

4 Diskussion

Künstliche neuronale Netze werden zunehmend als Ergänzung der statistischen Werkzeuge angesehen. Übersichtsarbeiten zur Verwendung neuronaler Netze für statistische Zwecke gibt es u.a. von B. Ripley [7, 132]. Zu den verschiedenen Methoden bei der Simulation neuronaler Netze gibt es zahlreiche wissenschaftliche Arbeiten [21, 120, 175].

Schwierig ist die Motivation einzuschätzen, warum künstliche neuronale Netze für statistische Analysen verwendet werden. Sicherlich gibt es dafür verschiedene Ursachen:

- weil die Berechnung machbar geworden ist
- weil nichtlineare Methoden 'modern' sind
- weil geringe Abweichung von der Linearität zu universellen Approximatoren führt
- weil damit Geld zu verdienen ist bzw. Forschungsprojekte gefördert werden

Gegen einen Einsatz künstlicher neuronaler Netze spricht die Tatsache, daß die Idee, künstliche neuronale Netze zu verwenden, dem Ausgangspunkt der Entwicklung statistischer Methoden widerspricht. Rechenmaschinen und Computer wurden ja gerade deshalb entwickelt, weil die Datenanalyse allein mit dem menschlichen Verstand so schlecht funktionierte. Dieses Argument zählt allerdings nicht, wenn künstliche neuronale Netze als flexible mathematische Methoden angewendet werden, wie beispielsweise in dieser Arbeit.

Immerhin hat die Anwendung nichtlinearer Methoden zu neuen Disziplinen wie *data mining*, *KDD* (knowledge discovery in databases), *machine learning* und *neural computation* geführt.

Die Anwendung künstlicher neuronaler Netze für die Analyse von Verlaufsdaten mit zensierten Beobachtungen beschränkt sich bisher im wesentlichen auf medizinische und wirtschaftliche Anwendungen. Daher sind auch die meisten Impulse für die Entwicklung und Überprüfung entsprechender Verfahren aus der medizinischen Informationsverarbeitung und Biometrie gekommen. Im Bereich der Sozialwissenschaften ist bisher keine Anwendung von künstlichen neuronalen Netzen für die Analyse von Verlaufsdaten zu finden.

In der statistischen und medizinischen Literatur werden die neuen Verfahren unterschiedlich diskutiert. Überwiegend optimistische Einschätzungen kommen aus dem anwendungsorientierten Bereich. Zum Teil sehr kritische Meinungen stammen aus überwiegend an der statistischen Theorie arbeitenden Institutionen.

4.1 Beispiele für die Anwendung künstlicher neuronaler Netze bei der Analyse von Verlaufsdaten

4.1.1 Simulierte Daten

Simulierte Daten haben den Vorteil, daß sich damit Modelleigenschaften sehr schön demonstrieren lassen. Sie haben aber auch den Nachteil, daß sie häufig genau auf eine Fragestellung zugeschnitten wurden und daß solche Daten in der Realität nicht vorkommen. Unter diesem Gesichtspunkt sind die Ergebnisse von Analysen mit simulierten Daten immer nur bedingt gültig.

Bei den in dieser Arbeit untersuchten simulierten Daten konnte gezeigt werden, daß das CART-Modell und das künstliche neuronale Netz sich von der Prognosequalität nicht unterscheiden, auch wenn es leichte Vorteile zugunsten der Analyse mit dem künstlichen neuronalen Netz gab. Das Cox-PH-Modell fiel in der Prognose-

sequalität leicht ab. Die Betrachtung von Interaktionen führte zu einer Verbesserung der Prognosequalität im Cox-PH-Modell. In diesem Fall hat die Analyse mit einem künstlichen neuronalen Netz zu dem Ergebnis geführt, daß in den Daten mehr Informationen enthalten sind, als durch das ursprüngliche Cox-PH-Modell beschrieben wurden.

Wenn das Wissen, wie die Daten generiert wurden, eingebracht wird, läßt sich die Qualität der Prognose-schätzung mit dem Cox-PH-Modell weiter verbessern. Es sollten also immer vorhandene Hintergrundinfor-mationen bei der Modellbildung berücksichtigt werden.

4.1.2 Daten von Melanopatienten

4.1.2.1 Prognostische Faktoren bei Melanopatienten

Der wichtigste Prognosefaktor beim malignen Melanom ist, wie bei anderen Tumoren, die Tumorausbreitung, oder häufig synonym verwendet, das Tumorstadium. Zur Tumorausbreitung gehören die Tumormasse des Primärtumors und das Auftreten von regionären Metastasen oder Fernmetastasen. Die Prognose bei Vorlie-gen eines Primärtumors ohne erkennbare Metastasierung kann als günstig angesehen werden. Bei Auftreten einer Metastasierung wird die Prognose als insgesamt ungünstig angesehen.

Mindestens 90 % aller malignen Melanome werden in Deutschland derzeit im Stadium des Primärtumors ohne erkennbare Metastasierung diagnostiziert [68]. Deshalb ist die Prognoseschätzung bei primären malignen Melanomen für therapeutische Entscheidungen und für die Planung der Nachsorge von großer Bedeutung.

Im Unterschied zu vielen anderen Tumoren wird für die Einordnung in verschiedene Tumorkategorien (pT) nicht der größte Durchmesser des Primärtumors gemessen, sondern die vertikale Eindringtiefe des Tumors (Tumordicke nach Breslow) als der entscheidende Parameter [20]. Auf der Tumordicke und der Tumoraus-breitung beruht die TNM-Einteilung der UICC sowie eine modifizierte TNM-Stadieneinteilung nach Empfeh-lungen der Deutschen Dermatologischen Gesellschaft [123].

Das Hinzuiehen des Invasionslevels nach Clark ist nur bei dünnen malignen Melanomen für die Unterschei-dung zwischen solchen mit einem Invasionslevel II und denen mit Invasionslevel III oder mehr von prognosti-scher Bedeutung [69]. Aus diesem Grund empfiehlt die Arbeitsgemeinschaft Dermatologische Onkologie und die Deutsche Dermatologische Gesellschaft, den Invasionslevel nach Clark nicht mehr routinemäßig zur Sta-dieneingruppierung heranzuziehen, wie es nach den Empfehlungen der UICC geschieht.

Ein weiteres histologisches Merkmal, das eine Bedeutung in der Prognoseschätzung hat, ist der histologische Subtyp des malignen Melanoms. Daneben haben auch klinische Faktoren, wie das Geschlecht und die Lokali-sation des Tumors, eine signifikante prognostische Bedeutung.

Die Lokalisation des Primärtumors als prognostischer Faktor wird in der Literatur kontrovers diskutiert. Es existieren verschiedene Vorschläge für die Definition von Gruppen mit erhöhtem Metastasierungs- und Ster-berisiko (7,62). Auf den Ersterhebungsbögen des Zentralregisters Malignes Melanom werden die Lokalisatio-nen Gesicht, sonstiger Kopf, Hals, Brust, Rücken, Unterbauch, Gesäß, Genitoanalregion, Oberarm, Unterarm, Hand, Oberschenkel, Unterschenkel, Fuß, Occult (verborgen, Primärtumor nicht auffindbar) und Schleimhaut unterschieden. Als prognostisch ungünstig erwies sich die Tumorklassifikation am Thorax, oberen Arm, Hals und behaarten Kopf, die mit der Abkürzung TANS (upper Trunk, upper Arm, Neck and Scalp) zusammenge-faßt wurde.

Die wichtigsten prognostischen Faktoren beim primären malignen Melanom ohne Metastasen sind nach neue-ren Studien [27] folgende [2, 68, 69, 76, 78, 81, 88, 104, 105, 111, 126, 138, 147, 152, 162]

- die Tumordicke nach Breslow

- der Invasionslevel nach Clark
- der klinisch-histologische Typ
- das Geschlecht
- die Tumorklassifikation (ungünstige Prognose für oberen Rumpf, Oberarme, Hals und behaarten Kopf).
- das Vorhandensein einer histologisch beschriebenen Ulzeration.

Diese Merkmale wurde immer wieder in verschiedenen Analysen beschrieben. Hinzu kommen möglicherweise neue Faktoren, die auf Entwicklungen der Biotechnologie der letzten Jahre beruhen und für die bisher keine langfristigen Ergebnisse vorliegen.

Bisher hat auch die Hinzuziehung von immunhistologischen Markern die Prognoseschätzungen nicht weiter verbessern können. Häufig sind diese Marker stark mit der Tumordicke korreliert, wie z.B. der Proliferationsmarker (Ki-67) [157].

4.1.2.2 Einordnung der Ergebnisse der Analyse mit künstlichen neuronalen Netzen im Vergleich zu statistischen Methoden

Im Vergleich zwischen künstlichen neuronalen Netzen und statistischen Methoden konnten die neuronalen Netze bei diesen Daten keine Vorteile erbringen. Bereits die CART-Analyse liefert Hinweise auf die Ursache. Die häufig angenommene Komplexität in der Struktur der Daten lag bei den hier analysierten Melanomdaten nicht vor. Die Tumordicke erwies sich klar als wichtigster Faktor. Die Hinzunahme weiterer Faktoren brachte nur beim Cox-PH-Modell eine minimale Verbesserung der Prognoseschätzung. Andererseits muß auch gesagt werden, daß die Ergebnisse der künstlichen neuronalen Netze nur wenig schlechter waren als die der statistischen Methoden.

4.1.3 Anwendungsbeispiele aus der medizinischen Forschung

Schwarzer analysierte die Anwendung künstlicher neuronaler Netze im Zeitraum vom 1991 bis 1995. In der Zwischenzeit sind zahlreiche weitere Anwendungen für Prognoseprobleme in der Literatur beschrieben.

Bellotti et al. [12] beschrieben 1997 die Anwendung von künstlichen neuronalen Netzen bei Patienten mit Lungenkrebs. In ihrer Schlußfolgerung formulierten sie vorsichtig, daß neuronale Netze für individuelle Prognoseschätzungen nützlich sein können. Gleichzeitig formulierten sie die Forderung, die bisherigen Ergebnisse an Dateien mit großen Fallzahlen zu überprüfen.

Von Biganzoli et al. stammt [14] der 1998 entwickelte Ansatz einer partiellen logistischen Regression unter Verwendung künstlicher neuronaler Netze. Durch Benutzung des Zeitintervalls als Eingabe für das Netz ist es möglich, diskrete Hazards als bedingte Wahrscheinlichkeiten zu schätzen. Die Autoren gehen wiederum von einer höheren Qualität der Prognoseschätzung bei komplexen Daten aus. Leider wurde in dieser Arbeit kein überzeugendes Beispiel dafür geliefert.

In einer Übersichtsarbeit beschrieb Bostwick 1998 den Einsatz neuronaler Netze für die Prognoseschätzung als vielversprechend [17]. Ohne konkrete Beispiele zu liefern, folgerte er, daß künstliche neuronale Netze mit größerer Genauigkeit bei der Kombination mehrerer Faktoren arbeiten. In der Schlußfolgerung ging er davon aus, daß künstliche neuronale Netze innerhalb der nächsten Jahre zu den Standardwerkzeugen bei der Datenanalyse gehören werden.

Burke et al. [26] kamen 1997 zu der Schlußfolgerung, daß mit künstlichen neuronalen Netzen die Genauigkeit der Prognoseschätzung verbessert werden kann. Im Vergleich, der zu diesen Ergebnissen geführt hat, unter-

suchten die Autoren allerdings das TNM-System zur Stadieneinteilung und ein künstliches neuronales Netz. Hätten die Autoren zum Vergleich ein Cox-PH-Modell anstelle des künstlichen neuronalen Netzes verwendet, wären sie mit hoher Wahrscheinlichkeit zu ähnlichen Ergebnissen gekommen.

Bottaci et al. untersuchten 1997 die Prognose von Patienten mit kolorektalen Karzinomen [18]. Bei dieser Analyse, die in der renommierten Zeitschrift 'Lancet' erschienen ist, haben die Autoren auf einen Vergleich mit statistischen Methoden verzichtet. Sie demonstrierten lediglich die Benutzung künstlicher neuronaler Netze für die Datenanalyse. Ausgehend davon, daß komplexe biologische Zusammenhänge nicht mit linearen statistischen Methoden beschrieben werden können, wurden in dieser Arbeit künstliche neuronale Netze verwendet.

Brown et al. analysierten 1997 den Einfluß der Handhabung von Zensierungen auf die Qualität der Prognose-schätzung [22]. Sie kamen zu dem Schluß, daß künstliche neuronale Netze bei richtiger Behandlung der Zensierungen eine robuste Methode zur modellunabhängigen Datenanalyse sind.

De Laurentis et al. untersuchten 1996 den Zusammenhang zwischen der histologischen Charakteristik des Primärtumors und dem Grad des axillären Lymphknotenbefalls bei Brustkrebs [42]. Bei dieser Analyse wurde kein überzeugender Vergleich zwischen statistischen Methoden und der Analyse mit künstlichen neuronalen Netzen vorgenommen. Es wurde aber gezeigt, daß mit neuronalen Netzen gute Ergebnisse erzielt werden können.

Doyle et al. untersuchten 1994 und 1995 die Ergebnisse von Lebertransplantationen [51]. Während sie bei der ersten Analyse bei 155 Implantationen noch 95% der Fälle richtig klassifizierten, waren es bei einer späteren Auswertung mit größeren Fallzahlen nur noch ca. 90%. Ein Vergleich mit statistischen Methoden stellten die Autoren nicht dar.[52]

In einer Publikation von 1998 verglichen Geddes et al. [70] die Prognoseschätzung erfahrener Ärzte mit den Ausgaben eines künstlichen neuronalen Netzes. Bei dieser Analyse wurden die Daten von 54 Patienten unter Berücksichtigung von 7 Faktoren analysiert. Die Daten wurden nicht in Trainings- und Testdaten geteilt. Die Autoren benutzten ein 'jack-knife-sampling' zur internalen Datenvalidierung. Zum Vergleich wurden die trainierten Daten wiederverwendet. Die Ergebnisse dieser Arbeit sind also sehr fragwürdig.

Groves et al. verglichen 1999 künstliche neuronale Netze und Cox-PH-Modelle [72]. Sie kamen zu dem Ergebnis, daß es keinen Unterschied in der Prognosequalität zwischen künstlichen neuronalen Netzen und dem Cox-PH-Modell gab. Den Wert der künstlichen neuronalen Netze beschrieben die Autoren damit, daß mit den verwendeten Modellen der maximale Informationsgehalt aus der Datenbank extrahiert wurde.

Die Arbeit von Luxhoj und Shyr von 1997 [103] ist eines der wenigen Beispiele, bei denen künstliche neuronale Netze für die Prognoseschätzung bei zensierten Daten außerhalb der Medizin angewendet wurden. Das Ergebnis dieser, aus dem Bereich der Ingenieurwissenschaften stammenden Arbeit ist eine bessere Prognose künstlicher neuronaler Netze gegenüber einem Cox-PH-Modell. Allerdings wurden bei dieser Arbeit nicht alle Möglichkeiten der statistischen Analyse ausgeschöpft.

Mariani et al. verglichen 1997 ein Cox-PH-Modell mit einem künstlichen neuronalen Netz bei der Untersuchung zu Prognosefaktoren bei einer speziellen Form des Brustkrebses [106]. Sie kamen zu dem Ergebnis, daß der prädiktive Wert von Cox-PH-Modell und künstlichem neuronalen Netz relativ niedrig war. Daraus schlossen sie, daß die tatsächlich relevanten prognostischen Faktoren nicht in den zur Verfügung stehenden Daten vorhanden waren.

Naguib veröffentlichte 1998 mit unterschiedlichen Autoren drei Arbeiten zur Analyse zensierter Verlaufsdaten mit künstlichen neuronalen Netzen [113, 114, 115] Obwohl die Daten von relativ häufig vorkommenden Tumo-

ren untersucht wurden, bezogen sich die Arbeiten auf 38, 41 und 81 Patienten. Mit diesen kleinen Fallzahlen lassen sich bei der Verwendung künstlicher neuronaler Netze zur Untersuchung mehrerer prognostischer Faktoren keine gültigen Ergebnisse erzielen.

In der Studie von Ortiz et al. wurde 1995 ein künstliches neuronales Netz für die Schätzung der 1-Jahres-Mortalität verwendet [124]. Dabei wurde das Überleben als binäre Größe betrachtet, die Beobachtungsdauer und Zensierungen wurden bei dieser Analyse von 95 Patienten nicht berücksichtigt. Berechtigterweise wird in der Schlußfolgerung vorsichtig formuliert, daß weitere Forschungsarbeit auf diesem Gebiet nötig ist.

Von Ragde et al. wurde 1998 eine Arbeit zur 10-Jahres-Rezidivfreiheit von Patienten mit Prostatakarzinom veröffentlicht [128]. Das erstaunliche bei dieser Arbeit ist die Verwendung eines künstlichen neuronalen Netzes als 'normale' Methode zur Datenanalyse. Ein Vergleich mit statistischen Methoden wurde trotz der eigentlich kleinen Fallzahl von 152 Patienten nicht vorgenommen.

Zernikow et al. versuchten 1998 mit künstlichen neuronalen Netzen eine Risikoabschätzung bei Frühgeborenen vorzunehmen [176]. Bei der Analyse einer relativ großen Fallzahl von 890 Kindern kamen sie zu dem Ergebnis, daß wegen der hohen Zahl von Fehlklassifikationen auch künstliche neuronale Netze keine echte Hilfe bei individuellen Therapieentscheidungen darstellen.

Einen ausführlichen Vergleich zwischen künstlichen neuronalen Netzen und statistischen Methoden lieferte R. Ripley am Beispiel von Brustkrebspatientinnen [133, 134]. Bei der Analyse von realen Daten konnte Ripley nie einen echten Vorteil von künstlichen neuronalen Netzen gegenüber statistischen Methoden zeigen. Die Ergebnisse der künstlichen neuronalen Netze sind mit den Resultaten moderner statistischer Verfahren vergleichbar.

Einen speziellen Ansatz zur Verwendung künstlicher neuronaler Netze verfolgt Ohno-Machado [118, 119]. Die Verwendung hierarchischer neuronaler Netze sollte eine erhöhte Flexibilität der Modellanpassung bringen. Allerdings ist die vorgeschlagene Methode für den Umgang mit zensierten Zeitangaben nicht optimal.

Singson et al. benutzten ein probabilistisches neuronales Netz für die Analyse beim therapeutischen Vorgehen von kolorektalen Adenokarzinomen [150]. Sie kamen beim Vergleich von multipler logistischer Regressionsanalyse und neuronalen Netzen zu dem Ergebnis, daß die neuronalen Netze nützliche Ergebnisse liefern, diese Ergebnisse aber an Daten mit großen Fallzahlen validiert werden müssen. In dieser Studie wurden die Daten von 99 Patienten analysiert, wobei die Angaben von 59 Patienten als Testdaten verwendet wurden.

Edwards et al. [55] untersuchten mit einem künstlichen neuronalen Netz die Daten von 81 Herzinfarkt-Patienten und kamen zu dem Schluß: Die Analyse mit künstlichen neuronalen Netzen nutzt die Daten für die Vorhersage der Mortalität effizienter als die multiple logistische Regressionsanalyse. Ein gut validiertes neuronales Netz könne im klinischen Management eine Rolle spielen.

Qureshi et al. untersuchten klinisch-pathologische und molekulare Marker bei Blasenkrebs [127]. Für diese prognostische Untersuchung wurden keine Überlebens- oder rezidivfreien Zeiten betrachtet, sondern das Problem auf ein binäres Problem für einen festen Zeitpunkt reduziert. Die Ergebnisse der Analyse mit neuronalen Netzen werden positiv gesehen und schneiden im Vergleich zur klinischen Erfahrung von Ärzten besser ab.

De Laurentiis et al. verglichen die Ergebnisse eines neuronalen Netzes mit der TNM-Klassifikation bei Brustkrebspatientinnen [41]. Bei diesem Vergleich wurde die Prognose für einen festen Zeitpunkt definiert. Zensierungen wurden nicht berücksichtigt. Zum Vergleich wurde ein sehr einfaches, lineares Modell verwendet (TNM-Klassifikation).

Bei den zuletzt zitierten Veröffentlichungen handelt es sich um Arbeiten, die in den letzten zwei Jahren verfaßt wurden. Obwohl die Methodik für eine aus statistischer Sicht angemessene Auswertungsstrategie existiert, werden bekannte und bereits kritisierte Fehler wiederholt.

Diese Auswertung der Literatur zeigt, daß immer wieder nach einem allgemeingültigen Auswertungsansatz gesucht wird. Teilweise wurden die statistischen Methoden bei weitem nicht ausgeschöpft, es wurde aber versucht, eine neue Methode anzuwenden.

4.2 Kritik an der Verwendung künstlicher neuronaler Netze bei der Analyse von Verlaufsdaten

Schwarzer et al. [148] haben die gegenwärtig verbreitete Verwendung künstlicher neuronaler Netze für die Prognoseschätzung im Zeitraum 1991-1995 systematisch analysiert und scharf kritisiert. Sie vergleichen die Situation, mit der gegen Ende der siebziger Jahre, als Statistikpakete für Nichtspezialisten zugänglich wurden und bei der Diskriminanzanalyse häufig Fehler auftraten. Sie beschreiben mehrere Arten von Fehlern, die mehrfach bei wissenschaftlichen Veröffentlichungen zu finden sind:

- Fehler bei der Schätzung der Wahrscheinlichkeit für Fehlklassifikationen

Die Auswahl des gültigen Modells anhand der Testdaten führt zu einer Unterschätzung der Fehlklassifikationsrate. Deshalb muß gefordert werden, daß die Modellauswahl entweder anhand von Validierungsdaten erfolgt, oder die Modelle a priori festgelegt werden. Außerdem war bei vielen Arbeiten die Fallzahl bei den Testdaten zu klein. So konnten instabile Schätzer entstehen.

- Anpassung implausibler Funktionen

Häufig werden künstliche neuronale Netze im Sinne einer black-box verwendet. Falls jedoch sehr viele versteckte Einheiten verwendet werden, kommt es zu einer Überanpassung des Netzes. Zur Feststellung der Überanpassung gibt es keine Universallösung. Es sollte aber immer versucht werden, die unter Statistikern anerkannte Regel zu verwenden, daß die Zahl der Parameter nie mehr als ein Zehntel der beobachteten Ereignisse beträgt.

- Fehlerhafte oder fehlende Beschreibung der Komplexität des Netzwerks

Diese Regel ist auch einigen Anwendern von neuronalen Netzen bekannt. Um Kritik zu umgehen oder wegen falscher Interpretation der Methode wurde häufig nur die Zahl der benutzten Faktoren, nicht aber die Zahl der verwendeten Neuronen genannt. Daher wurde die Wahrscheinlichkeit für eine Fehlklassifikation häufig unterschätzt.

- Benutzung unpassender statistischer Methoden für Vergleiche

Verschiedene Autoren kommen zu dem Ergebnis, daß künstliche neuronale Netze statistischen Methoden überlegen sind, da sie in der Lage sind, komplexe, nichtlineare Funktionen zwischen Eingabedaten und Ergebnissen zu beschreiben.

Die Netze müssen mit Methoden verglichen werden, die ähnlich flexibel sind wie z.B. CART, GAM (generalisierte additive Modelle) oder Regressionsmodelle mit quadratischen Interaktionstermen.

Zum Beispiel haben Ripley [134] und Vach [165] solche Vergleiche durchgeführt und sind zu dem Ergebnis gekommen, daß sich die Methoden in ihren Diskriminierungseigenschaften nicht unterscheiden.

- Falsche Anwendung künstlicher neuronaler Netze für die Analyse von Verlaufsdaten

In der medizinischen Literatur sind mehrere Beispiele für die Anwendung künstlicher neuronaler Netze zu finden, die zensierte Fälle nicht in ausreichender Weise berücksichtigen. So werden z.B. zensierte Fälle

bei der Analyse ausgeschlossen [25, 26], oder es wird die Zeit in Intervalle unterteilt, wobei die Intervallgröße nicht in das Modell aufgenommen wird [130, 131].

Die bis hier aufgeführten Punkte sind der Standpunkt von Schwarzer et al. [148]. Diese Kritikpunkte beschreiben die aktuelle Situation treffend. Sicherlich ist es problematisch, ein Modell zu finden, daß in der Lage ist, zensierte Fälle zu berücksichtigen. Der von Faraggi [59] stammende Vorschlag zur Benutzung des Cox-PH-Modells als Grundlage für ein Modell, bei dem die Maximum-Likelihood-Schätzung durch ein künstliches neuronales Netz ersetzt wurde, scheint den besten Kompromiß aus Flexibilität und Berücksichtigung der Besonderheiten zensierter Daten darzustellen.

4.2.1 Probleme beim Vergleich von Prognosemodellen

Beim Vergleich der Prognosemodelle mit unterschiedlichen methodischen Ansätzen sollten die Aspekte berücksichtigt werden, die auch bei der Konstruktion statistischer Tests zum Vergleich verschiedener Modelle einer Methode verwendet werden. Dazu zählen ein Maß für die Klassifikationsgüte, ein Maß für die Zahl der verwendeten Faktoren (Freiheitsgrade) und die analysierte Fallzahl. Zum gegenwärtigen Zeitpunkt ist kein Test bekannt, der diese Faktoren für unterschiedliche Methoden berücksichtigt.

4.2.2 Methoden zum Vergleich von Prognosemodellen

Da bisher keine 'fertigen' Verfahren für den Vergleich von Prognosemodellen existieren, wird ersatzweise auf eine Rangkorrelation zurückgegriffen, bei der die Klassifikationsgüte unabhängig von der Wahl des Modells bestimmt wird. Auch die in dem speziellen Fall der Untersuchung von Verlaufsdaten nötige Berücksichtigung zensierter Zeiten ist möglich. Unter Verwendung dieser Rangkorrelation können Konfidenzintervalle konstruiert werden. Eine andere Möglichkeit bietet die Ableitung von ROC-Kurven. Für ROC-Kurven wurden bereits Methoden entwickelt, um Unterschiede statistisch abzusichern. Im Rahmen dieser Arbeit war aber ein Vergleich der ROC-Kurven nicht nötig, da die Modelle sehr dicht beieinander lagen.

4.3 Neue Erkenntnisse durch die Verwendung künstlicher neuronaler Netze bei der Analyse von Verlaufsdaten

Etwas enttäuschend muß die Frage nach neuen wissenschaftlichen Erkenntnissen durch die Verwendung künstlicher neuronaler Netze beantwortet werden. In der im Rahmen dieser Arbeit vorgenommenen Literaturanalyse konnte in keiner Arbeit überzeugend gezeigt werden, daß die Verwendung künstlicher neuronaler Netze zu neuen wissenschaftlichen Erkenntnissen geführt hat, die mit statistischen Methoden nicht gefunden worden wären. Auch die Analysen in dieser Arbeit haben keine neuen Erkenntnisse zur Prognoseschätzung bei Melanompatienten gebracht.

Andererseits wurden in der wissenschaftlichen Diskussion um die Verwendung neuronaler Netze für die Datenanalyse die Schwachstellen der neuen Methoden aufgezeigt.

4.4 Weitere Forschung

Aus dem Bereich der künstlichen Intelligenz gibt es weitere interessante Ansätze, die auch für die Analyse von Verlaufsdaten in Frage kommen. Dazu gehören genetische Algorithmen und auf Fuzzy-Logik basierende neuronale Netze. Diese Methode wurde auch bereits für Klassifikationsaufgaben eingesetzt [50, 54].

Die bisherigen Diskussionen um die verschiedenen Modelle haben dazu geführt, daß in verschiedenen Arbeitsgruppen begonnen wurde, die Methoden für den Modellvergleich zu verfeinern. Bisher fehlt ein echter Test, um z.B. die Ergebnisse eines Cox-PH-Modell mit denen eines künstlichen neuronalen Netzes zu vergleichen. Die mehrfach verwendete ROC-Statistik lieferte nur Anhaltspunkte.