.6 Prognosemodelle und individuelle Prognose

Die Analyse verlaufsbezogener Daten mit unvollständigen Beobachtungen (zensierte Daten) ist ein Problem, das sich besonders häufig in der Medizin stellt. So wird z.B. in der Onkologie häufig die Frage nach der Bedeutung einzelner klinischer Faktoren für den Verlauf der Erkrankung gestellt.

Eine einfache Auswertung solcher Daten erfolgt heute üblicherweise mit Kaplan-Meier-Schätzungen und dem Log-Rank-Test. Mit diesen Methoden kann hinsichtlich des Zeitverlaufs die Bedeutung eines Faktors untersucht werden. Für die statistische Auswertung multifaktorieller Fragestellungen stehen viele verschiedene Methoden zur Verfügung. Wahrscheinlich werden Cox-Regressions-Modelle am häufigsten verwendet. Mit dem Cox-PH-Modell können gute Schätzungen der Regressionskoeffizienten, der Hazard-Ratios und von adjustierten Überlebenskurven bei einer Vielzahl von Datensituationen vorgenommen werden.

Eine neuere Methode zur Prognoseschätzung sind Klassifikations- und Regressionsbäume (CART) [19]. Im Unterschied zu linearen oder multiplikativen Modellen sind die Ergebnisse der CART-Analyse bei Verwendung von nicht stetigen Faktoren leichter zu interpretieren. Es können andere als multiplikative Interaktionen entdeckt werden. Um den Zusammenhang zwischen verschiedenen Faktoren und einer Zielvariablen zu beschreiben, werden Klassifikations- bzw. Regressionsregeln aufgestellt. Diese Regeln werden in der Form von binären Bäumen dargestellt. Weitere Hinweise und Beispiele zur Anwendung von CART-Methode wurden von Ciampi beschrieben [34, 35].

Allerdings sind CART-Bäume und das Cox-PH-Modell ungeeignet, komplexe, nichtlineare Zusammenhänge aufzudecken. Sobald Interaktionen zwischen Faktoren betrachtet werden sollen, müssen diese im Cox-PH-Modell als Faktor definiert und in das Modell aufgenommen werden. CART ist geeignet, Interaktionen zwischen Faktoren aufzudecken. Ein wichtiger Nachteil ist die ausschließliche Verwendung binärer Verzweigungen.

1.7 Methoden für die Analyse zensierter Verlaufsdaten

Bei Methoden für die Analyse zensierter Verlaufsdaten handelt es sich prinzipiell um Verfahren, deren Zielgröße eine Zeitperiode bis zum Eintreten eines definierten Ereignisses ist. Beispielsweise kann es sich hier um die Dauer der Arbeitslosigkeit bis zum Wiedereintritt in die Erwerbstätigkeit handeln oder bei medizinischen Daten um die Dauer zwischen dem ersten Auftreten einer Erkrankung und dem Tod. Weil die hier angesprochenen Verfahren häufig im genannten medizinischen Kontext verwendet werden, werden sie auch als Überlebensanalyse bezeichnet.

Obwohl wie im Beispiel der Arbeitslosigkeit das Ereignis ‚Wiedereintritt in die Erwerbstätigkeit‘ in den Biographien der untersuchten Personen mehrfach auftreten kann, sind die hier betrachteten Methoden nur für die Analyse bis zum Eintreten des ersten Ereignisses angelegt.

Das Besondere bei dieser Form der Verlaufsanalyse sind zensierte Beobachtungen. Von einer Zensierung wird ausgegangen, wenn innerhalb der Beobachtungszeit das interessierende Ereignis nicht eingetreten ist. Für Zensierungen sind drei Gründe möglich:

1. die Studie endet, ohne Eintritt des Ereignisses
2. eine weitere Verlaufsbeobachtung ist nicht möglich (lost to follow-up)
3. Herausfallen aus der Studie (z.B. wegen unerwünschter Arzneimittelnebenwirkungen)

In Tabelle 2 sind fiktive Verlaufsdaten dargestellt. Bei den Personen 4 und 6 wurde das Studieneende nach zwölf Wochen erreicht, ohne daß ein Ereignis eingetreten ist. Bei den Personen 2, 3, 5, 7, 8 und 9 ist das Ereignis innerhalb des Beobachtungszeitraums eingetreten. Für die Personen 1 und 10 liegen nach 3 bzw. 7 Wochen zensierte Daten vor.

Tabelle 2: Fiktive Verlaufsdaten von 10 Personen

Person	Beobachtungszeit (Wochen)	Ereignis	Faktor
1	3	0	0
2	4	1	0
3	7	1	1
4	12	0	1
5	8	1	0
6	12	0	1
7	4	1	0
8	6	1	1
9	8	1	1
10	7	0	0

1.7.1 Kaplan-Meier-Überlebenskurven und der Log-Rank-Test

Kaplan-Meier-Überlebenskurven sind die etablierte Form der Darstellung von zensierten Verlaufsdaten. Die Überlebenswahrscheinlichkeit S wird geschätzt mit

$$\hat{S}(t_j) = \hat{S}(t_{j-1}) \cdot \hat{P}(T > t_j | T \geq t_j)$$

oder analog als Produkt-Limit-Schätzung

$$\hat{S}(t_j) = \prod_{i=1}^j \hat{P}(T > t_j | T \geq t_j).$$

Für Tabelle 2 ergibt sich mit diesen Formeln:

```
*** Nonparametric Survival ***
Call: survfit(formula = Surv(Zeit, Ereignis), data = SDF1, na.action = na.omit)

  n events mean se(mean) median 0.95LCL 0.95UCL
10     6 7.89   1.01     8     6     NA
Call: survfit(formula = Surv(Zeit, Ereignis), data = SDF1, na.action = na.omit)

  time n.risk n.event survival std.err lower 95% CI upper 95% CI
  4     9     2   0.778   0.139   0.5485   1.000
  6     7     1   0.667   0.157   0.4200   1.000
  7     6     1   0.556   0.166   0.3097   0.997
  8     4     2   0.278   0.162   0.0888   0.869
```

Zu den Zeitpunkten 4, 6, 7 und 8 treten Ereignisse ein. Die Wahrscheinlichkeit, kein Ereignis bis zum Zeitpunkt 4 zu beobachten, wird aus dem Quotienten der bis zum Zeitpunkt 4 beobachteten Individuen abzüglich der Fälle, bei denen das Ereignis eingetreten ist ($9-2=7$) und der bis zum Zeitpunkt 4 beobachteten Individuen (9) berechnet. Diese Wahrscheinlichkeit wird mit der bis vor diesem Zeitpunkt geschätzten Überlebenswahrscheinlichkeit (in diesem Fall $\hat{S} = 1.0$) multipliziert. Damit ergibt sich eine Überlebenswahrscheinlichkeit \hat{S} von 0,778. Für den Zeitpunkt 6 resultiert $\hat{S} = (6/7) \cdot 0,778 = 0,667$ und $\hat{S} = (5/6) \cdot 0,667$ für den Zeitpunkt 7.

Diese Ergebnisse wurden mit dem Statistikpaket S-Plus erstellt. Gleiche Ergebnisse in etwas anderer Darstellung der Tabelle liefern andere Statistikpakete, wie z.B. SPSS, SAS oder JMP. Abbildung 5 zeigt die Kaplan-Meier-Kurve für diese Daten, die bei 1.0 beginnt und beim Zeitpunkt eines Ereignisses um $\hat{P}(T > \text{Zeit} | T \geq \text{Zeit})$ nach unten gesetzt wird.

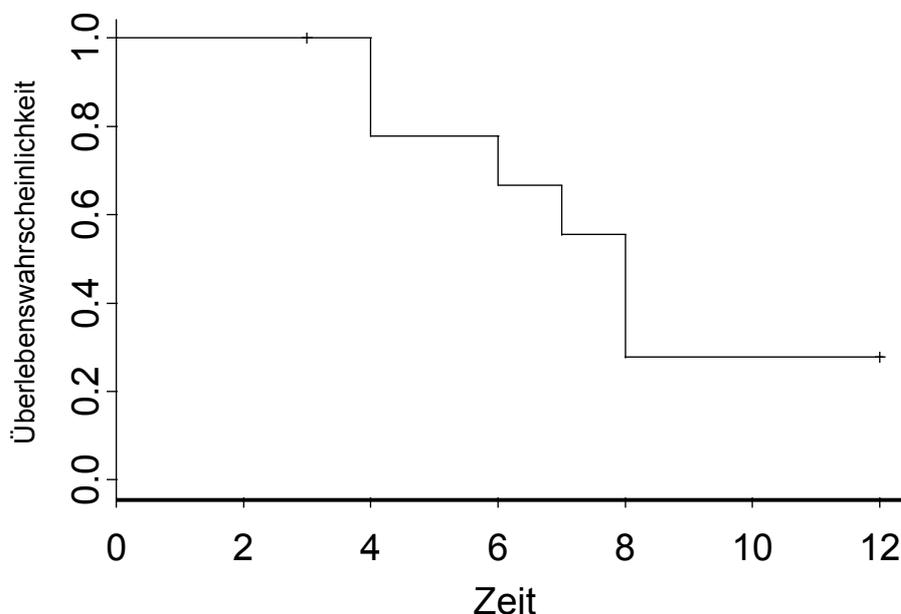


Abbildung 5: Beispiel einer Kaplan-Meier-Überlebenskurve

Bei der Analyse von zwei oder mehr Gruppen stellt sich die Frage, ob Unterschiede zufälliger Natur sind oder ob tatsächlich ein Effekt vorliegt. Zum Vergleich der Kurven kann der Log-Rank-Test herangezogen werden.

Der Log-Rank-Test ist ein χ^2 -Test für große Stichproben. Wie bei den meisten anderen χ^2 -Tests werden auch bei diesem Test Erwartungswerte mit beobachteten Werten verglichen. Dabei beruht der Vergleich auf den vollständigen Überlebenskurven. Dazu werden die beobachteten mit den erwarteten Zellhäufigkeiten verglichen, wobei die Zellen durch das Eintreten eines Ereignisses definiert sind. Die Teststatistik kann berechnet werden mit:

$$X^2 = \frac{(O_i - E_i)^2}{\hat{Var}(O_i - E_i)} \text{ mit}$$

$$O_i - E_i = \sum_j (m_{ij} - e_{ij})$$

i ... Gruppe

j ... Zeitpunkt von Ereignis j

Als Alternative zum Log-Rank-Test wurde der Peto-Test vorgeschlagen [125]. Im Unterschied zum Log-Rank-Test werden hier nicht alle Zeitintervalle gleichmäßig berücksichtigt, sondern gewichtet mit der Zahl der Beobachtungen zum Zeitpunkt j:

$$w(O_i - E_i) = \frac{\sum_j n_j (m_{ij} - e_{ij})}{\sum_j n_j}$$

Dadurch werden die Informationen am Anfang der Kurve stärker berücksichtigt und am Ende der Kurve, wo nur noch wenige Beobachtungen vorliegen, weniger beachtet.

Aus den gegebenen Formeln läßt sich ableiten, daß der Peto-Test im Vergleich zum Log-Rank-Test nicht konservativer sein muß. Es sind Konstellationen denkbar, bei denen der χ^2 -Wert größer oder kleiner ist, als beim Log-Rank-Test.

1.7.2 Klassifikations- und Regressionsbäume für zensierte Daten

Für die explorative Datenanalyse werden seit der Vorstellung dieser neuen Verfahren gern Klassifikations- und Regressionsbäume verwendet. Im wesentlichen wurden diese Ideen von Breimann, Friedman, Olshen und Stone entwickelt [19].

Prinzipiell funktionieren die Bäume nach folgendem Schema: Zuerst wird die Variable ausgesucht, mit der die Daten am besten in zwei Gruppen unterteilt werden können. Die Daten werden danach in die zwei Gruppen unterteilt und der Prozess wird rekursiv für jede Untergruppe wiederholt bis die Gruppen eine definierte Minimalgröße erreichen oder bis keine Verbesserung durch weitere Unterteilungen erreicht werden kann.

Das resultierende Modell ist in der Regel zu komplex und wie bei allen schrittweisen Verfahren stellt sich auch hier die Frage, unter welchen Bedingungen der schrittweise Prozess angehalten werden soll.

Der zweite Schritt der Prozedur besteht aus Kreuzvalidierungen, um den Baum zu beschneiden und ihn so unter Berücksichtigung der Variabilität in den Daten besser an die tatsächlich relevanten Informationen anzupassen.

1.7.3 Das Cox-Proportional-Hazards-Modell

Das Cox-Proportional-Hazard-Modell wird eingesetzt zur Modellierung von zensierten Daten und Kovariaten. Dem Modell liegt die Annahme zugrunde, daß die Hazard-Funktion h von der Zeit t und einem Vektor der Kovariaten

$$x_i = (x_{i1}, x_{i2}, \dots, x_{ip})', i=1, \dots, n \text{ mit}$$

$$h(t, x_i) = h_0(t) \exp(\beta' x_i)$$

abhängt, wobei $\beta = (\beta_1, \dots, \beta_p)'$ und $h_0(t)$ der von den Kovariaten unabhängige baseline hazard ist.

Durch Maximierung der partiellen Likelihood wird $\hat{\beta}$ geschätzt mit

$$L_c(\beta) = \prod_{i=1}^d \frac{h_0(t) \exp(\beta' x_i)}{\sum_l I(T_l \geq t) h_0(t) \exp(\beta' x_j)} = \prod_{i=1}^d \frac{\exp(\beta' x_i)}{\sum_l I(T_l \geq t) \exp(\beta' x_j)}.$$

Die Likelihood wird partiell genannt, weil das Produkt nur über die nichtzensierten Fälle berechnet wird [38]. Diese Methode hat den Nachteil, daß als Bedingung eine Unabhängigkeit der Zensierungen gefordert werden muß bzw. die Likelihood für zensierte Beobachtungen nicht von β abhängen darf. Bei Verwendung der Approximation von Breslow wird die bedingte Likelihood zu

$$L_c(\beta) = \prod_{i=1}^d \frac{\tau_i}{\sum_1^m \tau_j - \frac{i}{d} \sum_1^d \tau_j} \text{ mit } \tau_i = I(T_i \geq t) \exp(\beta' x_i) \text{ und}$$

d als Zahl der insgesamt m möglichen Ereignisse zum Zeitpunkt t .

Aus rechentechnischen Gründen wird die logarithmierte Likelihoodfunktion maximiert. Für die Maximierung werden die partiellen Ableitungen von L nach jedem Parameter verwendet. Die Lösung dieses Gleichungssystems wird über iterative Verfahren gefunden, z.B. mit dem Newton-Raphson-Verfahren [161].

Am besten läßt sich das Cox-Modell an einem einfachen Beispiel demonstrieren. Dazu werden wieder die Daten aus Tabelle 2 verwendet. S-Plus liefert als Ergebnis des Kommandos

```
coxph(formula=Surv(Zeit, Status)~F1, na.action=na.omit, eps=0.0001, iter.max=10, robust=F)
```

folgende Ergebnisse:

```
*** Cox Proportional Hazards ***

n= 10

      coef exp(coef) se(coef)  z    p
F1 -0.839   0.432   0.836 -1 0.32

      exp(coef) exp(-coef) lower .95 upper .95
F1   0.432      2.31   0.084   2.22

Rsquare= 0.094 (max possible= 0.878 )
Likelihood ratio test= 0.98 on 1 df,  p=0.321
Wald test               = 1.01 on 1 df,  p=0.315
Efficient score test = 1.06 on 1 df,  p=0.303
```

1.7.4 Erweiterungen des Cox-Modells

Um die bestehenden Einschränkungen für die Verwendung des Cox-Modells zu umgehen, wurden Erweiterungen entwickelt. Für die Analyse von Daten, bei denen die Proportionalitätsannahme verletzt ist, kann das stratifizierte Cox-Modell verwendet werden. Bei zeitanhängigen Kovariaten gibt es die Möglichkeit, diese im Modell zu berücksichtigen. Genauere Angaben zum Vorgehen dabei finden sich z.B. bei Kleinbaum [91] oder Venables & Ripley [167].

1.7.5 Verwendung von Splines in der Cox-Regressionsanalyse

Bei stetigen Variablen tritt in der Statistik häufig das Problem auf, diese Variablen möglichst genau zu beschreiben. Wenn diese Variablen eine komplexe Struktur aufweisen, sind zur Interpolation höhergradige Polynome nötig.

Ein anderer Ansatz ist die Verwendung von Splines. Mit Splines kann eine komplizierte Kurve intervallweise durch niedergradige Polynome interpoliert werden. Wie ihre englische Bezeichnung ausdrückt, sind sie mathematische Modelle von Kurvenlinealen.

Ein polynomialer Spline $S(t)$ vom Grad m ist definiert durch:

$S(t)$ stimmt auf jedem Intervall mit $[\mu, \mu+1)$ $\mu \in \mathbb{Z}$ mit einem Polynom $p_\mu \in \Pi_m, m \in \mathbb{N}$ überein und $S(t) \in C^{m-1}(\mathbb{R})$

Dabei ergibt sich bei $m=1$ ein Polygonzug, bei $m=2$ eine quadratischer Spline und bei $m=3$ ein kubischer Spline. Damit es einen 'glatten' Kurvenverlauf gibt, müssen Splines an den Intervallgrenzen Anschlußbedingungen erfüllen. Bei höhergradigen Splines müssen an den Intervallgrenzen die Funktionswerte und die Ableitungen identisch sein. Diese Forderung entspricht einer m -maligen Differenzierbarkeit von $S(t)$.

Zu einem Vektor β ergibt sich genau ein kubischer Spline $S(t)$ auf dem Intervall $[0, k)$ mit

$$S(t) = \sum_{\nu=0}^3 \beta_{\nu\eta} b_{\nu 3}(t - \eta), \quad t \in [\eta, \eta + 1], \eta = 0, 1, \dots, k - 1$$

Eine genauere Beschreibung der Definition und der Anwendung von Splines für die Analyse zensierter Verlaufsdaten ist z.B. bei Intrator et al. [83], bei Devlin et al. [47] und bei Koopberg et al. [93] zu finden.

1.7.6 Parametrische Modelle für zensierte Verlaufsdaten

Da im Rahmen dieser Arbeit keine parametrischen Modelle verwendet werden, soll an dieser Stelle auch nicht näher auf diese Ansätze eingegangen werden.

1.7.7 Neuronale Netze für die Analyse zensierter Verlaufsdaten

Für die Analyse zensierter Verlaufsdaten mit neuronalen Netzen werden die Variablen als Eingaben in das Netz verwendet. Die Zeit bis zum Eintritt eines Ereignisses entspricht der Ausgabe des Netzes. Somit können künstliche neuronale Netze als Erweiterung bekannter statistischer Modelle verwendet werden. Auf diese Weise können nichtlineare Funktionen angepaßt werden. Es können auch Modelle mit zeitabhängigen Faktoren untersucht werden.

Wegen der gut untersuchten Eigenschaften des Cox-Modells wird dieser Modellansatz auch für die Analyse mit neuronalen Netzen verwendet. Wie beim statistischen Modell muß auch hier davon ausgegangen werden,

daß die Hazards bei allen Fällen für die gesamte Beobachtungszeit proportional verlaufen. Das Modell ist definiert mit:

$$h(x, t) = h_0(t) \exp \eta(x)$$

$\eta(x)$ wird als lineare Ausgabe eines neuronalen Netzes simuliert. Die logarithmierte partielle Likelihood ist dann gegeben durch

$$\log(L) = \sum_r \left(\eta(x_r) - \log \sum_a \exp \eta(x_a) \right),$$

wobei r nur über die Ereignisse läuft und a über alle Personen mit Beobachtung zum Zeitpunkt dieses Ereignisses. Die partielle Likelihood ist invariant, wenn zu jedem der Scores $\eta(x)$ eine Konstante addiert wird.

1.8 Beispiele für die Verwendung neuronaler Netze bei der Analyse zensierter Verlaufsdaten

Die ersten Ansätze, neuronale Netze für die Analyse zensierter Daten zu verwenden, sind in der medizinischen Literatur zu finden. Die beiden ersten Arbeiten, bei denen künstliche neuronale Netze für die Analyse von Verlaufsdaten verwendet wurden, stammen von Ravdin und Clark [130, 131]. In beiden 1992 publizierten Arbeiten wurde die Prognose von Brustkrebspatientinnen untersucht. Kappen & Neijt veröffentlichten 1993 eine Arbeit über die Prognose bei Ovarialkarzinomen [87]. 1994 erschien von Bugliosi et al. eine Arbeit zur Prognose beim Myelom [24]. Diese ersten Arbeiten zu diesem Thema wurden später vor allen Dingen wegen des unbefriedigenden Umgangs mit zensierten Fällen kritisiert [133, 148]. Ab 1994 erschienen zunehmend mehr Publikationen, bei denen künstliche neuronale Netze für die Prognoseschätzung verwendet wurden.

In der Literatur lassen sich verschiedene Ansätze für die Verwendung neuronaler Netze zur Analyse von Verlaufsdaten finden [44, 59, 98].

1. Reduktion des Prognoseansatzes auf ein Klassifikationsproblem: Bei diesem Ansatz wird das Überleben zu einem einzelnen definierten Zeitpunkt betrachtet. Patienten mit zensierten Daten werden von der Analyse ausgeschlossen. Damit handelt es sich um ein binäres Klassifikationsproblem. Der Ausschluß zensierter Beobachtungen kann eine erhebliche Verzerrung der Ergebnisse bedeuten.
2. Eine Verfeinerung der ersten Methode stellt die Unterteilung der Überlebenszeiten in mehrere Gruppen dar. So könnte die Überlebenszeit beispielsweise in 'Tod nach einem Jahr', 'Tod nach 2 Jahren', 'Tod nach drei Jahren' und 'Tod nach 5 Jahren' unterteilt werden. Diese Daten können dann mit einem Klassifikationsnetzwerk für k Klassen untersucht werden. Da bei dieser Form der Codierung eine ordinal skalierte Zeitvariable erzeugt wird, bietet es sich an, ein 'proportional odds model' [132] zu verwenden. Auch bei diesen Ansätzen ist kein zufriedenstellender Umgang mit Zensierungen gefunden worden. Durch die Bildung von Klassen gehen unter Umständen wichtige Informationen verloren.
3. Es lassen sich in der Literatur verschiedene Ansätze zur Verwendung separater neuronaler Netze bei der Betrachtung für mehrere Zeitperioden finden.
4. Eine andere Möglichkeit ist die Modellierung bedingter Wahrscheinlichkeiten:

$$P(\text{Tod im Intervall } i \mid \text{Überleben in } i-1 \text{ Intervallen}, x) = g(\eta_i)$$

Dieser vierte Ansatz in Anlehnung an das Cox-Proportional-Hazards-Modell, wurde von Farragi und Simon [59] für die Verwendung neuronaler Netze vorgeschlagen.

1.9 Verwendung simulierter und realer Daten

Simulierte Daten werden häufig für die Überprüfung der Leistungsfähigkeit neuer statistischer Methoden verwendet [133, 167]. Auch für die vorliegende Arbeit wurden Daten erzeugt, um die verwendeten Methoden zu überprüfen. Der größte Nachteil simulierter Daten besteht darin, daß sie zu Verallgemeinerungen verleiten können, nur weil diese Daten besonders gut mit der neuen Methode verarbeitet werden können. Deshalb sollten auch immer reale Daten zur Überprüfung der Leistungsfähigkeit von verschiedenen Methoden verwendet werden.

1.10 Multivariable Modellbildung

Für die Modellbildung lassen sich Schritte angeben, die unabhängig von der Auswertungsmethode erforderlich sind, um valide Prognosemodelle und Prognoseschätzungen zu erhalten. Die hier angegebenen Schritte sind im wesentlichen vom "Tutorial in Biostatistics. Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors" [75] abgeleitet. Diese Richtlinien stellen eine Zusammenfassung des aktuellen wissenschaftlichen Standards dar. Sie sollten daher generell für die Analyse von Prognosedaten berücksichtigt werden. Zusätzlich zu den bei statistischen Analysen notwendigen Schritten wird ggf. auf die Besonderheiten bei der Simulation neuronaler Netze eingegangen.

1.10.1 Vorbereitende Schritte

Vor Beginn der Analyse müssen die zu untersuchenden Beziehungen zwischen Variablen, die Zielgrößen und potentielle Prädiktoren definiert werden. Ein häufig auftretendes Problem stellen Daten mit fehlenden Angaben dar. Das Löschen fehlender Angaben kann eine Verfälschung der Ergebnisse verursachen und ist daher keine geeignete Strategie. Es sollte versucht werden, fehlende Angaben mit Hilfe statistischer Verfahren zu ersetzen. Einfache Methoden dafür sind das Einsetzen des Mittelwerts, Medians oder Modalwerts. Bei Variablen, die mit anderen Prädiktoren korreliert sind, führt diese Methode zu Verzerrungen und ist nicht effizient. Angepaßte Regressionsmodelle zur Bestimmung eines Prädiktors in Abhängigkeit von allen anderen Prädiktoren stellen die Methode der Wahl dar. Harrell [75] hat einen Algorithmus von Kuhfeld [95] weiterentwickelt, der nicht nur mit nichtlinearen und nichtmonotonen Daten umgehen kann, sondern auch bei einem hohen Anteil fehlender Angaben gültige Ergebnisse liefert.

Es ist wichtig, ein Maximum an Information aus den Daten zu extrahieren. Aus diesem Grund und um eine hohe Reliabilität der Ergebnisse zu erreichen, sollten alle verfügbaren Informationen vollständig genutzt werden. Eine Verwendung stetiger Variablen ist der Einteilung in Kategorien vorzuziehen. Inhaltliches Wissen über existierende Zusammenhänge sollte bei der Modellbildung unbedingt berücksichtigt werden. Zeitiges Löschen von Faktoren mit nur geringer Chance, Informationen zum Modell beizutragen, hilft, eine Überanpassung zu vermeiden und führt damit zu einer besseren Generalisierbarkeit. Plausible Interaktionen sollten sorgfältig ausgewählt werden, um Probleme mit multiplen Parametern zu vermeiden.

1.10.2 Datenreduktion

Sorgfältig angewandte statistische Methoden sind ein hervorragendes Werkzeug, um prognostische Vorhersagen zu treffen. Wenn aber Modellannahmen verletzt sind oder wenn ein Modell falsch verwendet wird, sind die Modelleigenschaften unbefriedigend.

In der medizinischen Literatur gibt es viele Beispiele für die Analyse von 20-50 Faktoren bei Daten von ca. 100 Patienten. Hier wurde nicht berücksichtigt, daß ein signifikanzbasiertes ‚Heraussuchen‘ von Variablen nicht

zu reliablen Ergebnissen führen kann. Auch führt die schrittweise Variablenselektion zu dem Problem, wie mit multiplen Tests umgegangen werden soll. Zu diesem Thema gibt es zahlreiche Literaturstellen, wo die auftretenden Probleme ausführlich diskutiert werden [46, 58, 154, 159].

Wenn bei der Analyse die Zahl der untersuchten Variablen berücksichtigt werden muß, sind auch zusätzlich konstruierte Variablen z.B. für die Beschreibung von Interaktionen oder nichtlinearen Zusammenhängen zu berücksichtigen. Der Freiheitsgrad für die Modellbildung wird durch die Anzahl der untersuchten Faktoren (Spalten in der Designmatrix) bestimmt. Wenn graphische oder andere Formen der Analyse zur Unterstützung der Modellbildung herangezogen werden, ist es schwierig, einen p-Wert für das Modell zu bestimmen. Es sollte versucht werden, die effektive Zahl der betrachteten Parameter in Abhängigkeit von der Flexibilität der verwendeten Methode zu schätzen. Die Quantität p ist die effektive Anzahl der für die Betrachtung zugelassenen Faktoren bzw. die Zahl der Regressionkoeffizienten, die formell oder informell geschätzt wurden, ohne algebraische Restriktionen.

Um die Genauigkeit eines Modells zu erhöhen, ist es erforderlich, die Zahl der im Modell verwendeten Variablen zu reduzieren – es sei denn, der Stichprobenumfang ist sehr groß. Wenn nicht eine Technik verwendet wird, bei der die Verwendung zusätzlicher Faktoren bestraft wird, sollten Methoden für die Datenreduktion benutzt werden, bei der die eigentliche Zielvariable nicht verwendet wird. Bei der Verwendung von Regressionsmodellen sollten nicht mehr als $m/10$ Faktoren untersucht werden, wobei m als Zahl der nichtzensierten Verläufe in der Überlebensanalyse oder die schwächer besetzte Kategorie bei einer binären Zielvariable definiert ist. Bei $p > m/10$ sollte eine Technik zur Variablenreduktion, wie z.B. Hauptkomponenten-Variablenclustering verwendet werden. Zur Reduktion der Freiheitsgrade eines Modells sind Shrinking-Methoden [166] gut geeignet.

1.10.3 Überprüfung von Modellannahmen

Linearitätsannahme

In ihrer einfachsten Form wird bei allen Regressionsmodellen von einer linearen Abhängigkeit der Prädiktoren von der Zielvariablen ausgegangen. Bei der logistischen Regressionsanalyse für binäre Zielgrößen, ist die initiale Annahme, daß X linear abhängig ist zum log odds der Zielvariablen ($\log[P/(1-P)]$), wobei P die Wahrscheinlichkeit für das Zielereignis darstellt) für Gruppen, die nach X unterteilt werden. Im Cox-Proportional-Hazards-Modell für Überlebensrechnungen wird davon ausgegangen, daß zu jedem Zeitpunkt t , $\log[-\log(S(t))]$ und äquivalent $\log \lambda(t)$ linear von X abhängen. Dabei ist $S(t)$ die Wahrscheinlichkeit bis zum bestimmten Zeitpunkt t zu überleben und $\lambda(t)$ die Hazard-Rate, also die Rate für ein Ereignis zum Zeitpunkt t . Eine Verletzung der Linearitätsannahme ist z.B. eine U-förmige Verteilung beim Alter. Solche Verletzungen der Modellannahmen führen zu fehlerhaften und unbrauchbaren Prognosemodellen und somit zu unbrauchbaren Prognosen.

Eine direkte Möglichkeit zur Überprüfung der Linearitätsannahme und zur Bestimmung eventuell nötiger Transformationen von X besteht in der Zerlegung des Faktors in multiple Terme, die flexible Anpassung an die nötigen Beziehungen erlauben. Die zusätzlichen Terme können statistisch bezüglich der Linearitätsannahme getestet werden. Als zusätzliche Terme werden häufig X^2 und zum Teil auch höhere Potenzen von X in das Modell aufgenommen. Eine flexiblere Möglichkeit stellt die Verwendung von Splines dar. Bei der Verwen-

derung von Splines werden z.B. abschnittsweise logistische Regressionen gerechnet oder abschnittsweise kubische Polynome betrachtet (Siehe Abschnitt 1.7.5., S. 21)

Als Alternative können geglättete Plots der Residuen verwendet werden, um die Form der Funktion für jeden Prädiktor zu bestimmen. Bei der binären logistischen Regressionsanalyse können geglättete Plots der partialen Residuen verwendet werden, bei der Cox-PH-Regression können geglättete martingale Residuen Abweichungen aufdecken.

Additivitätsannahme

Eine weitere Annahme bei den Regressionsmodellen ist die Additivität der Effekte der Prädiktoren oder, mit anderen Worten das Nichtvorhandensein von Interaktionen. Es ist zu beachten, daß Interaktionen durchaus komplexes Aussehen haben können, und damit die im Modell benötigten Kreuzprodukte nicht unbedingt einfach aussehen müssen.

Die Zahl der möglichen Terme für Kreuzprodukte ist üblicherweise sehr groß, besonders wenn unter den Variablen stetige oder kategoriale Daten mit mehreren Ausprägungen vorkommen. Daher muß die Additivität festgelegt werden, bevor die Daten analysiert werden. Andererseits stellen der Fehler erster Art und Überanpassung ein wichtiges Problem dar. Als Kompromißlösung kann ein gepoolter Test auf Interaktionen verstanden werden. Zum Beispiel können in einer Analyse mit den Faktoren Alter, Geschlecht und Dosis jeweils alle Interaktionen mit Alter, mit Geschlecht und dann mit Dosis getestet werden. Ein kombinierter Test mit allen Interaktionen zweiter Ordnung ist ebenfalls sinnvoll. Wenn ein gepoolter Test keine signifikanten Ergebnisse liefert, sollte auf die Einbeziehung dieser Interaktionen verzichtet werden.

Verteilungsannahmen

Wenn entschieden wurde, welche Prädiktoren betrachtet werden und in welcher Form sie in das Modell aufgenommen werden, besitzen die meisten Modelle Annahmen zur Verteilung, die gleichfalls berücksichtigt werden müssen. Im Cox-PH-Modell gibt es keine Annahmen zur Verteilung der Überlebensfunktion $S(t)$ über die Zeit t , aber es muß davon ausgegangen werden, daß sich die Überlebenskurven von verschiedenen Individuen proportional verhalten. Ob diese Annahme erfüllt ist, kann z.B. unter Verwendung von Schoenfelds Residuen überprüft werden. Anders als das Cox-PH-Modell verfügen parametrische Modelle (z.B. Weibull-Modelle) sogar dann über Verteilungsannahmen, wenn keine Kovariablen betrachtet werden.

Wenn die Form der Überlebensfunktion $S(t)$ nicht zu den analysierten Daten paßt, resultieren falsche Prognoseergebnisse.

1.10.4 Quantifizierung der Vorhersagegenauigkeit

Es gibt mehrere Maße für die Vorhersagegenauigkeit von Prognosemodellen:

1. Die Quantifizierung des Nutzens eines Prädiktors oder eines Prognosemodells, das für das Screening verwendet wird, um Personen mit einem erhöhten Risiko zu identifizieren.

2. Überprüfung eines vorhandenen Modells auf Überanpassung oder unzureichende Anpassung an die Daten
3. Vergleich von konkurrierenden Modellen oder konkurrierenden Methoden

Wenn die Zielgröße eine stetige Variable ist, die vollständig gemessen werden kann (im Unterschied zu zensierten Daten bei der Überlebensanalyse), ist ein häufig verwendetes Maß der erwartete quadrierte Fehler der Schätzung. Dieses Maß ist definiert als die mittlere quadrierte Differenz der Abweichungen zwischen beobachteten und prognostizierten Werten, wenn das Experiment unendlich oft wiederholt wird und neue Schätzungen bei jeder Wiederholung vorgenommen werden.

Es gibt zwei andere Ausdrücke, um die Komponenten der Vorhersagegenauigkeit zu beschreiben: Kalibrierung und Diskriminierung. Wenn z.B. für eine Patientengruppe mit ähnlicher Prognose eine Mortalität von 20% prognostiziert wird und die tatsächliche Prognose gleichfalls 20% beträgt, kann die Prognose als gut kalibriert gelten. Diskriminierung mißt die Qualität eines Modells hinsichtlich der Fähigkeit, verschiedene Zielereignisse korrekt zu prognostizieren. Ein Wetterbericht, der jeden Tag eine Niederschlagswahrscheinlichkeit von 15% prognostiziert, ist in einer Region mit 55 Regentagen pro Jahr gut kalibriert, aber trotzdem ist diese Information nutzlos. Die Diskriminierung ist gut, wenn die Vorhersagen aus einem breiten Spektrum möglicher Prognosen getroffen werden bzw. wenn es an Tagen, für die Regen prognostiziert wurde, auch tatsächlich regnet.

Wenn ein Modell schlechte diskriminierende Eigenschaften besitzt, kann weder eine Adjustierung noch eine Kalibrierung das Modell korrigieren. Wenn aber die Diskriminierungseigenschaften gut sind, kann die Prognoseschätzung kalibriert werden, ohne die Diskriminierung zu verschlechtern. Kalibrierung der Prognosen bedeutet, sie zu verändern ohne die Rangfolge zu berühren.

1.10.5 Methoden für die Modellvalidierung

Es erscheint naheliegend, daß die Untersuchung der Prognoseleistungen eines Modells nicht anhand der Trainingsdaten erfolgen kann. Den strengsten Test eines solchen Modells stellt eine externe Überprüfung des Modells, also die Verwendung neuer Daten dar, die nicht für die Modellbildung verwendet wurden. Um die Generalisierungseigenschaften eines Modells zu überprüfen, sollte überlegt werden, unabhängig erhobene Daten zu verwenden.

Die wichtigsten Methoden für eine interne Überprüfung eines Modells stellen Datensplitting, Kreuzvalidierung und Bootstrapping dar [56].

Beim Datasplitting wird ein bestimmter Anteil der Daten für die Modellbildung verwendet, z.B. 2/3. Wenn das Modell ‚fertig‘ ist, wird es an den restlichen Daten überprüft. Die Stichprobe für die Validierung muß ausreichend groß sein, damit der Zusammenhang zwischen beobachteten und prognostizierten Ergebnissen mit ausreichender Genauigkeit beschrieben werden kann. Datasplitting ist relativ einfach zu bewerkstelligen, da alle Schritte zur Modellbildung nur ein mal durchgeführt werden müssen. Datasplitting ist auch vorteilhaft, wenn lohnend erscheint die Unterteilung nach geographischen Regionen oder unter zeitlichen Gesichtspunkten vorzunehmen, womit unter Umständen gute Generalisierungseigenschaften des Modells demonstriert werden können.

Kreuzvalidierung bedeutet mehrfaches Datasplitting. Um exakte Schätzungen unter Benutzung der Kreuzvalidierung zu erhalten, ist es nötig, mehr als 200 Modelle zu entwickeln und mit Ergebnissen die aus 200 Wiederholungen stammen, zu testen. Die Vorteile der Kreuzvalidierung gegenüber dem Datasplitting bestehen darin, daß die Größe der Stichproben für das Training des Modells wesentlich größer sein können. Damit

werden weniger Daten vom Modellbildungsprozess ausgeschlossen. Außerdem reduziert die Kreuzvalidierung die Variabilität, da sie nicht von einer einzelnen Aufteilung einer Stichprobe abhängt.

Bootstrapping stellt eine alternative Methode für die interne Validierung dar.

1.11 Kurze Übersicht über die Prognose bei Melanompatienten

In dieser Arbeit werden neben simulierten auch reale Daten verwendet. Wichtig für die Bewertung der Analysen mit neuronalen Netzen ist es, über Daten von möglichst vielen Fällen zu verfügen [75].

Ähnlich wie bei anderen malignen Erkrankungen ist das Stadium der wichtigste prognostische Faktor. Gegenwärtig werden in Deutschland bei ca. 90-95% der Patienten mit Melanom keine weiteren Metastasen bei der Erstdiagnose festgestellt. Die Prognose für Melanompatienten ohne Metastasierung liegt bei einer 5-Jahres-Überlebensrate von 79%.

Prognoseschätzungen, basierend auf klinischen und histologischen Parametern, geben zusätzliche Informationen zu den herkömmlichen Systemen für die Stadieneinteilung. Das ist vor allen Dingen wichtig für die Auswahl von Patienten für adjuvante Behandlungen oder für die Festlegung von Nachsorgestrategien, die auf einer individuellen Prognoseschätzung beruhen.

Verschiedene Prognosemodelle wurden mit dem Ziel vorgestellt, eine möglichst gute Vorhersage des Krankheitsverlaufs zu erreichen und um Patientengruppen mit niedrigem und hohem Risiko für ein Fortschreiten der Erkrankung zu definieren.

Die Überlebenswahrscheinlichkeit für Melanompatienten wird von klinischen und histologischen Tumormerkmalen beeinflusst. Die meisten Studien konnten einen wichtigen Einfluß der Tumordicke nach Breslow zeigen. Weitere wichtige prognostische Faktoren sind der Invasionslevel nach Clark, die histologisch gesicherte Ulzeration, der histologische Subtyp des Melanoms, Geschlecht, Alter und die anatomische Lokalisation des Primärtumors. Außerdem wurden u.a. noch die Mitoserate und tumorinfiltrierende Lymphozyten in einigen Studien als prognostische Faktoren identifiziert.

In neueren Arbeiten wird diskutiert, ob eine wenig invasive Operation eines speziellen Lymphknotens zu diagnostischen Zwecken (sentinel lymph node biopsie) zu einer weiteren Verbesserung der Prognoseschätzung führt [15] [149].

Aktuell verwendete Stadieneinteilungen, wie z.B. der UICC (International Union against Cancer) oder dem AJCC (American Joint Committee on Cancer) basieren auf wenigen, in den meisten Studien immer wieder bestätigten Faktoren: Vorhandensein von Metastasen (Satellitenmetastasen, Intransitmetastasen, regionale Lymphknotenmetastasen und Fernmetastasen) und bei Nichtvorhandensein von Metastasen zusätzlich Tumordicke und Invasionslevel. Auch bei metastasierten Melanomen werden weitere Faktoren für die Stadieneinteilung berücksichtigt. Auf die Prognose des metastasierten Melanoms wird aber im Rahmen dieser Arbeit nicht eingegangen.

In den folgenden Abschnitten werden Definitionen und Einteilungen beim Melanom der Haut erläutert.

Das maligne Melanom (schwarzer Hautkrebs) ist ein besonders bösartiger Tumor, der von den pigmentbildenden Zellen (Melanozyten oder Nävuszellen) der Haut, seltener der Schleimhaut, der Aderhaut des Auges oder den Hirnhäuten ausgehen kann. Maligne Melanome können sich spontan auf vorher völlig normaler Haut oder auf dem Boden eines vorbestehenden Muttermals (Nävuszellnävus, Leberfleck) entwickeln.

Ausgangspunkt des Wachstums der Tumorzellen ist in den meisten Fällen die Basalschicht der Oberhaut (Stratum basale der Epidermis). Je nach Ausbreitungsbestreben (Autonomiegrad) der malignen Tumorzellen kann das Wachstum zunächst

- horizontal-intraepidermal erfolgen (Lentigo-Maligna, in-situ-Melanom),
- horizontal-intraepidermal und superfiziell-dermal (superfiziell spreitendes Melanom) und
- vertikal-exo-endophytisch (noduläres Melanom) erfolgen.

Das maligne Melanom wird aufgrund dieser feingeweblichen (histologischen) Klassifikationen in nach Art des Wachstums unterscheidbare Typen unterteilt:

- Superfiziell spreitendes Melanom (SSM)
- Noduläres Melanom (NM)
- Lentigo-maligna-Melanom (LMM)
- Akrolentiginöses Melanom (ALM)
- Sonstige Melanome (z.B. Aderhaut, Schleimhaut)

Bei ca. 2% aller Patienten mit malignen Melanom kann der Primärtumor trotz ausführlicher Diagnostik nicht gefunden werden. In diesem Fall spricht man von einem primär okkulten Melanom.

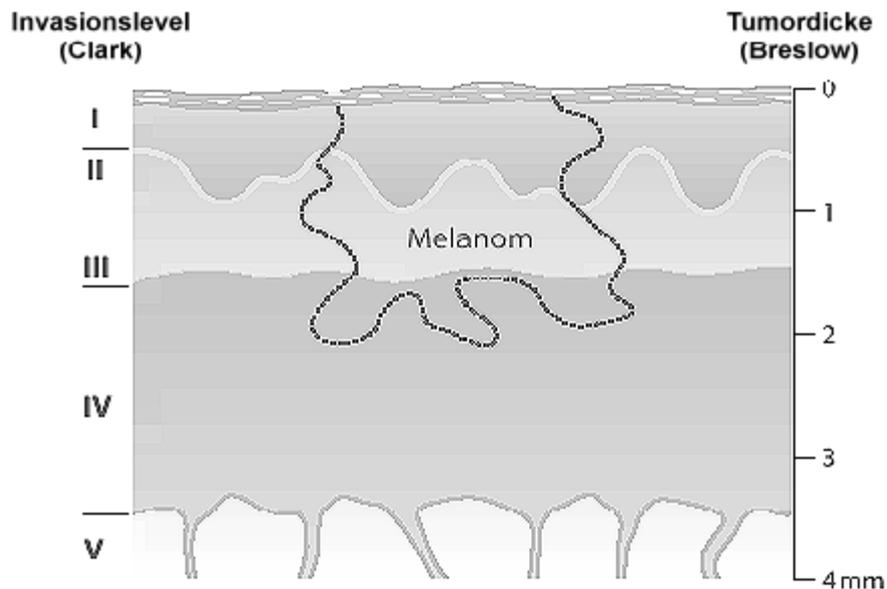


Abbildung 6: Modell der Haut mit den verschiedenen Hautschichten; Darstellung des Invasionslevels nach Clark und der Tumordicke nach Breslow (modifiziert nach Fritsch [67])

Im Rahmen der feingeweblichen (histologischen) Untersuchung des malignen Melanoms wird das Tiefenwachstum absolut in Millimeter (Tumordicke nach Breslow[20]) oder relativ in Bezug auf die Hautschichten (Invasionslevel nach Clark[36]) angegeben.

Tabelle 3: Kategorien für die TNM-Klassifikation

pT	Primärtumor	
pTx	Primärtumor kann nicht beurteilt werden	
pT0	Kein Primärtumor (Primärtumor okkult)	
pT1	Tumordicke ≤ 0,75 mm	bei fehlender Tumordicke: Level II
pT2	Tumordicke 0,76 - 1,5 mm	bei fehlender Tumordicke: Level III
pT3	Tumordicke 1,51 - 4,0 mm	bei fehlender Tumordicke: Level IV
pT4	Tumordicke > 4,0 mm	bei fehlender Tumordicke: Level V
pTa*	Satelliten-Metastasen innerhalb von 2 cm vom Primärtumor (bzw. Lokalrezidiv nach Entfernung mit Sicherheitsabstand)	
pTb*	In-transit-Metastasen vor der regionären Lymphknotenstation	

* Die Attribute a und b sollten, soweit vorhanden, jedem pT (pT1-4) hinzugefügt werden.

N	Regionäre Metastasen (regionäre Lymphknoten)
N1	Metastasen ≤ 3 cm in reg. Lymphknoten
N2	Metastasen > 3 cm in reg. Lymphknoten

Darüber hinaus ist die Zahl der befallenen Lymphknoten anzugeben.

M	Fernmetastasierung
M1(a)	Befall von Haut, Subkutis oder Lymphknoten jenseits der regionären LK
M1(b)	Viszerale Metastasen

Tabelle 4: Stadieneinteilung des malignen Melanoms und 5-Jahres-Überlebenswahrscheinlichkeit

Stadien		10-Jahres-Überleben		
I a	pT ₁ (≤0,75 mm)	N ₀	M ₀	> 95%
I b	pT ₂ (0,76-1,5 mm)	N ₀	M ₀	ca. 90%
II a	pT ₃ (1,5-4,0 mm)	N ₀	M ₀	ca. 70%
II b	pT ₄ (> 4,0 mm)	N ₀	M ₀	ca. 50%
III a	pT _a , pT _b	N ₀	M ₀	ca. 30%
III b	jedes pT	N ₁ , N ₂	M ₀	ca. 20%
IV	jedes pT	Jedes N	M ₁	ca. 5%

In Tabelle 3 ist eine Übersicht zur Einteilung in Stadien angegeben. Diese Einteilung entspricht den Empfehlungen der Deutschen Dermatologischen Gesellschaft.

In Tabelle 4 sind zu den Stadien die zu erwartenden Überlebenswahrscheinlichkeiten angegeben[68].

1.12 Fragestellungen und Hypothesen

1. Ist es möglich, unter Verwendung künstlicher neuronaler Netze bessere Prognosemodelle zu finden, als bei der Verwendung der etablierten statistischen Verfahren und Modelle?
2. Können künstliche neuronale Netze im Sinne einer black-box für die Datenanalyse verwendet werden?
3. Werden mit den neuen statistischen Verfahren Ideen zur explorativen Datenanalyse aus dem Bereich der neuronalen Netze übernommen? Werden statistische Verfahren für die Bewertung der Ergebnisse neuronaler Netze verwendet?
4. Können künstliche neuronale Netze analog zur statistischen Regressionanalyse verwendet werden?
5. Sind künstliche neuronale Netze als Methode für die Datenanalyse geeignet?
6. Lässt sich mit Analysemethoden aus dem Bereich der *Künstlichen Intelligenz* eine bessere Qualität der Prognoseschätzung erreichen als mit statistischen Methoden?