

Aus der Klinik für Urologie
der Medizinischen Fakultät Charité – Universitätsmedizin Berlin

DISSERTATION

**Validierung eines artifiziellen neuronalen Netzwerkes zur
Vorhersage von Prostatabiopsieergebnissen anhand
prospektiver Daten**

Zur Erlangung des akademischen Grades
Doctor medicinae (Dr. med.)

vorgelegt der Medizinischen Fakultät
Charité – Universitätsmedizin Berlin

von

Martin Bender

aus Krefeld

Gutachter: 1. Priv.-Doz. Dr. med. C. Stephan
2. Priv.-Doz. Dr. T. Brinkmann
3. Priv.-Doz. Dr. med. F. Christoph

Datum der Promotion: 24. Februar 2012

Inhaltsverzeichnis

Inhaltsverzeichnis	I
Tabellenverzeichnis	III
Abbildungsverzeichnis	IV
Abkürzungsverzeichnis	V
1 EINLEITUNG	1
1.1 Allgemeine Angaben zum Prostatakarzinom (PCa)	1
1.2 Diagnostische Parameter des PCa	4
1.2.1 Die digital-rektale Untersuchung	4
1.2.2 Das Prostataspezifische Antigen (PSA)	4
1.2.2.1 Freies PSA und komplexiertes PSA	6
1.2.2.2 Weitere molekulare Formen des fPSA	7
1.2.3 Steigerung der PCa-Spezifität des PSA	8
1.2.3.1 Altersspezifische Referenzgrenzen	8
1.2.3.2 Die PSA-Anstiegsgeschwindigkeit	9
1.2.3.3 Die PSA-Dichte	10
1.2.4 Neuere Marker	10
1.2.4.1 Neuere Serummarker	10
1.2.4.2 Neuere Urinmarker	12
1.3 Multivariate Modelle in der PCa-Diagnostik	13
1.3.1 Artificielle neuronale Netzwerke (ANN)	13
1.3.1.1 ANN in der PCa-Diagnostik	14
1.3.2 Nomogramme	14
1.3.3 Vergleich von ANN und Nomogrammen	15
2 AUFGABENSTELLUNG	16
3 MATERIAL UND METHODEN	18
3.1 Das Patientenkollektiv	18
3.1.1 Die „Interne Validierung“ Kohorte	18
3.1.2 Die „ProstataClass“ Kohorte	18
3.2 PSA-Bestimmung	19
3.3 Klinische Untersuchungen	19
3.3.1 Prostatavolumenbestimmung	19
3.3.2 Digital-rektale Untersuchung	19
3.4 Artificielle neuronale Netzwerke	20
3.5 Statistische Methoden	22
4 ERGEBNISSE	24
4.1 Vergleich der Kohorten	24
4.2 Ergebnisse der ROC-Analysen	27
4.3 Spezifitätenvergleich bei 90 % und 95 % Sensitivität	31
4.4 Gültigkeit der PCa-Vorhersage	31

5	DISKUSSION	37
6	ZUSAMMENFASSUNG	47
7	LITERATUR	49
	Eidstattliche Erklärung	i
	Lebenslauf	ii
	Publikationsliste	iii
	Danksagung	iv

Tabellenverzeichnis

Tabelle 1:	Vergleich der Ursprungskohorten der fünf artifiziellen neuronalen Netzwerke	22
Tabelle 2:	Vergleich der „Interne Validierung“ Kohorte und der „ProstataClass“ Kohorte	26
Tabelle 3:	ROC-Kurvenanalyse für tPSA, %fPSA und die fünf artifiziellen neuronalen Netzwerke	27

Abbildungsverzeichnis

- Abb. 1:** Molekulare Formen des PSA **8**
- Abb. 2** Schematische Darstellung eines artifiziiellen neuronalen Netzwerkes zur Vorhersage des PCa-Risikos **21**
- Abb. 3:** ROC-Kurven von tPSA, %fPSA und ANN „ProstataClass“ (ANNpc) für die „Interne Validierung“ Kohorte mit den entsprechenden AUC-Werten **29**
- Abb. 4:** ROC-Kurven von tPSA, %fPSA und ANN „Interne Validierung“ (ANNiv) für die „Interne Validierung“ Kohorte mit den entsprechenden AUC-Werten **29**
- Abb. 5:** ROC-Kurven von ANN „Gesamt“ (ANNall), Multi Layer Perceptron (ANNmlp) und ANN „Hamburg“ (ANNham) für die „Interne Validierung“ Kohorte mit den entsprechenden AUC-Werten **30**
- Abb. 6:** Kalibrierungen mit Ausgleichskurven des ANN „ProstataClass“ (ANNpc) für die „ProstataClass“ Kohorte (a) und für die „Interne Validierung“ Kohorte (b) **32**
- Abb. 7:** Kalibrierungen mit Ausgleichskurven des ANN „Interne Validierung“ (ANNiv) für die „ProstataClass“ Kohorte (a) und für die „Interne Validierung“ Kohorte (b) **33**
- Abb. 8:** Kalibrierungen mit Ausgleichskurven des ANN „Gesamt“ (ANNall) für die „ProstataClass“ Kohorte (a) und für die „Interne Validierung“ Kohorte (b) **34**
- Abb. 9:** Kalibrierungen mit Ausgleichskurven des Multi Layer Perceptrons (ANNmlp) für die „ProstataClass“ Kohorte (a) und für die „Interne Validierung“ Kohorte (b) **35**
- Abb. 10:** Kalibrierungen mit Ausgleichskurven des ANN Hamburg (ANNham) für die „ProstataClass“ Kohorte (a) und für die „Interne Validierung“ Kohorte (b) **36**

Abkürzungsverzeichnis

%fPSA	Quotient des fPSA zum tPSA (fPSA/tPSA)
Abb.	Abbildung
ACT	Alpha-1-Antichymotrypsin
ANN	artifizielles neuronales Netzwerk
ANNall	ANN „Gesamt“ (ANN basierend auf den Daten beider Kohorten)
ANNham	ANN Hamburg (ANN basierend auf Daten von Chun et al.)
ANNiv	ANN „Interne Validierung“
ANNmlp	Multi Layer Perceptron (ANN von Finne et al.)
ANNpc	ANN “ProstataClass”
ANXA3	Annexin A3
AUC	Fläche unter der ROC-Kurve (engl.: area under the curve)
BPH	benigne Prostatahyperplasie
bPSA	benignes PSA (auch BPH assoziiertes PSA, BPHA)
cPSA	komplexiertes PSA (Summe aus PSA-ACT und PSA-API)
Da	Dalton
DRU	digital-rektale Untersuchung
ERSPC	European Randomized Study of Screening for Prostate Cancer (dt.: europäische randomisierte Studie zum Prostatakrebs-Screening)
fPSA	freies PSA
GOLM	Golgi-Membranprotein 1
ICC	Intraklassenkorrelationskoeffizient (engl.: intra class correlation coefficient)
IGF-1	insulinanaloger Wachstumsfaktor 1 (engl.: insulin-like growth factor)
inPSA	inaktives PSA
KLK	humanes Kallikrein
<i>KLK</i>	humanes Kallikrein-Gen
LOO	Leave-One-Out-Kreuzvalidierung
LR	logistische Regression
MIC-1	Makrophagen-Inhibitor-Zytokin-1
MIF	Makrophagenmigrations-inhibierender-Faktor

N	Anzahl der Patienten
NEM	kein Anhalt auf Malignität (engl.: no evidence of malignancy)
PCa	Prostatakarzinom (engl.: prostate carcinoma)
PCA3	Prostatakrebs Antigen 3 (engl.: prostate cancer antigen 3)
proPSA	Vorstufe des enzymatisch aktiven PSA
PSA	Prostataspezifisches Antigen
PSA-A2M	an Alpha-2-Makroglobulin gebundenes PSA
PSA-ACT	an Alpha-1-Antichymotrypsin gebundenes PSA
PSA-API	an Alpha-1-Proteaseinhibitor gebundenes PSA
PSAV	PSA-Anstiegsgeschwindigkeit (engl.: PSA-Velocity)
PSP94	prostatishes Sekretionsprotein 94
ROC	receiver operating characteristic
RV	residuelle Varianz
tPSA	Gesamt-PSA (engl.: total PSA)
TRUS	transrektaler Ultraschall
USA	Vereinigte Staaten von Amerika (engl.: United States of America)

1 Einleitung

1.1 Allgemeine Angaben zum Prostatakarzinom (PCa)

In Deutschland ist das Prostatakarzinom (PCa) die häufigste bösartige Neubildung beim Mann. Am PCa versterben hierzulande gegenwärtig etwa 11 000 bis 12 000 Männer pro Jahr. Insgesamt stellt das PCa mit 10 % die dritthäufigste Todesursache unter den Neoplasien dar. Mit 69 Jahren liegt das mittlere Erkrankungsalter in Deutschland im Bereich von Neoplasien allgemein, wobei die Inzidenz altersabhängig ist und das Risiko zu erkranken mit zunehmendem Lebensalter wächst. Der Erkrankungsgipfel liegt zwischen dem 75. und 79. Lebensjahr (1).

Insgesamt zeigte das PCa im Zeitraum von 1980 bis 2005 eine Zunahme seiner Inzidenz von mehr als 100 % (1). Dies ist vermutlich der erhöhten Lebenserwartung sowie den verbesserten diagnostischen Möglichkeiten zuzuschreiben. Seit der Etablierung des prostataspezifischen Antigens (PSA) als Parameter zur PCa-Vorsorge, werden signifikant mehr inapparente, langsam wachsende Karzinome entdeckt.

In den Vereinigten Staaten von Amerika (USA, engl.: United States of America) wurde 2009 schätzungsweise bei 192 280 Männern ein PCa erstdiagnostiziert. Das PCa ist mit 25 % aller zu erwartenden neoplastischen Neuerkrankungen die häufigste Krebserkrankung beim Mann. Im Jahr 2009 verstarben in den USA schätzungsweise 27 360 Männer an einem PCa. Damit stellt das PCa zusammen mit dem Kolon- und Rektumkarzinom, und nach dem Lungenkarzinom, die zweithäufigste Todesursache unter den Neoplasien des Mannes dar (2).

Bis heute ist sehr wenig über die Ursachen der malignen Entartung der Prostata und mögliche Risikofaktoren für die Entstehung eines PCa bekannt. Als ausreichend belegte Risikofaktoren gelten Alter, ethnische Herkunft und familiäre Belastung. So werden 63 % aller PCa bei Patienten diagnostiziert, die bereits 65 Jahre oder älter sind. Weltweit ist die Inzidenz für Prostatakrebs bei afroamerikanischen Männern, sowie jamaikanischen Männern mit afrikanischer Abstammung am höchsten. In der asiatischen und südamerikanischen Bevölkerung hingegen ist die Inzidenz wesentlich geringer (2;3). Familiäre Belastung konnte in zahlreichen Studien mit einem erhöhten PCa-Risiko assoziiert werden. Als Gründe für die epidemiologisch belegten Risikofaktoren werden unterschiedliche endogene und exogene Ursachen diskutiert

(z.B. diätetische Ursachen, Übergewicht, Rauchen, hormonelle Einflüsse, genetische Disposition und genetische Alterationen), wobei die genauen Zusammenhänge weitgehend unklar sind (4).

Die Erkennung des PCa in einem frühen Stadium ist aufgrund der typischen Lokalisation in der peripheren Zone der Prostata erschwert. Es kommt in der Regel erst im fortgeschrittenen Stadium zu wegweisenden Symptomen wie Dysurie, Nykturie, Pollakisurie oder Anurie. Diese Symptome sind zudem häufiger auf eine benigne Prostatahyperplasie (BPH) als auf ein PCa zurückzuführen. Auch die erektile Dysfunktion, die aufgrund einer mechanischen Schädigung der Nervenfasern entstehen kann, tritt üblicherweise erst im späteren Stadium der Erkrankung auf. In selteneren Fällen werden Hämaturie und Hämatospermie beobachtet (5).

Einige Patienten zeigen bei Erstdiagnose bereits Symptome einer Metastasierung, ohne dass der Primärtumor im Voraus zu Beschwerden geführt hat. Aufgrund der typischen Metastasierungswege können solche okkulten Tumoren durch Skelettschmerzen, Kopfschmerzen, aber auch durch regionale Lymphabflussstörungen und spinale neurologische Ausfälle klinisch auffällig werden.

Die Prognose des PCa ist durch den Grad der Differenzierung der Tumorzellen und die Ausbreitung des Karzinoms beeinflusst. Lokal begrenzte Tumoren, die die Kapsel der Prostata nicht überschreiten, liegen derzeit bei über 90 % der Patienten zum Diagnosezeitpunkt vor. Diese können mit kurativem Ansatz therapiert werden und fünf Jahre nach Diagnosestellung haben diese Patienten eine Überlebenswahrscheinlichkeit von nahezu 100 %. Liegt zum Zeitpunkt der Diagnosestellung bereits ein kapselüberschreitendes Wachstum vor oder finden sich sogar Metastasen, ist die Prognose deutlich schlechter (2;3). Dies macht die Notwendigkeit geeigneter Früherkennungsverfahren deutlich.

Seit Anfang der Siebzigerjahre existieren Vorsorgeprogramme, welche die Verbesserung der PCa-Früherkennung zum Ziel haben. Bis zur klinischen Etablierung des PSA beinhalteten diese Programme lediglich die urologische Anamnese und die digital-rektale Untersuchung (DRU). In den USA wurde das PSA 1994 von der „Food and Drugs Administration“ zur Diagnose des PCa zugelassen, nachdem es bereits 1986 als Marker zur Verlaufskontrolle zugelassen worden war (6). Infolge zeigte sich ein deutlicher Anstieg der Anzahl diagnostizierter Neuerkrankungen, da seitdem mehr

inapparente Karzinome entdeckt werden. Außerdem stieg der Anteil der lokal begrenzten Tumoren bei Diagnosestellung (7).

Fraglich bleibt, ob die Screeningprogramme tatsächlich die Mortalität senken und wie groß die Anzahl der überbehandelten PCa-Patienten ist. Die Ergebnisse der European Randomised Study of Screening for Prostate Cancer (ERSPC) zeigen eine Reduktion der PCa bedingten Sterblichkeit durch Teilnahme an einem Screeningprogramm um 20 % (8). Nach der mathematischen Korrektur für die Störgrößen Nichtteilnahme und Kontaminierung wird die Mortalität sogar um 31 % gesenkt (9). Dem widersprechen die Ergebnisse einer randomisierten Studie aus den USA, die keine signifikante Senkung der Mortalität für die Screening-Population nachweisen konnte (10). Als mögliche Erklärung für die fehlende Mortalitätssenkung diskutieren die Autoren der amerikanischen Studie das hohe Maß an Vorsorge in der Kontrollgruppe, die höheren PSA-Grenzwerte verglichen mit der ERSPC, eine Selektions-Bias durch der Studie vorangegangene PSA-Tests, sowie eine stetige Verbesserung der Therapieoptionen über den Zeitraum der Studie (10). Zudem wurden in der Studie aus den USA nicht alle auffälligen PSA-Befunde histologisch abgeklärt. Im Vergleich zur ERSPC war die Teilnehmerzahl geringer und die Verlaufsbeobachtung kürzer.

Der Senkung der Sterblichkeit steht das Problem der Überbehandlung gegenüber, das aus der steigenden Anzahl inapparent entdeckter Karzinome resultiert, die in vielen Fällen langsam wachsen und nicht die spätere Todesursache des Patienten sind. Dies betrifft in besonders hohem Maße ältere Patienten. Draisma et al. berichten, dass jährliche Screeninguntersuchungen bei der Altersgruppe von 55 bis 67 Jahre zu einer Überbehandlungsrate von 46 % bis 57 % führen (11). Das Fortführen von Screeninguntersuchungen bis zum 75. Lebensjahr führt mindestens zu zwei klinisch irrelevanten Diagnosen pro klinisch relevanter Diagnose (11). Es erleiden folglich wenigstens zwei Patienten eine mit Nebenwirkungen und Risiken behaftete Überbehandlung, damit ein PCa-Patient adäquat behandelt werden kann. Dies entspricht einer Überbehandlungsrate von mindestens 66 %.

1.2 Diagnostische Parameter des PCa

1.2.1 Die digital-rektale Untersuchung

Derzeit empfiehlt die American Cancer Society Männern mit einer verbleibenden Lebenserwartung von mindestens zehn Jahren, sich ab dem 50. Lebensjahr im Rahmen der Krebsvorsorge einmal jährlich einer DRU, sowie einer PSA-Wert-Bestimmung zu unterziehen. Risikopatienten (bei einem engen Verwandten ist die Diagnose PCa gesichert) und Hochrisikopatienten (bei mindestens zwei engen Verwandten ist die Diagnose PCa gesichert) wird bereits ab dem 45. bzw. 40. Lebensjahr eine jährliche Vorsorgeuntersuchung mit DRU und PSA-Wert-Bestimmung empfohlen (12).

Nach der deutschen EBM-Leitlinie aus dem Jahr 2002 wird hierzulande Männern zwischen dem 50. und 75. Lebensjahr eine jährliche Vorsorgeuntersuchung mit DRU und PSA-Wert-Bestimmung empfohlen. Bei familiärer Belastung werden jährliche Vorsorgeuntersuchungen bereits ab 45 Jahren empfohlen (13).

Ein suspekter Tastbefund liegt vor, wenn die Vorsteherdrüse eine unregelmäßige höckerige Oberfläche und Verhärtungen aufweist. Fehlende Verschieblichkeit gegen die Rektumschleimhaut, Asymmetrie der beiden Prostatalappen, sowie mangelnde Abgrenzbarkeit gegenüber dem umliegenden Gewebe deutet zusätzlich auf einen malignen Prozess. Eine vergrößerte, aber prall elastische Drüse ist hingegen typisch für eine BPH. Nachteile der DRU sind die subjektive Bewertung durch den Untersucher, sowie die geringe Sensitivität bei niedriggradigen Neoplasien. Etwa 70 % der Tumoren, die aufgrund eines suspekten Tastbefundes diagnostiziert werden, sind bereits kapselüberschreitend oder sogar bereits metastasiert. Für Patienten mit einem derartigen Befund stehen nur eingeschränkt kurative Therapieoptionen zur Verfügung.

1.2.2 Das Prostataspezifische Antigen (PSA)

Das PSA ist derzeit der wichtigste Tumormarker für das PCa. Die Bestimmung erfolgt routinemäßig zur Karzinomvorsorge, zum Tumorstaging und zur Rezidivkontrolle (14;15).

Das PSA wurde von Hara et al. bereits Ende der Siebzigerjahre des 20. Jahrhunderts

im Seminalplasma nachgewiesen und unter dem Namen "γ-seminoprotein" beschrieben (16). Im Jahr 1979 wurde es von Wang et al. erstmals als prostataspezifische Substanz und potentieller Tumormarker für das PCa erwähnt (17). Nachdem PSA im Serum von Patienten mit PCa nachgewiesen werden konnte (18) und schließlich geeignete PSA-Bestimmungsverfahren verfügbar waren, entwickelte es sich zum derzeit wichtigsten und am weitesten verbreiteten Tumormarker in der Urologie.

Das PSA wird der Gruppe der Kallikreine zugeordnet und auch als humanes Kallikrein (KLK) 3 bezeichnet. Die bisher bekannten 15 humanen Kallikreine werden von einer gemeinsamen Region des Chromosoms 19 codiert und weisen neben den genetischen Gemeinsamkeiten auch funktionelle Ähnlichkeit als Serinproteasen auf (19).

Biochemisch handelt es sich beim PSA um ein Glycoprotein aus 237 Aminosäuren und einer Kohlenhydratkette mit einem Gesamtmolekulargewicht von 28.430 Da. In den Epithelzellen der Prostata wird ein Präprotein mit 261 Aminosäuren synthetisiert, welches nach Prozessierung noch 244 Aminosäuren aufweist und als proPSA bezeichnet wird. Dieses wird ins Drüsenlumen sezerniert und durch proteolytische Abspaltung von weiteren 7 Aminosäuren entsteht die enzymatisch aktive Serinprotease PSA (20). Die enzymatisch aktive Form des PSA kann Peptidbindungen an den Carboxylgruppen von Leucin und Tyrosin spalten (21) und verflüssigt die Samenflüssigkeit nach der Ejakulation durch Proteolyse der Semenogeline I und II, sowie des Fibronektins (22).

Beim gesunden Mann gelangt das PSA zum überwiegenden Teil ins Sperma, wodurch die physiologische PSA-Konzentration im Sperma ca. 0,5 – 3 g/L beträgt. Die PSA-Serumkonzentration liegt bei gesunden Männern normalerweise unter 4 µg/L, da bei einer gesunden Drüse nur sehr geringe Mengen in den systemischen Kreislauf gelangen. Sie ist damit ungefähr eine Million Mal geringer als die Konzentration im Sperma (22).

Im Falle eines PCa kommt es im Drüsengewebe zu einer Architekturstörung, die mit einer Zerstörung der Basalmembran der Epithelzellen einhergeht. Dies kann in der Folge dazu führen, dass größere Mengen PSA ins Serum gelangen. Dieser Prozess ist jedoch keinesfalls spezifisch für das PCa. Auch nach Trauma, Prostatitis und bei starkem Wachstum der Drüse, im Rahmen der BPH, kommt es zur Störung der physiologischen Sekretion. Zudem führen Geschlechtsverkehr, transurethrale Dauer-

katheterisierung, sowie diagnostische oder therapeutische Manipulationen an der Drüse zu einer Erhöhung der PSA-Serumkonzentration. Diese nicht neoplastischen Einflüsse auf den PSA-Wert führen zu einem niedrigen positiv prädiktiven Wert und folglich zu einer großen Anzahl unnötiger Folgeuntersuchungen. Dies schränkt den Nutzen des PSA als alleinigen Screeningparameter stark ein (23).

1.2.2.1 Freies PSA und komplexiertes PSA

In den frühen Neunzigerjahren wiesen Lilja et al. und Stenman et al. unabhängig voneinander zwei molekulare Hauptformen des PSA nach (24;25). Etwa 60 - 95 % des Gesamt-PSA (engl.: total PSA, tPSA) ist an Alpha-1-Antichymotrypsin (ACT) gebunden und wird als PSA-Alpha-1-Antichymotrypsin (PSA-ACT) bezeichnet, während 5 - 35 % des PSA ungebunden als freies PSA (fPSA) vorliegen (24-26). Hierbei ist besonders der Quotient des fPSA zum tPSA (fPSA/tPSA), das so genannte prozentuale freie PSA (%fPSA), von diagnostischer Bedeutung. PCa-Patienten weisen in der Regel einen niedrigeren %fPSA-Wert auf als BPH-Patienten (24-26). Dies hat zu einer weiteren Verbesserung der diagnostischen Spezifität des PCa-Screenings geführt, wodurch etwa 20 - 25 % der unnötigen Prostatabiopsien bei PSA-Werten zwischen 4 – 10 µg/L vermieden werden können (27-31). Eine Meta-Analyse von Lee et al. kommt zu dem Ergebnis, dass die Anwendung von %fPSA im tPSA-Bereich von 4 – 10 µg/L nur bei sehr niedrigen %fPSA-Werten einen zusätzlichen Nutzen zum tPSA zeigt (32). Im niedrigen tPSA-Bereich unter 4 µg/L wird die Relevanz des %fPSA für die Diagnostik des PCa kritisch diskutiert. Einige Autoren konnten auch in diesem Bereich einen Vorteil für %fPSA gegenüber tPSA zeigen (33-35). Andere Autoren hingegen lassen Einschränkungen für die Verwendbarkeit von %fPSA bezüglich der Vermeidung von unnötigen Biopsien in diesem tPSA-Bereich erkennen, konstatieren allerdings Vorteile für die Vorhersage des PCa-Risikos (36) bzw. für die Einschätzung der Aggressivität eines Karzinoms (37). Ungeachtet einiger Einschränkungen ist das %fPSA derzeit der wichtigste Parameter zur Ergänzung des PSA und hat eine gewisse Bedeutung in der Routinediagnostik erhalten (38).

Neben dem PSA-ACT konnten noch die beiden weiteren PSA-Komplexe PSA-Alpha-2-Makroglobulin (PSA-A2M) und PSA-Alpha-1-Protease Inhibitor (PSA-API) identifiziert werden, welche jedoch wegen ihres geringen Anteils am tPSA und der

schwierigen Nachweisverfahren nur von eingeschränktem klinischen Nutzen sind (39-41). Als komplexiertes PSA (cPSA) bezeichnet man die Summe von PSA-ACT und PSA-API, die mit einem direkten Nachweisverfahren bestimmt werden kann. Hierbei wird PSA-A2M allerdings nicht gemessen (41). Das cPSA konnte das %fPSA jedoch nicht ersetzen, wie anfänglich vermutet wurde (42). Lediglich der Quotient aus cPSA und tPSA (cPSA/tPSA) zeigte einen mit %fPSA vergleichbaren Nutzen zur Ergänzung des PSA-Wertes (43;44).

1.2.2.2 Weitere molekulare Formen des fPSA

Im menschlichen Serum können verschiedene molekulare Formen des fPSA gemessen werden: das bereits erwähnte proPSA, das benigne oder BPH assoziierte PSA (bPSA oder BPHA) und das inaktive PSA (inPSA) (45).

Das proPSA bildet die Vorstufe des fPSA und liegt in verschiedenen Formen vor. Das (-7)proPSA ist die native Vorstufe des fPSA. Es weist verglichen mit fPSA sieben zusätzliche Aminosäuren auf, die bei der Aktivierung zum fPSA abgespalten werden. Neben dem (-7)proPSA können verkürzte Formen mit weniger zusätzlichen Aminosäuren nachgewiesen werden, die als (-1)- bis (-6)proPSA bezeichnet werden, je nach Anzahl der zusätzlichen Aminosäuren (46). Das (-2)proPSA und das (-4)proPSA sind mit dem PCa assoziiert (47;48). Seit die Möglichkeit zum einzelnen Nachweis von (-2)proPSA besteht, kann das Verhältnis von (-2)proPSA zum fPSA (%(-2)proPSA) zur Diagnose des PCa genutzt werden (49). Neuere Daten mit einem automatisierten Messsystem zum %(-2)proPSA zeigen eine Steigerung der diagnostischen Genauigkeit verglichen mit %fPSA (50;51). Die (-7)- und (-5)proPSA Formen, für die ebenfalls ein Testsystem existiert, konnten keine Verbesserung der diagnostischen Genauigkeit im Vergleich zu tPSA und %fPSA zeigen (52;53).

Neben den verschiedenen Formen des proPSA existiert eine weitere molekulare Form des fPSA, das bPSA. Das bPSA entspricht in seiner Länge dem fPSA, ist jedoch durch eine innere Spaltung in der Aminosäurenkette inaktiviert. Dieser Bruch findet sich zwischen den Aminosäuren Lys182 und Lys145 (54). Ein erhöhter bPSA-Wert ist mit der BPH assoziiert. Dennoch konnte weder bPSA, noch der Quotient bPSA/tPSA die diagnostische Genauigkeit des %fPSA verbessern (55).

Das inPSA ist eine weitere molekulare Form des fPSA. Es besitzt keine enzymatische Aktivität, obschon es strukturell intakt ist und keine Spaltungen innerhalb des Moleküls aufweist. Die Serumkonzentration von inPSA zeigt keine signifikanten Unterschiede zwischen PCa-Patienten und Patienten ohne Anhalt für Malignität (engl.: no evidence of malignancy, NEM), allerdings ist eine Erhöhung des Quotienten inPSA/fPSA mit einem PCa assoziiert (56;57).

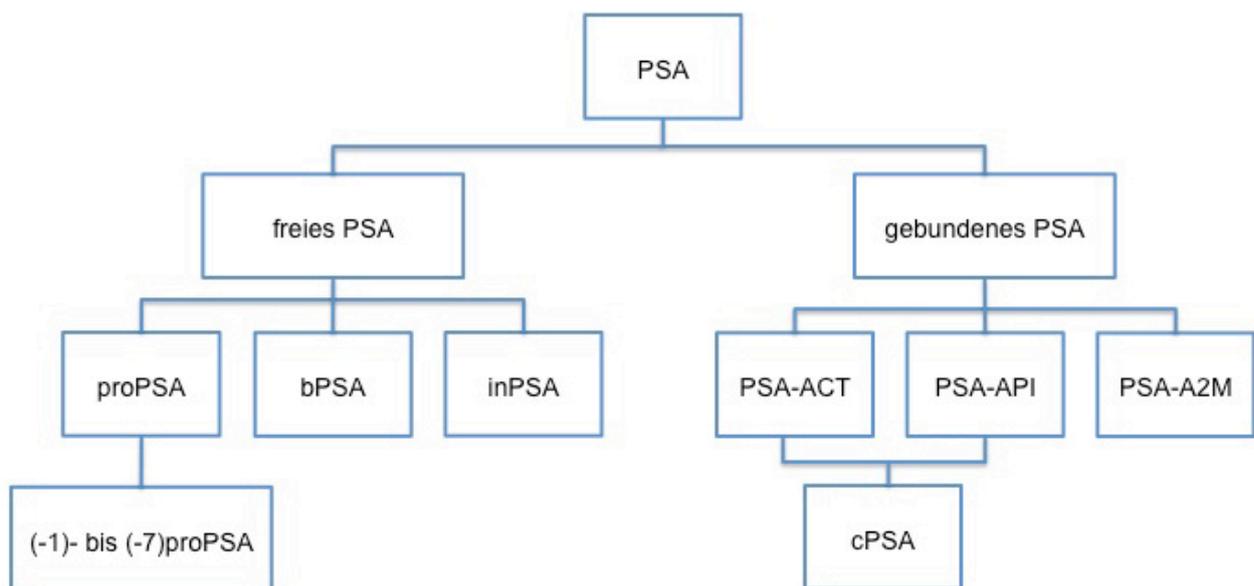


Abb. 1: Molekulare Formen des PSA (abgewandelt nach einer Abbildung von Stephan et al. (58)).

1.2.3 Steigerung der PCa-Spezifität des PSA

1.2.3.1 Altersspezifische Referenzgrenzen

Mit zunehmendem Lebensalter nimmt das Volumen der Vorsteherdrüse zu, wodurch es zu einem Anstieg der PSA-Werte im Serum kommt. Dabei ist die altersabhängige Steigerung des mittleren Prostatavolumens im Wesentlichen auf die Prävalenz der BPH in der jeweiligen Altersgruppe zurückzuführen. Eine statistisch signifikante positive

Korrelation zwischen Lebensalter und PSA-Werten im Serum ($r = 0,43$, $p < 0,0001$) konnte von Oesterling et al. nachgewiesen werden (59). Die PSA-Werte steigen um etwa $0,04 \mu\text{g/L}$ pro Jahr, dies entspricht einer Steigerung von etwa $3,2\%$. Weil diese Erhöhung der PSA-Werte jedoch nicht auf einer malignen Entartung basiert, sondern einem physiologischen Prozess des Alterns gleichkommt, ist es wichtig, dies in die Interpretation der PSA-Werte mit einzubeziehen. Zur Vereinfachung der klinische Anwendung gelten folgende altersspezifische Referenzgrenzen: für Männer unter 50 Jahren gilt ein Grenzwert von $2.5 \mu\text{g/L}$, mit 50 - 59 Jahren liegt der Referenzbereich unter $3.5 \mu\text{g/L}$, mit 60 - 69 Jahren unter $4.5 \mu\text{g/L}$ und mit 70 - 79 Jahren unter $6.5 \mu\text{g/L}$ (59).

Obgleich die Anwendung altersspezifischer Referenzgrenzen zu einer Verringerung der Sensitivität führt, sind diese bei der Interpretation der PSA-Konzentration im Serum wegen der verbesserten Spezifität von Bedeutung (60). In Leitlinien zur PCa-Diagnostik fanden die alterabhängigen PSA-Grenzwerte bisher jedoch kaum Beachtung.

1.2.3.2 Die PSA-Anstiegsgeschwindigkeit

Zur weiteren Verbesserung der Aussagekraft des PSA-Wertes dient die PSA-Anstiegsgeschwindigkeit. Dieser auch als PSA-Velocity (PSAV) bezeichnete Parameter beschreibt den Anstieg der PSA-Konzentration in der Zeiteinheit. Eine erhöhte PSAV kann auf das Vorliegen eines PCa hindeuten und erklärt sich aus dem malignen Wachstum der Drüse, sowie der verstärkten PSA-Sekretion ins Blut. Carter et al. fanden signifikant erhöhte PSAV-Werte bei PCa-Patienten verglichen mit BPH-Patienten und Patienten einer Kontrollgruppe (61). Verschiedene Folgestudien mit kürzeren Messabständen erbrachten widersprüchliche Ergebnisse (62-64). Die Tatsache, dass große intraindividuelle Unterschiede bezüglich der PSA-Serumwerte bestehen, schränkt den Nutzen der Methode ein. Es ergibt sich die Notwendigkeit mehrerer Messungen über einen Zeitraum von 2 Jahren, damit eine annehmbare diagnostische Genauigkeit erreicht wird (65;66). Zusätzlich kann durch Nutzung verschiedener Testsysteme die Aussagekraft der PSAV gemindert werden (67).

1.2.3.3 Die PSA-Dichte

Die Grundannahme, dass die Erhöhung des PSA-Wertes beim PCa vornehmlich auf die Störung der „Prostata-Blut-Schranke“, bei einer BPH hingegen auf die stark erhöhte Zellzahl zurückzuführen ist, hat zur Einführung der PSA-Dichte (engl.: PSA-Density) in die PCa-Diagnostik geführt. Die PSA-Dichte ist definiert als der Quotient tPSA/Prostatavolumen. Folglich deutet eine Erhöhung der PSA-Dichte auf einen malignen Prozess hin. Diese Korrelation konnte durch verschiedene Studien belegt werden (68-70). Einschränkungen dieses Verfahrens stellen die geringe Genauigkeit der Volumenbestimmung mittels transrektalen Ultraschalls (TRUS), sowie das intraindividuell unterschiedliche Epithel-Struma-Verhältnis dar. Die Volumenbestimmung mittels TRUS ist stark untersucherabhängig und bereits geringe Messfehler führen zu erheblichen Abweichungen des errechneten Prostatavolumens vom tatsächlichen Prostatavolumen (71). Das Verhältnis von Epithel zu Struma ist für die Aussagekraft der PSA-Dichte mitentscheidend, da das Struma der Prostata kein PSA bildet. Die starken intraindividuellen Unterschiede bezüglich dieses Verhältnisses sind eine entscheidende Fehlerquelle und schränken den Nutzen der PSA-Dichte ein (72).

1.2.4 Neuere Marker

1.2.4.1 Neuere Serummarker

Neben dem humanen Kallikrein 3 (KLK3), welches besser bekannt ist unter der Bezeichnung PSA, sind noch 14 weitere Kallikreine bekannt (19). Die 15 Kallikrein-Gene werden als *KLK1* bis *KLK15* (*kursiv*) bezeichnet, während die dazugehörigen Proteine mit KLK1-15 benannt werden (73). Obgleich der Nutzen von KLK2 im Rahmen der PCa-Diagnostik gesichert werden konnte (74;75), wird KLK2 nicht in der Routinediagnostik des PCa eingesetzt. Daneben gibt es Hinweise darauf, dass das KLK11 in Kombination mit tPSA und %fPSA zur weiteren Spezifitätssteigerung der Früherkennung des Prostatakarzinoms beitragen kann (76). Aus derzeitiger Sicht könnten einzelne Vertreter der Kallikreinfamilie für die Prognoseabschätzung des PCa Bedeutung erlangen (77;78).

Weitere Marker sind das Makrophagen-Inhibitor-Zytokin-1 (MIC-1) und der Makrophagenmigration-inhibitorische-Faktor (MIF). MIC-1 zeigte zunächst in einer

Arbeit von Brown et al. eine signifikante Steigerung der Spezifität verglichen mit %fPSA, woraus eine mögliche Vermeidung von 27 % unnötiger Prostatabiopsien resultieren könnte (79). Ein verringerter MIC-1 deutet der Studie zufolge auf ein erhöhtes PCa-Risiko. Eine weitere Studie von Stephan et al. konnte die signifikante Steigerung im Vergleich zu %fPSA nicht reproduzieren, zeigte aber für MIC-1, MIF und KLK11 als Parameter in artifiziellen neuronalen Netzwerken (ANN) eine signifikante Spezifitätssteigerung (80). Die Datenlage in Bezug auf MIF ist widersprüchlich: Eine Studie mit 509 Probanden zeigte erhöhte MIF-Werte bei den PCa-Patienten verglichen mit BPH-Patienten (81). In weiteren Studien wurden signifikant erniedrigte MIF-Werte bei den Patienten mit histologisch gesichertem PCa verglichen mit den gesunden Patienten bzw. den BPH-Patienten gefunden (80;82).

Tahir et al. liefern Daten zu einem weiteren potentiellen Marker für das PCa, dem Caveolin-1 (83). Beim Caveolin-1 handelt es sich um ein Protein, das an molekularen Transportprozessen, Signaltransduktion und Zelladhäsion beteiligt ist und dessen Konzentration in 102 PCa-Patienten verglichen mit 107 BPH-Patienten und 81 gesunden Probanden signifikant erhöht war (83). Eine weitere Arbeit derselben Arbeitsgruppe zeigte ein erhöhtes Risiko für ein biochemisches Rezidiv nach radikaler Prostatektomie bei Karzinompatienten mit erhöhten präoperativen Caveolin-1-Werten (84). Die Ergebnisse sind jedoch bislang noch von keiner anderen Arbeitsgruppe reproduziert worden.

Ein potentieller Marker für das Risiko an einem PCa zu erkranken soll laut einer Studie von Stattin et al. der insulinanaloge Wachstumsfaktor 1 sein (engl.: insulin-like growth faktor 1, IGF-1), welcher die Proliferation der Zellen im Tumorgewebe fördert und deren Apoptose hemmt (85). In Folgestudien konnten die positiven Ergebnisse anhand von Screeningdaten dagegen nicht bestätigt werden (86;87). Eine Metaanalyse, die 12 prospektive Studien zusammenfasst, kommt zu dem Ergebnis, dass hohe IGF-1 Serumwerte mit einem moderat erhöhten Risiko an Prostatakrebs zu erkranken verbunden sind (88).

Das prostatistische Sekretionsprotein 94 (PSP94) ist ein potentieller Marker für die postoperative Rezidivprognose und zur Differenzierung des Aggressivitätsgrades des PCa. Erste Ergebnisse aus den Neunzigerjahren zeigten keine Unterschiede bezüglich der PSP94 Serumkonzentrationen zwischen PCa-Patienten, BPH-Patienten und einer

Kontrollgruppe (89). Allerdings konnten in einer aktuelleren Studie mit 1212 Probanden signifikant geringere PSP94-Konzentrationen bei den Karzinompatienten gemessen werden (90). Zusätzlich war bei den PCa-Patienten mit einer aggressiven Form, welche mittels tPSA und %fPSA nicht identifiziert werden konnten, anhand der PSP94-Serumwerte eine Differenzierung möglich (90). Reeves et al. bestätigen, dass freies und gebundenes PSP94, sowie das PSP94-bindende Protein unabhängige Prognoseparameter für das postoperative Rezidiv beim PCa sind (91).

1.2.4.2 Neuere Urinmarker

Das Protein Annexin A3 (ANXA3) zeigt Expressionsunterschiede in gesundem und karzinomatösem Prostatagewebe. Es treten zusätzlich Unterschiede der Expression von ANXA3 zwischen Karzinomgewebe und Gewebe mit intraepithelialer Neoplasie auf (92). Kollermann et al. konnten mittels ANXA3-Bestimmung das Rezidivrisiko von 969 PCa-Patienten ermitteln und die Probanden in eine Hochrisikogruppe und eine Niedrigrisikogruppe einteilen (93). Die Kombination von tPSA und ANXA3 verbessert die diagnostische Genauigkeit verglichen mit dem %fPSA. Zusätzlich hat ANXA3 eine höhere Sensitivität bei Patienten mit unauffälliger DRU und niedrigen PSA-Werten als das tPSA selbst (94).

Im PCa-Gewebe kommt es zu einer etwa 66-fach verstärkten Expression des Prostatakrebs Antigen 3 (engl.: prostate cancer antigen 3, PCA3), bei dem es sich um nicht-kodierende RNA handelt (95). Mehrere Arbeiten konnten eine Korrelation zwischen der Expression von PCA3 und der Wahrscheinlichkeit für ein positives Biopsieergebnis nachweisen (96;97). Darüber hinaus zeigt PCA3 keine Korrelation zum Drüsenvolumen (97). Eine Korrelation mit dem Karzinomvolumen konnte nur teilweise gezeigt werden (98;99), wobei in Studien mit einer geringeren Probandenzahl jedoch keine Korrelation beobachtet wurde (100;101).

Das Golgi-Membranprotein 1 (engl.: golgi membrane protein 1, GOLM1) wird in den Epithelzellen der Prostata exprimiert und lässt sich im Urin nachweisen. In einer Studie von Varambally et al. verbesserte GOLM1 verglichen mit tPSA die Diagnostik des PCa signifikant (102), weshalb es als potentieller zukünftiger Marker in Betracht kommt.

1.3 Multivariate Modelle in der PCa-Diagnostik

Die einzelnen Parameter zur Diagnose des PCa zeigen teilweise nur relativ geringe Spezifitäten. Dies führt zu vielen falsch-positiven Ergebnissen, welche wiederum teure und teilweise invasive Untersuchungen zur Folge haben. Um dieser Problematik entgegenzuwirken, werden verstärkt multivariate Auswertungsmodelle verwendet, die es ermöglichen mehrere Parameter in die Bewertung einzubeziehen. Im Wesentlichen existieren zwei multivariate Modellformen: ANN und auf logistischer Regression (LR) basierende Nomogramme.

1.3.1 Artificielle neuronale Netzwerke (ANN)

ANN wurden bereits in den späten Fünfzigerjahren entwickelt. Zunächst wurden einfache neuronale Netzwerke, sogenannte Perzeptronen von Rosenblatt (103), Widrow und Hoff (104), sowie von Widrow und Sterns (105) konstruiert, die auf das Lösen einfacher linearer Probleme beschränkt waren. Im Jahre 1974 gelang es Werbos, erste nicht-lineare Netzwerke zu entwickeln (106).

Grundsätzlich ähneln sich die Funktionsweisen künstlicher Neuronennetze und biologischer neuronaler Netzwerke. Im biologischen Vorbild findet man eine große Zahl untereinander verschalteter Einzelneurone, die befähigt sind, mittels Dendriten eingehende Signale zu erfassen. Liegen die Eingangssignale in ausreichender Stärke vor, können die Neurone ein Ausgangssignal generieren und über das Axon abgeben. Die Leistungsfähigkeit des Nervensystems liegt hierbei in der Komplexität und Variabilität der Verschaltungen. Das Nervensystem hat die Fähigkeit, die Gewichtung einzelner Verschaltungen zu verändern. Erweisen sich Verschaltungen als richtig bzw. nützlich, wird die Leitfähigkeit erhöht. Erweisen sich Verschaltungen als falsch bzw. nicht-nützlich, wird die Leitfähigkeit reduziert. Dies bildet das biologische Korrelat zum Lernprozess. Im künstlichen Neuronennetz finden sich ebenfalls Verschaltungspunkte, die in verschiedenen Ebenen angeordnet sind. Die Gewichtung der Verbindungen zwischen diesen Knotenpunkten ist, wie beim biologischen Vorbild, entscheidend für die Informationsverarbeitung. Auch im künstlichen Netzwerk müssen die Verschaltungen zunächst in einem Lernprozess gewichtet werden (107). Letztlich werden die Beziehungen zwischen den Eingangs- und Ausgangsparametern durch das Training

vom Netzwerk erlernt. Weil die vielen Vorhersageparameter für das PCa keine linearen Zusammenhänge aufweisen, versprechen künstliche neuronale Netzwerke durch den nicht-linearen Ansatz, besonders gut für die PCa-Vorhersage geeignet zu sein (108).

1.3.1.1 ANN in der PCa-Diagnostik

Seit im Jahre 1994 Snow et al. das erste ANN zur Vorhersage von Biopsieergebnissen und zur Prognose bei PCa konstruiert haben (109), kommen diese multivariaten Modelle zunehmend in der Diagnostik des Prostatakrebs zum Einsatz. Verschiedene Forschungsgruppen haben unterschiedliche Netzwerke erstellt (110-115). Zusammenfassende Arbeiten zeigen eine signifikante Steigerung der Spezifität bei der PCa-Diagnostik (38;116;117). Beispielsweise zeigte das vom Institut für Urologie der Charité – Universitätsmedizin Berlin multizentrisch entwickelte ANN „ProstataClass“ eine Erhöhung der Spezifität verglichen mit %fPSA um 20 - 22 %. Es ließen sich durch die Anwendung dieses auf prospektiven Daten von 1996 - 2001 entwickelten ANN bis zu 30 % unnötige Biopsien der Prostata vermeiden (114).

Nachdem die klinische Einführung von %fPSA die Vorhersage eines PCa bereits deutlich verbessert hatte (29), brachte der Einsatz verschiedener ANN-Modelle, welche %fPSA einbeziehen, eine weitere Spezifitätssteigerung des %fPSA um bis zu 50 % (110-112;115).

1.3.2 Nomogramme

Neben den künstlichen neuronalen Netzwerken sind Nomogramme als weitere multivariate Modelle zur Bewertung des Karzinomrisikos anhand verschiedener Parameter einsetzbar. Es handelt sich bei Nomogrammen um die grafische Darstellung mathematischer Formeln, die es ermöglicht das Ergebnis der Bewertung mehrerer Variablen abzulesen. Hierzu werden die Diagnoseparameter auf verschiedene Skalen aufgetragen. Das Karzinomrisiko kann schließlich auf der Ergebnisskala abgelesen werden. Die üblicherweise für die Krebsdiagnostik verwendeten Nomogramme basieren auf dem statistischen Modell der multivariaten logistischen Regressionsanalyse (118;119).

Das erste Nomogramm zur Vorhersage eines PCa wurde von Eastham et al. im Jahr 1999 entwickelt (120). Dieses Nomogramm ist, trotz seiner hohen diagnostischen Genauigkeit, nur eingeschränkt nutzbar, da es ausschließlich für Patienten mit suspektem Tastbefund und PSA-Werten von 0 – 4 µg/L gilt. Diese Kriterien erfüllt jedoch nur ein kleiner Teil der urologischen Patienten (121). Weitere Nomogramme, die in der Folge von Garzotto et al. und Karakiewicz et al. erarbeitet wurden, weisen eine wesentliche Einschränkung auf: Diese Modelle basieren auf der Sextantenbiopsie, die nicht den heutigen Standards entspricht (122;123). Das Nomogramm von Karakiewicz et al. zeigte bei Anwendung auf eine Population mit erweitertem Biopsieschema eine 7 % geringere diagnostische Genauigkeit. Chun et al. entwickelten und validierten daraufhin Nomogramme anhand von Patienten, bei denen ein erweitertes Biopsieschema genutzt wurde (121;124). Es folgten 2008 weitere Studien mit Nomogrammen, die ein erweitertes Biopsieschema nutzten (125).

1.3.3 Vergleich von ANN und Nomogrammen

In Übersichtsarbeiten und Vergleichsstudien haben einige Autoren versucht, für ANN bzw. Nomogramme Vorteile in der Karzinomvorhersage zu zeigen. Die Ergebnisse sind jedoch sehr widersprüchlich: Einige Übersichtsarbeiten (119;126) und Vergleichsstudien (125;127;128) zeigen Vorteile für die Anwendung von Nomogrammen, welche z.B. eine größere diagnostische Genauigkeit aufweisen als ANN-Modelle. Andere Studien (111-113) und Übersichtsarbeiten (129;130) sehen hingegen Vorteile auf Seiten der ANN bzw. keine Unterschiede. Bei der Betrachtung von Studien mit sehr großen Studienpopulationen zeigt sich in 7 von 8 Studien kein Unterschied zwischen beiden Modellen (129). Dies ist ein starker Hinweis darauf, dass aus mathematisch-methodischer Sicht keines der beiden Modelle zu bevorzugen ist. Vielmehr unterstützen die Daten die These, dass unterschiedliches Abschneiden der Modelle auf verschiedene Eingangsparameter, Populationsunterschiede und Nichtberücksichtigung der Methodenabhängigkeit der PSA-Werte zurückzuführen ist (112;127;128;131).

2 Aufgabenstellung

Im Rahmen der Früherkennung des PCa hat sich die Bestimmung der PSA-Serumkonzentration als Standard etabliert. Allerdings führt die breite Anwendung der PSA-Messung zu einer erheblichen Anzahl falsch-positiver Ergebnisse. Etwa 60 - 80 % aller Prostatastanzbiopsien zeigen keinen malignen Befund (15). Die auf diese Weise entstehenden unnötigen Kosten für das Gesundheitssystem, sowie die individuellen Risiken für den jeweiligen Patienten machen eine weitere Steigerung des positiv prädiktiven Wertes für den PSA-Wert nötig. Hierzu werden neben verschiedenen molekularen Formen des PSA auch multivariate Modelle verwendet. Diverse Studien zeigen für den Einsatz von ANN eine Steigerung der Spezifität der PCa-Vorhersage (110-112;114-116). Externe Validierungen der verwendeten Netzwerke sind jedoch nicht ausreichend vorhanden und die Ergebnisse der wenigen vorhandenen externen Validierungen zeigen im Vergleich zu der Population mit der das jeweilige Modell erstellt wurde meist ein schlechteres Abschneiden für die externe Population (116).

Eine interne Validierung des ersten ANN „ProstataClass“, welches anhand von prospektiven Daten von 859 Patienten im Zeitraum von März 1996 bis März 2001 aufgebaut wurde, zeigte für 85 im Zeitraum von März bis Oktober 2001 separat getestete Patienten vergleichbar gute Daten (132). Während bei der 10fachen Kreuzvalidierung der 859 Patienten die Anzahl der korrekt vom ANN klassifizierten Patienten zwischen 78,5 % und 81 % lag, konnten 69 der 85 separat getesteten Patienten korrekt vom ANN identifiziert werden. Dies entspricht einem Anteil von 81,2 % (132). Aufgrund eines Wechsels des PSA-Testsystems in der Charité wurde anhand von retrospektiven Daten (von 2001 - 2004) ein zweites ANN „ProstataClass“ (ANNpc) für das neue PSA-Testsystem aufgebaut (133).

Ziel der vorliegenden Arbeit ist jetzt, das ANNpc mit unabhängigen, prospektiven Daten von 393 Patienten zu validieren. Ein wesentlicher Aspekt ist dabei die Vergleichbarkeit der retrospektiven und prospektiven Validierungskohorte.

Ferner werden in dieser Arbeit die Ergebnisse des ANNpc bezüglich der Bewertung der „Interne Validierung“ Kohorte und der „ProstataClass“ Kohorte mit den Ergebnissen zweier neuer ANN und zweier Netzwerke, die auf externen Daten von zwei anderen Arbeitsgruppen basieren, verglichen, um Erkenntnisse zum Abschneiden verschiedener

Netzwerke für verschiedene Kohorten zu erlangen.

3 Material und Methoden

3.1 Das Patientenkollektiv

3.1.1 Die „Interne Validierung“ Kohorte

Insgesamt wurden prospektive Daten von 393 Patienten im Zeitraum von Februar 2007 bis Mai 2008 erhoben und ausgewertet. Diese Patienten werden im Folgenden als „Interne Validierung“ Kohorte zusammengefasst. Bei allen Patienten handelte es sich um urologisch eingewiesene Patienten der Klinik und Poliklinik für Urologie der Charité – Universitätsmedizin Berlin, Campus Charité Mitte. Einweisungsgründe waren erhöhte PSA-Werte, BPH-Symptome, der klinische Verdacht auf ein PCa oder ein bereits histologisch gesichertes PCa.

Bei allen 393 Studienteilnehmern erfolgte erstmalig oder wiederholt die Prostatastanzbiopsie (8 - 12fach) zur histologischen Diagnosestellung. Das Prostatavolumen wurde mittels TRUS ermittelt. Das Lebensalter und der DRU-Status, sowie die Laborparameter tPSA und %fPSA waren die weiteren Parameter, die von allen Patienten eruiert worden sind.

Zum Zeitpunkt der Erhebung der Daten hatte keiner der Patienten medikamentöse oder operative Therapien in Bezug auf eine Erkrankung der Prostata erfahren.

Bei 229 der insgesamt 393 Patienten wurde anhand der Stanzbiopsie die Diagnose PCa histologisch gesichert. Bei den übrigen 164 Patienten konnte mittels Stanzbiopsie kein Anhalt auf eine maligne Erkrankung gefunden werden.

3.1.2 Die „ProstataClass“ Kohorte

Bei den 787 Patienten der „ProstataClass“ Kohorte handelt es sich um Patienten, deren Daten retrospektiv ermittelt wurden. Die verwendeten Serumproben waren zwischen 2001 und 2004 gesammelt und archiviert worden.

3.2 PSA-Bestimmung

Für die „Interne Validierung“ Kohorte wurden die verwandten Blutproben vor bzw. mindestens drei Wochen nach jeglicher diagnostischer oder therapeutischer Manipulation an der Prostata entnommen und innerhalb eines Zeitraumes von zwei bis drei Stunden nach Blutentnahme zentrifugiert. Die auf diese Weise gewonnenen Serumproben wurden entweder am gleichen Tag analysiert oder bei -20° C gelagert und innerhalb von maximal 48 Stunden analysiert.

Die Bestimmung von tPSA und fPSA erfolgten an vollautomatischen Immunoassay-Analysegeräten unter Beachtung der Betriebsanweisungen der Hersteller. Hierbei kam der Elecsys 2010[®]-Test der Firma Roche (Roche Diagnostics GmbH, Penzberg, Deutschland) zum Einsatz (133). Alle Serumproben wurden vorab jeglicher diagnostischer oder therapeutischer Maßnahmen entnommen und bei -80 °C gelagert, bis sie schließlich analysiert wurden. Analytische Details der Testsysteme wurden bereits 2006 ausführlich beschrieben (67).

3.3 Klinische Untersuchungen

3.3.1 Prostatavolumenbestimmung

Die Prostatavolumenbestimmung wurde mittels TRUS bei allen 393 Patienten unter Verwendung der Ellipsoid-Formel ($\text{Höhe} \cdot \text{Breite} \cdot \text{Länge} \cdot \pi/6$) vorgenommen. Der TRUS wurde ausschließlich von erfahrenen Untersuchern mit dem Combison 330 (Kretz Technik, Zipf, Österreich) durchgeführt.

3.3.2 Digital-rektale Untersuchung

An allen 393 Patienten der „Interne Validierung“ Kohorte erfolgte eine DRU. Ein Tastbefund wurde als nicht karzinomverdächtig gewertet, wenn die Drüse vergrößert, gut abgrenzbar, ohne Verhärtungen und mit glatter Oberfläche getastet wurde. Derartige Befunde wurden als DRU negativ (DRU=0) definiert. Karzinomverdächtige Befunde hingegen wurden als DRU positiv (DRU=1) definiert. Als karzinomverdächtig galten Vorsteherdrüsen, die eines oder mehrere der folgenden Kriterien aufwiesen:

Gewebeverhärtungen, mangelnde Abgrenzbarkeit zur Umgebung oder palpable Knoten.

3.4 Artificielle neuronale Netzwerke

Das online verfügbare ANN „ProstataClass“ (ANN_{PC}) basiert auf retrospektiven Daten der „ProstataClass“ Kohorte. Die Eingangsvariablen sind tPSA, %fPSA, Prostatavolumen, DRU-Status und Alter (133). Es handelt sich um ein Feed-Forward-Back-Propagation-Netzwerk, dessen Eingangsebene aus fünf Neuronen für die Variablen tPSA, %fPSA, Alter, DRU-Status und Prostatavolumen besteht. Die Konstruktion erfolgte mit der MATLAB Neuronal Network Toolbox (MathWorks Inc., Natick, MA, USA). Ein Neuron bildet die Ausgangsebene und kann Werte von 0 (geringstes PCa-Risiko) bis 1 (höchstes PCa-Risiko) ausgeben. Die Eingangsebene und die Ausgangsebene werden über drei Neurone einer verborgenen Ebene verbunden. Eine Verbesserung der Generalisierbarkeit des Netzwerkes wurde mittels Bayesscher-Regulierung erreicht, die bereits von Finne et al. erfolgreich angewandt wurde (112). Diese Art der Regulierung verhindert die Überanpassung (engl. overfitting) des Modells an die Trainingsdaten (134). Überanpassung beschreibt eine Steigerung der Leistung des Netzwerkes in Bezug auf die Trainingspopulation bei gleichzeitiger Leistungsabnahme bezüglich einer vollkommen randomisierten Population. Ursächlich für die mögliche Überanpassung eines Netzwerkes ist die Berücksichtigung von zufälligen Eigenschaften der Trainingspopulation, die keinen kausalen Zusammenhang mit der Fragestellung (z.B. Erkrankung) aufweisen. Zur Validierung des Netzwerkes wurde die Leave-One-Out-Kreuzvalidierung (LOO) angewandt. Bei dieser Methode wird das Netzwerk N-fach trainiert und N-fach validiert, wobei N die Anzahl der Probanden ist. Dazu wird das Netzwerk jeweils mit den Daten von N - 1 Patienten trainiert und anschließend anhand des ausgeschlossenen Patienten geprüft. Dies lässt sich N-mal wiederholen, sodass jeder Patient einmal zur Validierung genutzt wird (112;133;135).

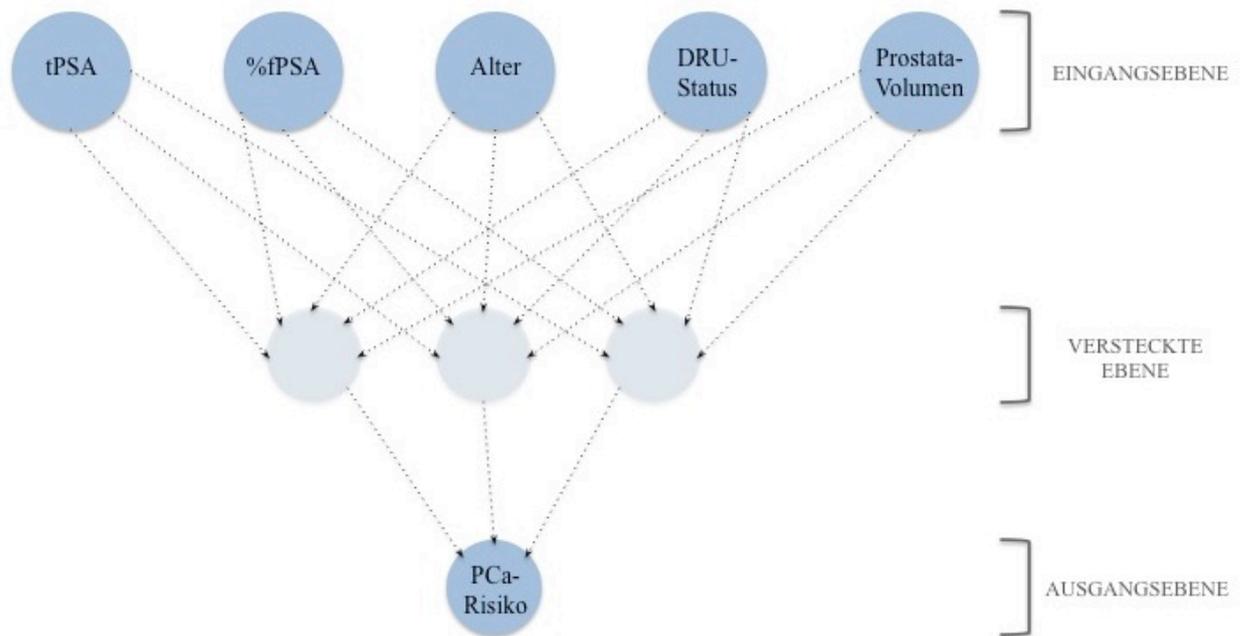


Abb. 2 Schematische Darstellung eines künstlichen neuronalen Netzwerkes zur Vorhersage des PCa-Risikos. (abgewandelt nach einer Abbildung von Anagnostou et al. (107))

Ergänzend zum ANN „ProstataClass“ wurde auf Basis der prospektiven Daten der „Interne Validierung“ Kohorte das ANN „Interne Validierung“ (ANNiv) erstellt. Hierbei kam ebenfalls das Programm MATLAB zur Anwendung. Das ANNiv ist, wie das ANNpc ein Feed-Forward-Back-Propagation-Netzwerk mit fünf Eingangsneuronen für die Variablen tPSA, %fPSA, Alter, DRU-Status und Prostata-Volumen. Die versteckte Ebene besteht gleichermaßen aus drei Neuronen, die mit einem Ausgangsneuron verbunden sind. Das Ausgangsneuron kann auch bei diesem Netzwerk Werte von 0 (geringstes Karzinomrisiko) bis 1 (höchstes Karzinomrisiko) annehmen. Zum Training und zur Validierung des ANNiv kam ebenso die LOO zur Anwendung.

Letztlich wurden die Daten beider Kohorten zusammengenommen und zur Grundlage eines weiteren Netzwerkes gemacht: Das ANN „Gesamt“ (ANNall) wurde in gleicher Weise wie das ANNpc und das ANNiv konstruiert, trainiert und validiert.

Neben den drei internen ANN werden zwei externe Netzwerke genutzt. Das Multi Layer Perceptron (ANNmpl) wurde von Finne et al. entwickelt und basiert auf Daten von 656 Patienten der ERSPC (112). Es setzt sich zusammen aus einer auf vier Neuronen

aufgebauten Eingangsebene, einer aus zwei Neuronen bestehenden verborgenen Zwischenebene und einer Ausgangsebene, die von einem Neuron gebildet wird. Die Eingangsvariablen des ANNmlp sind tPSA, %fPSA, Alter und Prostatavolumen. Das Ausgangsneuron kann Werte von 0 (geringstes Risiko) bis 1 (höchstes Risiko) annehmen. Das Netzwerk unterscheidet sich von den internen Netzwerken dadurch, dass es auf einer Screeningpopulation, einem anderen PSA Assay (Prostatus PSA, EG&G-Wallac, Turku, Finnland) und nur auf vier, anstatt fünf Eingangsparametern basiert (112).

Das ANN Hamburg (ANNham) wurde mit den Daten einer Studie von Chun et al. konstruiert und unterscheidet sich ebenfalls von den internen künstlichen neuronalen Netzwerken. Es basiert auf Daten, die mittels Abbott Axym PSA Assay (Abbott Park, IL, USA) erhoben wurden. Die Eingangsvariablen sind die gleichen wie bei den internen Netzwerken (128).

Tabelle 1: Vergleich der Ursprungskohorten der fünf artifiziellen neuronalen Netzwerke.

	ANNpc	ANNiv	ANNall	ANNmlp	ANNham
Anzahl N	787	393	1180	656	3980
Art der Kohorte	retrospektiv / eingewiesen	prospektiv / eingewiesen	gemischt / eingewiesen	prospektiv / Screening	retrospektiv / eingewiesen
tPSA / fPSA - Assay	Elecsys	Elecsys	Elecsys	Prostatus	AxSYM
tPSA Bereich (µg/L)	1 – 24,4	1,04 – 22,8	1 – 24,4	4 - 10	2 - 20

3.5 Statistische Methoden

Zur diagnostischen Validitätsprüfung wurden die Flächen unter der Kurve (engl.: area under the curve, AUC) der ROC- (receiver operating characteristic) Kurven, sowie die Spezifitäten bei den Sensitivitätsgrenzen von 90 % bzw. 95 % verglichen (136). Beide

Verfahren dienen zur Beurteilung der Diskriminationsfähigkeit der Modelle und messen somit, wie gut „Gesunde“ von „Kranken“ getrennt werden. Das Programm GraphROC 2.1 für Microsoft Windows kam hierbei zur Anwendung (137).

Zusätzlich zu der Analyse der ROC-Kurven und dem Vergleich der Spezifitäten erfolgte eine qualitative Bewertung der diagnostischen Modelle mittels Kalibrierung (Vergleich der vorhergesagten PCa-Wahrscheinlichkeiten mit den histologisch bestätigten PCa-Raten). Hierzu wurde zunächst die „ProstataClass“ Kohorte bzw. die „Interne Validierung“ Kohorte aufsteigend nach der vorhergesagten PCa-Wahrscheinlichkeit sortiert und eine Unterteilung in jeweils 20 Subgruppen (5 % der Kohorte in jeder Gruppe) vorgenommen. Daraus ergibt sich für die „ProstataClass“ Kohorte eine Gruppengröße von 39 Patienten und für die „Interne Validierung“ Kohorte eine Gruppengröße von 20 Patienten. Die grafische Darstellung der mittleren vorhergesagte PCa-Wahrscheinlichkeit in jeder Gruppe und der jeweils dazugehörenden beobachteten PCa-Rate ermöglicht eine genaue visuelle Bewertung der Übereinstimmung zwischen beiden Werten. Der aus den Wertepaaren berechnete Intraklassenkorrelationskoeffizient (engl.: intra class correlation coefficient, ICC) beschreibt die Übereinstimmung zwischen den gepaarten Gruppen und spiegelt deshalb die Richtigkeit der vorausgesagten Wahrscheinlichkeiten für alle Gruppen wider. Ein Wert von 1 beschreibt vollständige Übereinstimmung in allen Gruppen (138).

Zudem wurde die residuelle Varianz (RV) für die einzelnen Kalibrierungen errechnet. Diese beschreibt das Abweichen der Punkte von der 45° Linie und bietet ein Gesamtmaß für die Übereinstimmung der vorhergesagten PCa-Wahrscheinlichkeiten und der beobachteten PCa-Raten. Je geringer die RV ist, desto größer ist die Gesamtkonkordanz.

Zur Verdeutlichung der Beziehung zwischen der 45° Linie und den Datenpunkten sind in der graphischen Darstellung Ausgleichskurven dargestellt.

Die statistische Auswertung erfolgte mittels SPSS 17.0 für Microsoft Windows (SPSS, Chicago, USA) unter Anwendung des Mann-Whitney U Testes und des Chi-Quadrat Testes. Als statistisch signifikant wurde eine Überschreitungswahrscheinlichkeit von $P < 0,05$ angesehen. Die Abbildungen wurden mit dem Programm MATLAB (MathWorks Inc., Natick, MA, USA) erstellt.

4 Ergebnisse

4.1 Vergleich der Kohorten

Beim Vergleich der beiden Kohorten zeigt sich fast kein Unterschied bezüglich der prozentualen Verteilung von PCa-Patienten und NEM-Patienten. Mit 467 Karzinompatienten und 320 NEM-Patienten liegt der Anteil der Krebspatienten in der „ProstataClass“ Kohorte bei 59,3 %. Im Vergleich dazu entsprechen die 229 PCa-Patienten der „Interne Validierung“ Kohorte einem Anteil von 58,3 %.

Mit Ausnahme von tPSA können sowohl innerhalb der „Interne Validierung“ Kohorte als auch innerhalb der „ProstataClass“ Kohorte für alle Variablen signifikante Unterschiede zwischen den PCa-Patienten und den Patienten ohne Anhalt auf Malignität nachgewiesen werden ($P < 0,0001$). Für das tPSA in der „Interne Validierung“ Kohorte beträgt der tPSA Median Wert der PCa-Patienten 5,8 $\mu\text{g/L}$ und der der NEM-Patienten 6,12 $\mu\text{g/L}$ ($P = 0,36$).

Zwischen den beiden Kohorten kann für die Parameter Alter ($P = 0,35$) und Prostatavolumen ($P = 0,77$) kein signifikanter Unterschied festgestellt werden (siehe Tabelle 2). Der Vergleich einerseits nur der PCa-Patienten und andererseits nur der NEM-Patienten zwischen der „ProstataClass“ und „Interne Validierung“ Kohorte zeigte für die Variablen Alter und Prostatavolumen ebenfalls keine signifikanten Unterschiede zwischen den beiden Kohorten. Die P-Werte für die Karzinompatienten sind 0,42 beim Altersvergleich und 0,77 beim Prostatavolumenvergleich, sowie für die NEM-Patienten 0,9 beim Vergleich des Alters und 0,79 beim Vergleich des Drüsenvolumens.

Allerdings zeigten sich für die Parameter tPSA, %fPSA und DRU-Status signifikante Unterschiede sowohl zwischen den gesamten Kohorten (P von 0,017 bis $< 0,0001$) als auch beim einzelnen Vergleich von PCa-Patienten und NEM-Patienten zwischen den Kohorten. Die P-Werte liegen für die Karzinompatienten in allen Fällen unter 0,0001 und für die NEM-Patienten ergeben sich P-Werte von 0,04 bis $< 0,0001$.

Die %fPSA Mittel- und Medianwerte der „Interne Validierung“ Kohorte sind mit 17,4 % bzw. 15,8 % um 3,7 % bzw. 3,6 % höher als die Mittel- und Medianwerte der „ProstataClass“ Kohorte (13,7 % bzw. 12,2 %). Bezüglich der DRU-Befunde finden sich ebenfalls signifikante Unterschiede. In der „Interne Validierung“ Kohorte ist die Anzahl

der positiven DRU-Befunde etwa ein Drittel geringer verglichen mit der „ProstataClass“ Kohorte. Dies gilt sowohl für den Vergleich der gesamten Kohorten (13 % gegenüber 39,8 %) als auch für den separaten Vergleich von PCa-Patienten (19,7 % gegenüber 60,6 %) bzw. NEM-Patienten (3,7 % gegenüber 9,4 %).

Tabelle 2: Mittelwerte, Mediane und Intervalle für Alter, tPSA, %fPSA, Prostatavolumen, sowie Anzahl der positiven (verdächtigen) DRU-Befunde gruppiert nach Gesamt, PCa und NEM für beide Kohorten.

Variable	„Interne Validierung“ Kohorte			„ProstataClass“ Kohorte			P Wert [#]
	Gesamt	PCa	NEM	Gesamt	PCa	NEM	
Anzahl N	393	229	164	787	467	320	
Alter (Jahre)							
Mittelwert	64	63	66	64	63	66	0,353
Median	65	64	66.5	64	63	66	
Intervall	45-88	45-83	45-88	38-85	43-79	38-85	
tPSA (µg/l)							
Mittelwert	6,9	7,07	6,68	7,59	8,74	5,92	0,017
Median	5,99*	5,84*	6,12	6,67	7,59	5,05	
Intervall	1,04-22,8	1,07-22,5	1,04-22,8	1-24,4	1,18-24,4	1-24,3	
%fPSA (%)							
Mittelwert	17,4	14,6	21,2	13,7	11	17,7	< 0,0001
Median	15,8*	13,3*	19,2*	12,2	10,1	16,7	
Intervall	1,1-56,6	1,1-54,6	3-56,6	2,07-45,8	2,07-36,9	2,13-45,8	
Volumen (ml)							
Mittelwert	44,4	37,9	53,5	43,8	37,2	53,4	0,771
Median	37,7	32	50	38	34	46,5	
Intervall	7-159	7-145	15-159	10-180	10-110	13-180	
positive DRU (%)	51* (13%)*	45* (19,7%)*	6* (3,7%)*	313 (39,8%)	283 (60,6%)	30 (9,38%)	< 0,0001

* signifikanter Unterschied zur „ProstataClass“ Kohorte

P-Wert für den Vergleich der gesamten „Interne Validierung“ Kohorte mit der gesamten „ProstataClass“ Kohorte

4.2 Ergebnisse der ROC-Analysen

Zum Vergleich der diagnostischen Genauigkeit aller Parameter und der ANN untereinander in Bezug auf die verschiedenen Kohorten sind die AUC-Werte der ROC-Kurven für tPSA, %fPSA und die ANN-Modelle in Tabelle 3 aufgeführt. Neben dem ANNpc, dem ANNiv und dem ANNall sind die AUC-Werte für zwei externe ANN-Modelle, eines auf Screeningdaten aus Finnland (ANNmlp) (112), das andere auf einer großen Population eingewiesener Patienten (ANNham) (128) basierend, in der Tabelle 3 angegeben. Ergänzend sind in Abbildung (Abb.) 3 - 5 die ROC-Kurven der fünf neuronalen Netzwerke für die „Interne Validierung“ Kohorte jeweils den ROC-Kurven von tPSA und %fPSA gegenübergestellt.

Tabelle 3: ROC-Kurvenanalyse für tPSA, %fPSA und die fünf artifiziellen neuronalen Netzwerke.

Kohorte	Parameter		Artifizielle neuronale Netzwerke				
	tPSA	%fPSA	ANNpc	ANNiv	ANNall	ANNmlp	ANNham
AUC-Werte und Standardfehler in (Klammern)							
„Interne Validierung“ Kohorte	0,527* [#] (0,031)	0,726 (0,026)	0,747 (0,025)	0,754* (0,025)	0,774* (0,024)	0,726 (0,026)	0,729 (0,026)
„ProstataClass“ Kohorte	0,7* [#] (0,019)	0,782 [#] (0,016)	0,883* (0,012)	0,873* (0,013)	0,866* [#] (0,013)	0,856* [#] (0,013)	0,882* (0,012)
Spezifität bei 95% Sensitivität							
„Interne Validierung“ Kohorte	15,9%	22,6%	22,6%	20,7%	25,0%	22,6%	23,2%
„ProstataClass“ Kohorte	27,8%	27,5%	48,1%*	46,9%*	40,6%*	42,5%*	53,8%*
Spezifität bei 90% Sensitivität							
„Interne Validierung“ Kohorte	23,8%	31,7%	31,1%	37,2%	40,9%	30,5%	33,5%
„ProstataClass“ Kohorte	39,4%	44,1%	64,1%*	57,2%*	55,3%*	60,0%*	67,8%*

* signifikanter Unterschied verglichen mit %fPSA

signifikanter Unterschied verglichen mit ANNpc (nur für AUC-Werte errechnet)

Die Betrachtung der AUC-Werte für die „Interne Validierung“ Kohorte zeigt einen signifikant höheren Wert für %fPSA (0,726) im Vergleich zu tPSA (0,527) ($P < 0,0001$). Dieses wird in den Abb. 3 - 5 verdeutlicht.

In der „Interne Validierung“ Kohorte können ausschließlich das ANNiv ($P = 0,021$) und das ANNall ($P = 0,0008$) eine signifikante Verbesserung der diagnostischen Genauigkeit verglichen mit %fPSA erreichen. Weder das ANNpc ($P = 0,085$), noch die beiden auf externen Daten basierenden ANN-Modelle ANNmpl ($P = 0,99$) und ANNham ($P = 0,88$) zeigen signifikant höhere AUC-Werte im Vergleich zu %fPSA. Die entsprechenden ROC-Kurven sind in Abb. 3 für das ANNpc und in Abb. 5 für das ANNmpl und das ANNham dargestellt. Der Vergleich aller ANN-Modelle miteinander zeigt in keinem Fall signifikante Unterschiede bezüglich der AUC-Werte (P -Werte zwischen 0,09 und 0,79). Das ANNall erreicht für die „Interne Validierung“ Kohorte mit 0,774 den größten AUC-Wert gefolgt vom ANNiv, welches einen AUC-Wert von 0,754 erreicht. Mit einem AUC-Wert von 0,747 zeigt das ANNpc für die prospektive Kohorte keine signifikante Verbesserung verglichen mit %fPSA. Die beiden externen ANN erreichen mit AUC-Werten von 0,726 (ANNmpl) und 0,729 (ANNham) ebenfalls keine verbesserte Diskriminationsfähigkeit im Vergleich zu %fPSA.

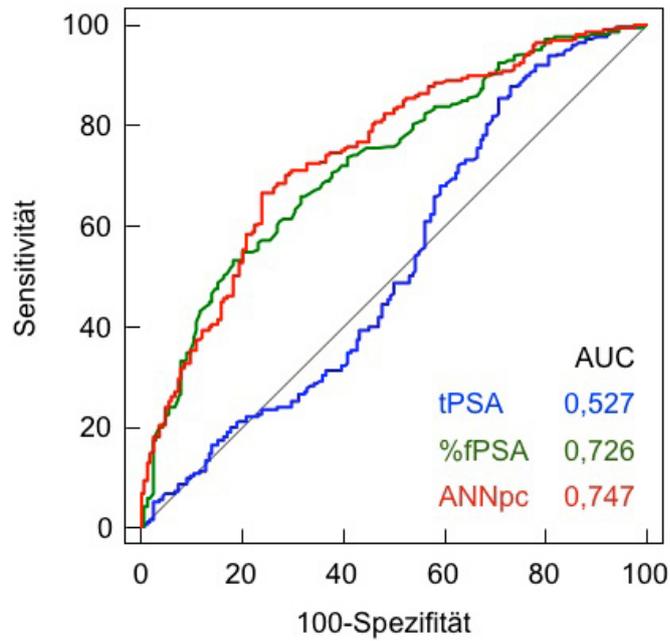


Abb. 3: ROC-Kurven von tPSA, %fPSA und ANN „ProstataClass“ (ANNpc) für die „Interne Validierung“ Kohorte mit den entsprechenden AUC-Werten.

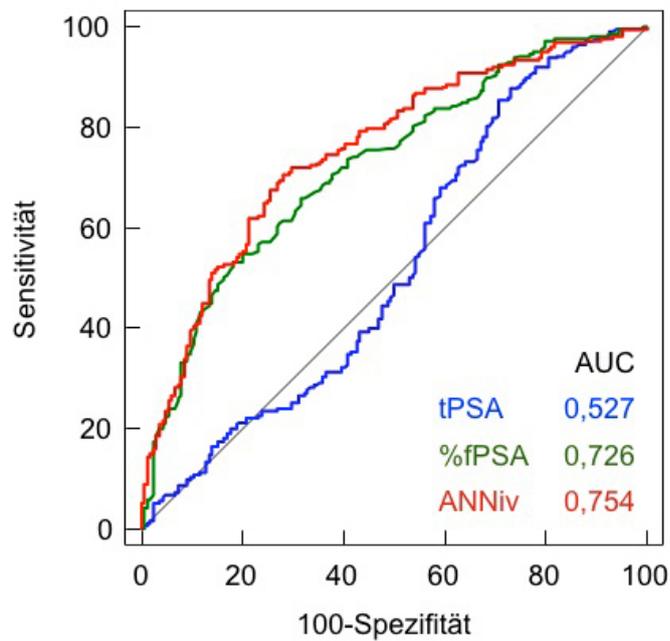


Abb. 4: ROC-Kurven von tPSA, %fPSA und ANN „Interne Validierung“ (ANNiv) für die „Interne Validierung“ Kohorte mit den entsprechenden AUC-Werten.

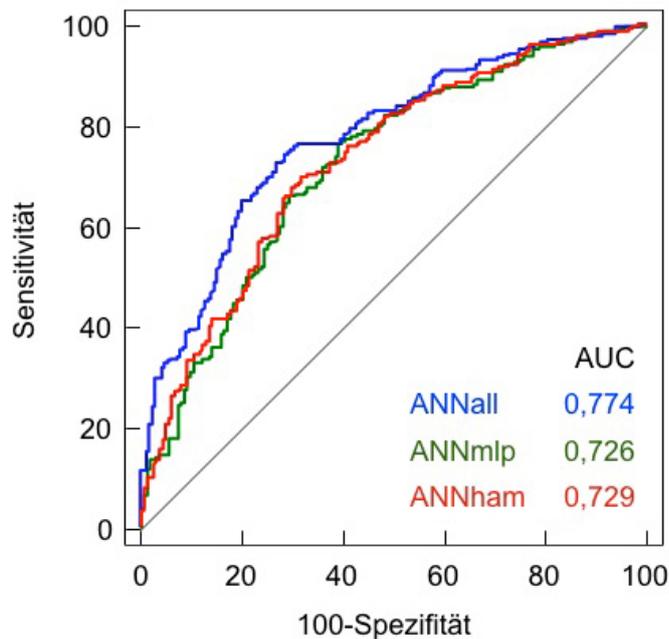


Abb. 5: ROC-Kurven von ANN „Gesamt (ANNall), Multi Layer Perceptron (ANNmlp) und ANN „Hamburg“ (ANNham) für die „Interne Validierung“ Kohorte mit den entsprechenden AUC-Werten.

Für die „ProstataClass“ Kohorte gilt ebenfalls, dass der AUC-Wert von %fPSA (0,782) signifikant höher ist ($P < 0,0001$) als der AUC-Wert des tPSA (0,7). Allerdings erreichen bei dieser Kohorte alle fünf ANN-Modelle AUC-Werte zwischen 0,856 und 0,883. Somit zeigen alle fünf ANN eine signifikante Steigerung der diagnostischen Genauigkeit verglichen mit %fPSA (P -Werte $< 0,0001$), welches nur einen AUC-Wert von 0,782 erreicht. Bei der detaillierten Analyse der ROC-Kurven für die „ProstataClass“ Kohorte wird deutlich, dass das ANNpc mit einem AUC-Wert von 0,883 und das ANNham mit einem AUC-Wert von 0,882 am besten abschneiden, wobei kein signifikanter Unterschied zwischen den beiden ANN-Modellen besteht ($P = 0,92$). Das ANNiv, welches auf der „Interne Validierung“ Kohorte basiert, erreicht einen AUC-Wert von 0,873, womit es ebenfalls keinen signifikanten Unterschied zum ANNpc zeigt ($P = 0,118$). Andererseits erreichen das ANNall und das ANNmlp nur AUC-Werte von 0,866 bzw. 0,856 und schneiden somit signifikant schlechter ab als das ANNpc ($P = 0,02$ bzw. $P < 0,0001$).

4.3 Spezifitätenvergleich bei 90 % und 95 % Sensitivität

Der Vergleich der Spezifitäten der ANN mit %fPSA zeigt für die „Interne Validierung“ Kohorte keine signifikanten Vorteile auf Seiten der Netzwerke, weder bei 90 % ($P = 0,11 - 0,92$) noch bei 95 % Sensitivität ($P = 0,13 - 0,9$). Des Weiteren bestätigen die Daten des Spezifitätenvergleichs das sehr ähnliche Abschneiden aller Netzwerke. Bei der klinisch wichtigen Sensitivitätsgrenze von 95 % zeigen die Spezifitäten der Netzwerke (20,7 % bis 25 %) nur marginale Unterschiede. Für die „ProstataClass“ Kohorte hingegen finden sich für die ANN im Vergleich zum %fPSA signifikant höhere Spezifitäten (P von 0,0042 bis $<0,0001$) für den klinisch relevanten Grenzwert bei 95 % Sensitivität. Die Spezifität der Netzwerke ist im Vergleich zum %fPSA um 13 % bis 26 % höher. Zusätzlich bestätigen die Spezifitäten das sehr ähnliche Abschneiden aller Netzwerke in der „ProstataClass“ Kohorte (siehe Tabelle 3).

Das insgesamt sehr ähnliche Abschneiden der verschiedenen ANN für jeweils dieselbe Kohorte ist eines der Hauptergebnisse der vorliegenden Studie. Diesbezüglich ist anzumerken, dass sowohl das ANN_{pc} als auch das ANN_{iv} und das ANN_{all} auf Daten basieren, die mittels Elecsys Assay der Firma Roche gemessen wurden. Das ANN_{mlp} und das ANN_{ham} hingegen basieren auf Daten, die mittels Prostatus Assay (EG & G-Wallac, Turku, Finnland) bzw. AxSYM Assay (Abbott, Abbott Park, IL, USA) erhoben wurden.

4.4 Gültigkeit der PCa-Vorhersage

Um die Qualität der Netzwerke in Bezug auf die Prostatakrebsvorhersage zu bewerten, werden neben der Analyse der ROC-Kurven die vorausgesagten PCa-Wahrscheinlichkeiten mit den durch Biopsie bestätigten tatsächlichen PCa-Raten verglichen (Kalibrierung).

Abb. 6a zeigt relativ große Übereinstimmungen zwischen den Wahrscheinlichkeiten und den PCa-Raten für das ANN_{pc} bezüglich der „ProstataClass“ Kohorte. Die geringe RV von 0,007 und ein ICC von 0,971 bestätigen den visuellen Eindruck, dass das ANN_{pc} gute Vorhersagen für diese Kohorte macht. In Abb. 6b weicht die Ausgleichskurve deutlich von der 45° Linie ab. Auch eine schwächere RV von 0,067 und ein ICC von nur

0,609 bestätigen das schlechtere Abschneiden des „ProstataClass“ Netzwerkes bezüglich der „Interne Validierung“ Kohorte. Beide Abbildungen zeigen ein Abweichen der Kurve nach oben links, dies entspricht einem Unterschätzen der PCa-Wahrscheinlichkeit durch das Netzwerk.

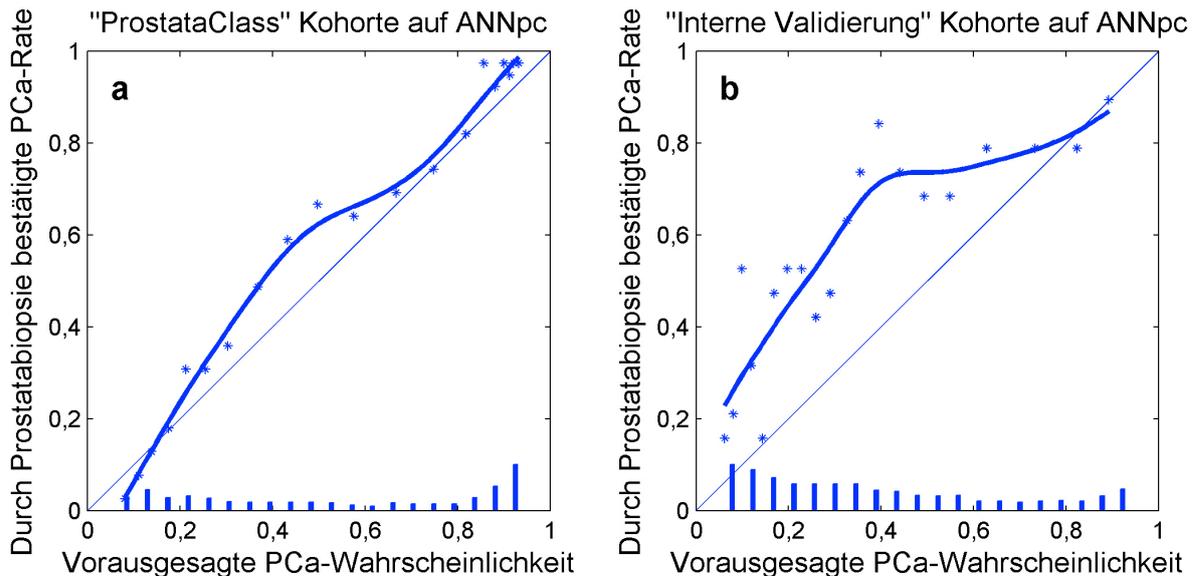


Abb. 6: Die Kalibrierungen mit Ausgleichskurven des ANN „ProstataClass“ (ANNpc) für die „ProstataClass“ Kohorte (a) und für die „Interne Validierung“ Kohorte (b) zeigen den Einfluss der Kohorte auf das Abschneiden des Netzwerkes. (a) AUC (Fläche unter der ROC-Kurve) = 0,883; RV (residuelle Varianz) = 0,007; ICC (Intraklassenkorrelationskoeffizient) = 0,971. (b) AUC = 0,747; RV = 0,067; ICC = 0,609.

Die Ergebnisse für das ANNiv sind in Abb. 7a und Abb. 7b dargestellt, die ein gegenteiliges Bild zeigen. Wie erwartet zeigt Abb. 7a ein schlechtes Abschneiden des „Interne Validierung“ Netzwerkes für die „ProstataClass“ Kohorte. Die Kalibrierung ergibt eine RV von 0,025 und einen ICC von 0,851. Die Kurve weicht besonders im Bereich der niedrigen Karzinomwahrscheinlichkeiten deutlich von der 45° Linie ab und zeigt, dass das Risiko in diesem Bereich überschätzt wird. Für die „Interne Validierung“ Kohorte ergibt sich ein anderes Bild: Abb. 7b zeigt eine Ausgleichskurve, die nur sehr

gering von der 45° Linie abweicht. Eine sehr niedrige RV von 0,009 und ein ICC von 0,901 belegen die hohe Validität der Vorhersage durch das ANNiv in Bezug auf die „Interne Validierung“ Kohorte.

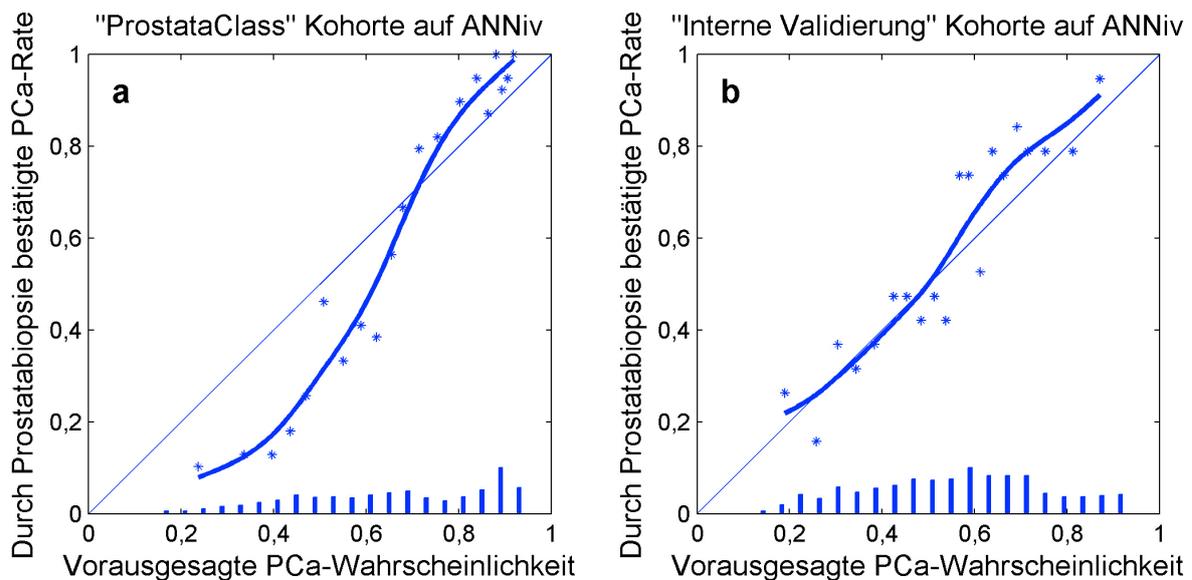


Abb. 7: Die Kalibrierungen mit Ausgleichskurven des ANN „Interne Validierung“ (ANNiv) für die „ProstataClass“ Kohorte (a) und für die „Interne Validierung“ Kohorte (b) zeigen den Einfluss der Kohorte auf das Abschneiden des Netzwerkes. (a) AUC (Fläche unter der ROC-Kurve) = 0,873; RV (residuelle Varianz) = 0,025; ICC (Intraklassenkorrelationskoeffizient) = 0,851. (b) AUC = 0,754; RV = 0,009; ICC = 0,901.

Abb. 8a und Abb. 8b zeigen die Kalibrierung für das ANNall. Für die „ProstataClass“ Kohorte (Abb. 8a) ergibt der Vergleich von vorausgesagten Karzinomwahrscheinlichkeiten mit bestätigten PCa-Raten eine RV von 0,046 und einen relativ niedrigen ICC von 0,684. Dies wird auch durch den visuellen Eindruck bestätigt. Die Kurve weicht stark von der 45° Linie ab. Das Karzinomrisiko wird dabei größtenteils überschätzt. Für die „Interne Validierung“ Kohorte sind die Ergebnisse in Abb. 8b dargestellt. Die Vorhersagen stimmen hier zwar etwas genauer mit den beobachteten Karzinomraten überein (RV = 0,024; ICC = 0,766), dennoch wird das PCa-Risiko für die meisten

Untergruppen überschätzt.

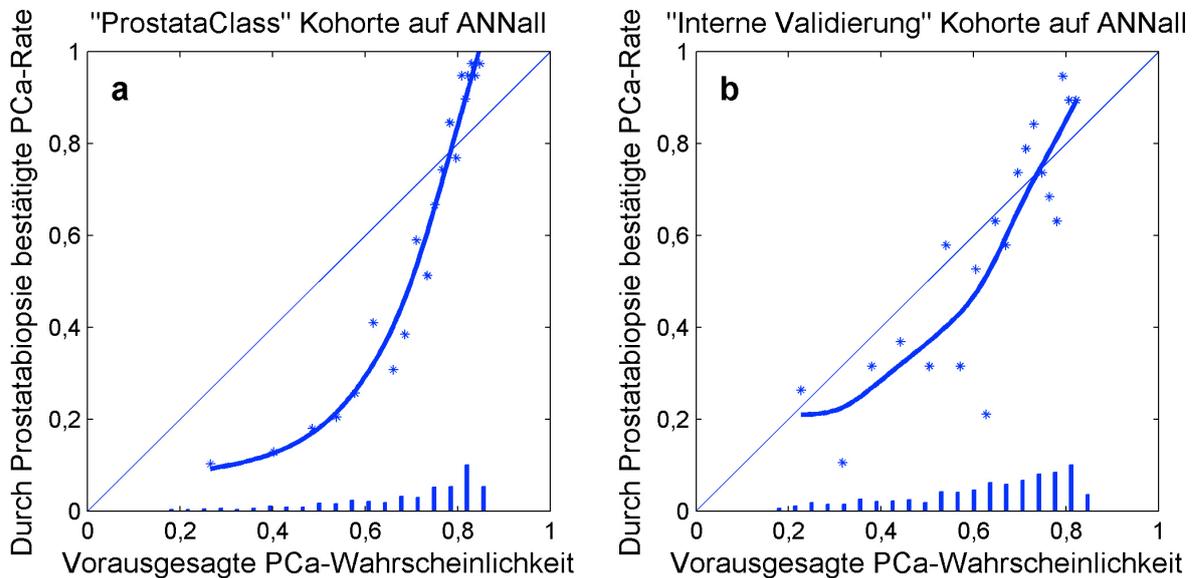


Abb. 8: Die Kalibrierungen mit Ausgleichskurven des ANN „Gesamt“ (ANNall) für die „ProstataClass“ Kohorte (a) und für die „Interne Validierung“ Kohorte (b) zeigen den Einfluss der Kohorte auf das Abschneiden des Netzwerkes. (a) AUC (Fläche unter der ROC-Kurve) = 0,866; RV (residuelle Varianz) = 0,046; ICC (Intraklassenkorrelationskoeffizient) = 0,684. (b) AUC = 0,774; RV = 0,024; ICC = 0,766.

Die Kalibrierungen für das externe ANNmlp zeigen sowohl für die „ProstataClass“ Kohorte (Abb. 9a) als auch für die „Interne Validierung“ Kohorte (Abb. 9b) große RV-Werte (RV = 0,090 bzw. RV = 0,110) und kleine ICC-Werte (ICC = 0,542 bzw. ICC = 0,327). Demnach zeigt das ANNmlp bei beiden Kohorten eine nur unzureichende Genauigkeit bei der Vorhersage eines PCa. Zusätzlich zeigt das Abweichen der Kurve, dass die Karzinomwahrscheinlichkeiten für fast alle Gruppen unterhalb der bestätigten PCa-Raten liegen. Das Risiko eines positiven Biopsieergebnisses wird demnach von dem Modell weitgehend zu niedrig eingeschätzt.

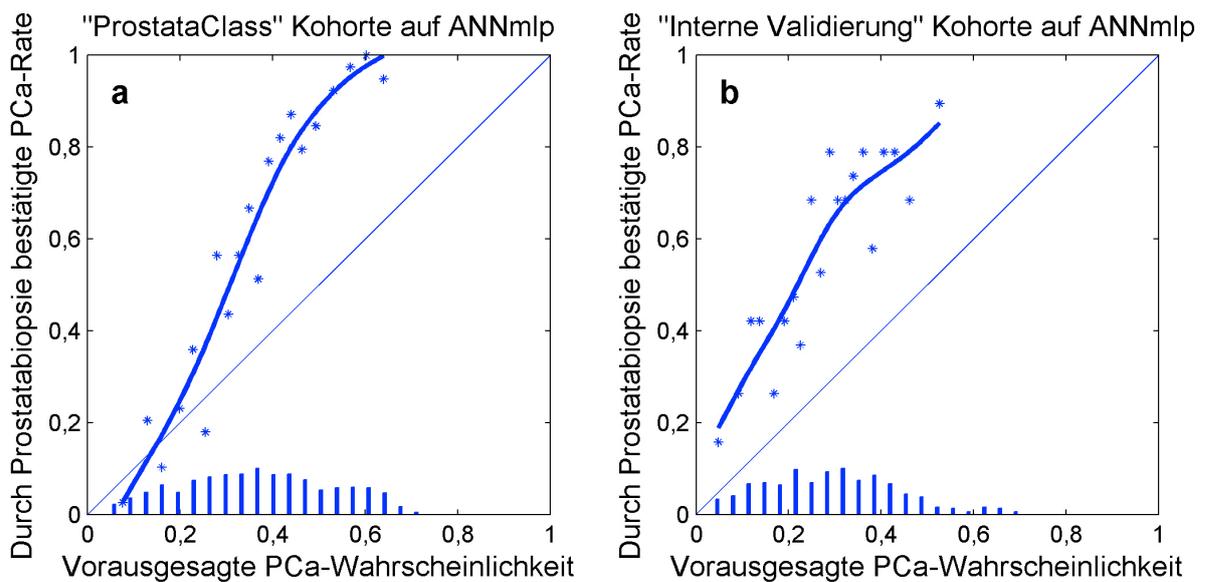


Abb. 9: Die Kalibrierungen mit Ausgleichskurven des Multi Layer Perceptrons (ANNmlp) für die „ProstataClass“ Kohorte (a) und für die „Interne Validierung“ Kohorte (b) zeigen den Einfluss der Kohorte auf das Abschneiden des Netzwerkes. (a) AUC (Fläche unter der ROC-Kurve) = 0,856; RV (residuelle Varianz) = 0,090; ICC (Intraklassenkorrelationskoeffizient) = 0,542. (b) AUC = 0,726; RV = 0,110; ICC = 0,327.

Für das zweite externe Netzwerk zeigt Abb. 10a vornehmlich ein Abweichen der Kurve nach unten rechts. Folglich prognostiziert das ANNham für die Patienten der „ProstataClass“ Kohorte meist ein zu hohes PCa-Risiko. Lediglich im Hochrisikobereich kehrt sich dieser Sachverhalt um. Die RV liegt bei 0,035 und der ICC bei 0,773. Bei der Bewertung der „Internen Validierung“ Kohorte (Abb. 10b) hingegen liegen die vom ANNham vorausgesagten Wahrscheinlichkeiten nahe bei den histologisch gesicherten PCa-Raten, was durch eine niedrige RV von 0,008 und einen ICC von 0,888 belegt wird.

Obgleich alle ANN-Modelle für eine jeweilige Kohorte sehr ähnliche AUC-Werte aufweisen, zeigt der Vergleich der vorausgesagten PCa-Wahrscheinlichkeiten und bestätigten Karzinomraten aber erhebliche Unterschiede im Kurvenverlauf und somit auch bei den RV- und den ICC-Werten.

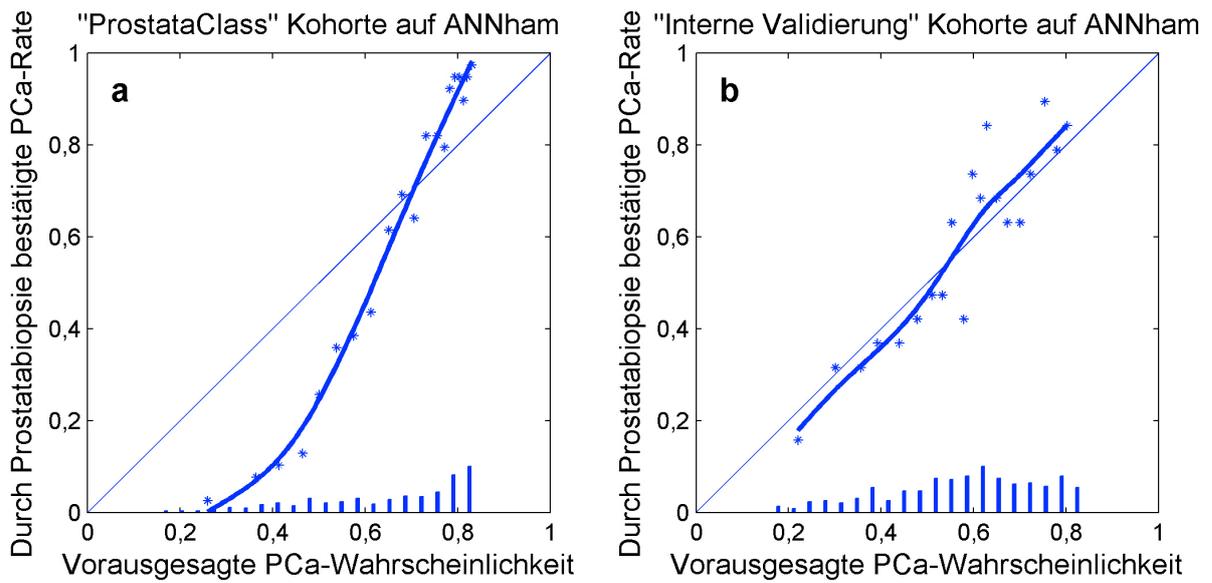


Abb. 10: Die Kalibrierungen mit Ausgleichskurven des ANN Hamburg (ANNham) für die „ProstataClass“ Kohorte (a) und für die „Interne Validierung“ Kohorte (b) zeigen den Einfluss der Kohorte auf das Abschneiden des Netzwerkes. (a) AUC (Fläche unter der ROC-Kurve) = 0,882; RV (residuelle Varianz) = 0,035; ICC (Intraklassenkorrelationskoeffizient) = 0,773. (b) AUC = 0,730; RV = 0,008; ICC = 0,888.

5 Diskussion

Bereits seit der klinischen Etablierung des PSA Anfang der 90er Jahre fand eine kontroverse Diskussion um das PSA als Screeningparameter statt (6). Die breite Anwendung des PSA in der Krebsvorsorge hat den Anteil der lokal begrenzten Tumoren bei Diagnosestellung deutlich erhöht. Allerdings ist die Anzahl der diagnostizierten Neuerkrankungen ebenfalls stark gestiegen, weil seither mehr inapparente Karzinome entdeckt werden (7).

Ob sich die Mortalität durch Screeningprogramme mittels PSA tatsächlich senken lässt und wie viele PCa-Patienten als Folge der Vorsorge überbehandelt werden, steht derzeit im Mittelpunkt der Diskussion. Die ERSPC-Ergebnisse zeigen eine Reduktion der Mortalität durch Teilnahme an der Krebsvorsorge um 20 % bzw. 31 % (8;9). Allerdings widersprechen die Ergebnisse der europäischen Screeningstudie den Ergebnissen einer Studie aus den USA, die keine signifikante Senkung der Mortalität zeigen konnte (10).

Die Überbehandlung von Karzinompatienten, die durch Früherkennung entdeckt werden, möglicherweise jedoch nicht therapiert werden müssten, ist besonders in höherem Alter ein Problem. Bis zu 57 % der PCa-Patienten zwischen 55 bis 67 Jahre werden überbehandelt (11). Wird die Prostatakrebsvorsorge zwischen dem 68. und 75. Lebensjahr fortgeführt, kommt es pro klinisch relevanter Diagnose zu zwei Überbehandlungen (11).

In den letzten Jahren werden verstärkt multivariate Modelle wie ANN oder Nomogramme eingesetzt, um die Früherkennung des PCa zu verbessern und unnötige Biopsien zu vermeiden: d.h. nicht mehr jeder Patient mit einem erhöhten PSA-Wert, sondern nur derjenige, meist jüngere Patient, mit einem möglicherweise aggressiveren Karzinom würde, ggf. auch schon bei einem geringeren PSA-Wert oder PSA-Anstieg, biopsiert werden. Hingegen könnte man bei älteren Patienten mit einem moderat erhöhten PSA-Wert und einem höheren Drüsenvolumen auf die Biopsie oder Wiederholungsbiopsie verzichten. Somit könnte dann konsequenterweise auch eine Überbehandlung von Karzinompatienten vermieden werden. Es ist jedoch fraglich, ob derartige multivariate Modelle auf Populationen angewandt werden können, die nicht der Population entsprechen, auf der sie basieren (139). Der überwiegende Teil der ANN

und Nomogramme ist bisher nicht extern validiert, wodurch der Nutzen in unabhängigen Populationen nicht ausreichend beurteilt werden kann. Betrachtet man die wenigen Ergebnisse externer Validierungen von Nomogrammen oder ANN, finden sich in 13 von 16 Fällen erniedrigte AUC-Werte (116). Obschon es sich in den meisten Fällen nur um geringfügig schlechteres Abschneiden der Modelle handelt, sinkt die diagnostische Genauigkeit in einigen wenigen Validierungen um bis zu 10 % (128;140;141). Derart große Unterschiede konnten bei der Validierung eines ANN in nur einem Zentrum gezeigt werden, während die Ergebnisse dieses Netzwerkes für eine Multi-Center-Population kaum von den Ergebnissen für die Ausgangspopulation abweichen (141). Dies unterstützt die Annahme, dass die Leistung eines Netzwerkes stark von der zu bewertenden Population abhängt. Einzelne Populationen können möglicherweise große Unterschiede zur Ursprungspopulation des jeweiligen ANN aufweisen, wodurch das ANN für diese Populationen schlechtere Ergebnisse zeigt. Diese populationsabhängigen Unterschiede werden in multizentrischen Studien ausgeglichen, weshalb die ursprünglichen Ergebnisse des ANN bestätigt werden. Des Weiteren zeigt sich die Populationsabhängigkeit der Netzwerke darin, dass ANN-Modelle, die auf lokalen Populationen und dem lokal verwendete PSA-Assay basieren, für die entsprechende lokale Population bessere Ergebnisse hervorbringen als Netzwerke, die auf multizentrischen Populationen konstruiert wurden (128;140;141). Eine externe Validierung eines Nomogramms, welches auf 10-fach Stanzbiopsien basiert, deutet darauf hin, dass multivariate Modelle, die auf einer größeren Anzahl an Biopsien basieren, bessere Ergebnisse in externen Validierungen zeigen. Auf 6-fach Biopsien basierende Nomogramme haben laut dieser Studie geringere diagnostische Aussagekraft bei Patienten, die ein erweitertes Biopsieschema erhalten haben, als Nomogramme, die auf 10-fach Biopsien basieren (121).

Das ANN_{pc}, welches auf 6- bis 8-fach Biopsien basiert, hat in der vorliegenden Studie eine geringere diagnostische Aussagekraft für die 10-fach biopsierte „Interne Validierung“ Kohorte als für die 6- bis 8-fach biopsierte „ProstataClass“ Kohorte. Dies bestätigt die Annahme, dass die Aussagekraft eines multivariaten Modells für Patienten, die verglichen mit der Ausgangspopulation des Modells eine erhöhte Biopsiezahl erhalten haben, abnimmt.

Eine Studie von Canto et al. zeigt allerdings, dass eine erhöhte Biopsiezahl zu geringeren AUC-Werten für das %fPSA führt (142).

Die vorliegende Arbeit zeigt geringere AUC-Werte für die „Interne Validierung“ Kohorte, deren Daten auf 10-fach Biopsien basieren als für die „ProstataClass“ Kohorte, welche auf 6- bis 8-fach Biopsien basiert. Dies gilt für tPSA, %fPSA und alle ANN, wobei der Unterschied für das tPSA am deutlichsten ist. Der Abstand zwischen den AUC-Werten von tPSA und %fPSA ist bei der „Interne Validierung“ Kohorte größer als bei der „ProstataClass“ Kohorte. Die Ergebnisse dieser Arbeit bestätigen die Annahme, dass eine erhöhte Biopsiezahl Einfluss auf die AUC-Werte der verschiedenen Parameter bzw. ANN hat.

Die Ziele der vorliegenden Arbeit können nur teilweise erreicht werden. Wie Tabelle 3 und Abb. 6b deutlich zeigen, kann die Evaluierung des ANNpc anhand der prospektiven Daten von 393 Patienten der „Interne Validierung“ Kohorte die Erwartungen nicht erfüllen. Das Netzwerk kann die signifikante Steigerung der Diskriminationsfähigkeit verglichen mit %fPSA in der „ProstataClass“ Kohorte jetzt in der prospektiven „Interne Validierung“ Kohorte nicht wiederholen. Es können keine signifikant höheren AUC-Werte ermittelt werden.

Als mögliche Erklärung für das schlechtere Abschneiden des ANNpc bezüglich der prospektiven Kohorte kommen die Unterschiede in der Zusammensetzung der Kohorten in Betracht. Im Wesentlichen unterscheiden sich die „ProstataClass“ Kohorte und die „Interne Validierung“ Kohorte anhand ihres Anteils positiver DRU-Ergebnisse und anhand der mittleren %fPSA-Werte.

Der prozentuale Anteil der Patienten mit karzinomverdächtiger DRU liegt für die „Interne Validierung“ Kohorte bei 13 %. Dies entspricht nur etwa einem Drittel des 39,8 % Anteils der Patienten mit positiver DRU für die „ProstataClass“ Kohorte. Es ist wesentlich, dass der Status der DRU ein starker und PSA-unabhängiger Parameter für die Netzwerke ist. Die Reduktion des Anteils der positiven Tastbefunde führt daher zum Unterschätzen des PCa-Risikos für die prospektive Kohorte, wie Abb. 6b zeigt. Das relativ gute Abschneiden des ANNham für die „Interne Validierung“ Kohorte ist besonders bei der Kalibrierung (Abb. 10b) erkennbar und vermutlich auch auf den geringen Anteil positiver DRU-Befunde in der Ursprungspopulation des ANNham von 5,2 % zurückzuführen ([128](#)). Es stellt sich die Frage, ob ein subjektives Verfahren wie die DRU überhaupt in multivariate Modelle einfließen sollte, da dieser Parameter offenbar großen Einfluss auf das Ergebnis hat. Im Umkehrschluss belegt der große

Einfluss die Wichtigkeit PSA-unabhängiger Parameter in multivariaten Modellen. In Zukunft könnten anstelle des DRU-Status neuere Serummarker und auch Urinmarker größere Bedeutung in der Diagnostik des PCa erlangen (38).

Die %fPSA Mittel- und Medianwerte sind zwischen der retrospektiven und prospektiven Kohorte ebenfalls signifikant unterschiedlich. Diese Tatsache kann ebenfalls als mitursächlich für das schwächere Ergebnis des ANNpc bei der internen Validierung sein. In der prospektiven Kohorte sind der %fPSA-Mittelwert 3,7 % und der %fPSA-Medianwert 3,6 % höher als in der retrospektiven Kohorte. Erhöhte %fPSA-Werte deuten auf ein geringeres PCa-Risiko hin und führen dazu, dass das ANNpc eine zu geringe Karzinomwahrscheinlichkeit prognostiziert. Dies wird in Abb. 6b durch das Abweichen der Ausgleichkurve nach oben links graphisch wiedergegeben. Die Gründe für diesen wesentlichen Unterschied bleiben unklar. Für die Bestimmung der tPSA- und fPSA-Werte für beide Kohorten wurde das Elecsys tPSA und fPSA Assay genutzt, wodurch mögliche Unterschiede aufgrund verschiedener PSA-Assays ausgeschlossen werden können. Hinsichtlich des Umgangs mit den Serumproben unterscheiden sich die Kohorten: Die Daten der „ProstataClass“ Kohorte wurden anhand archivierter Serumproben gewonnen, wogegen es sich bei den Daten der „Interne Validierung“ Kohorte um prospektiv gewonnene Daten handelt. Die Serumproben der „Interne Validierung“ Kohorte wurden am Tag der Blutentnahme analysiert oder maximal 48 Stunden bei -20°C gelagert, bevor die Messungen erfolgten. Zwar konnten Scaramuzzino et al. hinsichtlich der Verwendbarkeit von bei -70°C archivierten Serumproben zur Bestimmung von tPSA und fPSA ausreichende Stabilität der Moleküle für einen Zeitraum von fünf Jahren zeigen (143), dennoch besteht ein signifikanter Unterschied zwischen den Kohorten, der doch am ehesten durch die fPSA-Instabilität zu erklären ist.

Vergleiche des Abschneidens unterschiedlicher ANN, die auf unterschiedlichen Kohorten basieren, bezüglich einer unabhängigen Kohorte sind selten. Diese Arbeit vergleicht fünf ANN miteinander. Das ANNpc basiert auf retrospektiven Daten einer urologisch eingewiesenen Population, das ANNiv auf einer prospektiven urologisch eingewiesenen Population und das ANNall auf diesen beiden Populationen. Die beiden externen Modelle wurden auf Basis einer finnischen Screeningpopulation (ANNmlp) und einer retrospektiven urologisch eingewiesenen Population (ANNham) konstruiert. Zusätzlich wurden die beiden externen Netzwerke anhand von Daten, die mit anderen

PSA-Assays erhoben wurden, trainiert.

Der Vergleich der Netzwerke zeigt für die „Interne Validierung“ Kohorte ein sehr ähnliches Abschneiden aller Modelle. Dies wird besonders bei der Betrachtung der Spezifitäten an der klinisch relevanten Sensitivitätsgrenze von 95 % deutlich (siehe Tabelle 3). Hier liegen die Spezifitäten der Netzwerke zwischen 20,7 % und 25,0 %. Verglichen mit %fPSA (22,6 %) können die multivariaten Modelle allerdings keine Spezifitätssteigerung an der 95 % Sensitivitätsgrenze vorweisen.

In der „ProstataClass“ Kohorte steigern dagegen alle fünf Netzwerke die Spezifität der PCa-Voraussage signifikant im Vergleich zu %fPSA sowohl bei 95 % als auch bei 90 % Sensitivität. Das überlegene Abschneiden der ANN verglichen mit %fPSA lässt sich für diese Kohorte sowohl anhand des Spezifitätsvergleichs als auch anhand der ROC-Kurvenanalyse belegen. Alle fünf ANN-Modelle erreichen signifikant höhere AUC-Werte als das %fPSA. Der Vergleich der Netzwerke untereinander zeigt, dass die verschiedenen Netzwerke auch für die „ProstataClass“ Kohorte ähnlich abschneiden. Die Ergebnisse dieser Arbeit bestätigen eine Studie von Stephan et al., die nahelegt, dass verschiedene multivariate Modelle, auf die selbe Kohorte angewandt, sehr ähnlich abschneiden ([141](#)).

Als ein wesentliches Ergebnis dieser Arbeit sind die fehlenden deutlichen Unterschiede zwischen den auf dem Elecsys-Assay basierenden ANN-Modellen (ANNpc, ANNiv und ANNall) und den beiden auf anderen Assays basierenden externen ANN-Modellen (MLP und ANN Hamburg) zu verzeichnen. Diese hier erzielten Resultate widersprechen somit früheren Studien, die eine deutliche Abhängigkeit der diagnostischen Genauigkeit vom verwendeten PSA- bzw. fPSA-Assay zeigen ([133;140](#)). Es ist festzustellen, dass der Einfluss der Kohorte auf die Ergebnisse der verschiedenen ANN-Modelle offenbar größer ist, als der Einfluss des gewählten PSA-Testsystems. Ferner belegt das ähnliche Abschneiden der verschiedenen Modelle für jeweils dieselbe Kohorte, dass das bessere Ergebnis des ANNpc für die „ProstataClass“ Kohorte nicht durch Überanpassung (overfitting) des Netzwerkes auf seine Trainingspopulation zu Stande kommt, sondern alle Netzwerke durch die Kohortenunterschiede ähnlich beeinflusst werden. Dem Problem der Überanpassung kann folglich mit der Bayesschen-Regulierung ausreichend begegnet werden.

Insgesamt zeigen die AUC-Werte und die Spezifitäten an den typischen Sensitivitäts-

grenzen, dass alle Netzwerke für die „ProstataClass“ Kohorte bessere Voraussagen treffen als für die „Interne Validierung“ Kohorte. Allerdings bleibt die Steigerung der diagnostischen Genauigkeit verglichen mit tPSA für die „ProstataClass“ Kohorte verhältnismäßig gering (Steigerung der AUC $< 0,2$). Dies ist vor allem auf das gute Abschneiden von tPSA bezüglich dieser Kohorte (AUC = 0,7) zurückzuführen. Im Fall der „Interne Validierung“ Kohorte zeigen die Netzwerke eine deutlichere Verbesserung der diagnostischen Genauigkeit im Vergleich zu tPSA (Steigerung der AUC zwischen 0,2 und 0,24). Dort hatte das tPSA jedoch eine deutlich geringere AUC von 0,527.

Die beiden Kohorten lassen sich insgesamt gut vergleichen. Es liegen keine signifikanten Unterschiede bei Alter und Prostatavolumen vor. Der Anteil der Karzinompatienten an der jeweiligen Population unterscheidet sich nur um 1 % und auch die typischen Unterschiede zwischen PCa-Patienten und NEM-Patienten finden sich in beiden Kohorten. Eine Ausnahme stellen die tPSA-Werte dar, die in der „Interne Validierung“ Kohorte keine signifikanten Unterschiede zwischen Karzinompatienten und Patienten ohne Anhalt auf Malignität zeigen. Das ANNiv, welches auf dieser Kohorte basiert, zeigt dessen ungeachtet ähnliche Ergebnisse im Vergleich zu den anderen Netzwerken. Dies lässt darauf schließen, dass die tPSA-Werte alleine keinen wesentlichen Einfluss auf die Voraussage des Karzinomrisikos innerhalb der ANN haben.

Für die Beurteilung von multivariaten Modellen oder neuen Screeningparametern, die der Risikoeinschätzung für eine Erkrankung dienen, ist der Vergleich der Flächen unter der ROC-Kurve (AUC) das meistgenutzte statistische Verfahren (144-146). Die AUC-Werte beschreiben aber ausschließlich, wie gut das Modell zwischen „krank“ und „nicht krank“ unterscheiden kann. Die ROC-Kurvenanalyse gibt keine Informationen über die vorhergesagten Wahrscheinlichkeiten. Beispielsweise könnte ein multivariates Modell in allen „krank“-Fällen einen Wahrscheinlichkeitswert von 0,52 und in allen „nicht krank“-Fällen von 0,51 angeben und damit ideale Diskriminationsfähigkeit aufweisen, obschon die angegebenen Wahrscheinlichkeiten klinisch nicht besonders hilfreich wären (144). Weil aber gerade die Wahrscheinlichkeit im klinischen Alltag von großer Bedeutung ist, reicht der Vergleich von AUC-Werten zur Evaluierung von Vorhersagemodellen nicht aus. Zumal invasive Folgeuntersuchungen aus dem Ergebnis resultieren könnten. Folglich sollte eine klinische Evaluation von Markern oder Modellen sowohl die Analyse der ROC-Kurven als auch eine Methode zur Beurteilung

der vorausgesagten Wahrscheinlichkeiten beinhalten.

Der Vergleich der vorausgesagten PCa-Wahrscheinlichkeiten mit den histologisch gesicherten Karzinomraten (siehe Abb. 6 – 10) zeigt, dass es neben dem Vergleich der AUC-Werte weiterer Evaluationsmethoden bedarf. Obgleich die Netzwerke sehr ähnliche AUC-Werte zeigen, sind in den Kalibrierungskurven deutliche Unterschiede zwischen den Netzwerken ersichtlich. Erwartungsgemäß schneiden die Netzwerke in den Kohorten mit denen sie trainiert wurden besser ab, als die ANN-Modelle, die auf anderen Kohorten basieren. Interessanterweise erreicht das ANN_{all} für die „Interne Validierung“ Kohorte einen höheren AUC-Wert (AUC = 0,774) als das ausschließlich auf dieser Kohorte konstruierte ANN_{iv} (AUC = 0,754). Erst die Kalibrierungskurven zeigen das bessere Abschneiden des ANN_{iv} (ICC = 0,901; RV = 0,009) gegenüber dem ANN_{all} (ICC = 0,766; RV = 0,024) für diese Kohorte. Das ANN_{pc} bleibt bei der Validierung anhand der prospektiven Daten der „Interne Validierung“ Kohorte hinter den Erwartungen zurück, was sowohl die ROC-Kurven-Analyse (AUC = 0,75) als auch die Kalibrierung (ICC = 0,609; RV = 0,067) belegen. Verglichen mit der „ProstataClass“ Kohorte schneidet das ANN_{pc} bei der prospektiven „Interne Validierung“ Kohorte deutlich ungünstiger ab. Die beiden externen ANN_{mlp} und ANN_{ham} haben für die „Interne Validierung“ Kohorte beinahe identische AUC-Werte (0,726 bzw. 0,729), dennoch zeigen die Kalibrierungskurven deutliche Unterschiede zwischen den Netzwerken bezüglich der Übereinstimmung von vorausgesagten Wahrscheinlichkeiten und bestätigten PCa-Raten. Während das ANN_{ham} für die „Interne Validierung“ Kohorte einen ICC von 0,773 (RV = 0,008) erreicht, liegt der ICC des ANN_{mlp} bei sehr niedrigen 0,327 (RV = 0,110). Dieser Vorteil seitens des ANN_{ham} (128) ist am ehesten auf die größere Ähnlichkeit der zugrunde liegenden urologisch eingewiesenen Population zurückzuführen. Die Screeningpopulation des ANN_{mlp} (112) weist geringere Ähnlichkeit mit der „Interne Validierung“ Kohorte auf und schneidet daher schlechter ab. Solche wesentlichen Erkenntnisse bleiben in Studien verborgen, die multivariate Modelle ausschließlich anhand von AUC-Werten validieren (147).

Die vorliegende Arbeit verdeutlicht die Wichtigkeit zusätzlicher Evaluationsmethoden zur Ergänzung der ROC-Kurvenanalyse. Die Konkordanz zwischen den vorausgesagten PCa-Wahrscheinlichkeiten und den histologisch bestätigten Karzinomraten erweist sich als aussagekräftiges Bewertungsmodell mit großer klinischer Relevanz und sollte zur Ergänzung der ROC-Kurvenanalyse genutzt werden.

Der Vergleich der AUC-Werte des %fPSA und tPSA für die prospektive „Interne Validierung“ Kohorte zeigt signifikant höhere AUC-Werte für das %fPSA. Auch bezüglich der Spezifitäten ist das %fPSA dem tPSA überlegen. Somit bestätigen die Daten dieser Arbeit erneut die Ergebnisse früherer Arbeiten, die die Überlegenheit von %fPSA gegenüber tPSA in der Diagnostik des PCa gezeigt haben (30;31;148;149). Daten dazu wurden bereits mehrfach in Übersichtsarbeiten zusammengetragen (15;28;38;44;56).

Die vorliegende Arbeit zeichnet sich besonders durch das prospektive Studiendesign bezüglich der „Interne Validierung“ Kohorte aus. Im Bereich der Entwicklung und Validierung multivariater Modelle wird zumeist auf retrospektiv erhobene Daten zurückgegriffen. Die Tatsache, dass für die „Interne Validierung“ Kohorte unabhängige, prospektive Daten erhoben wurden, stellt diese Arbeit qualitativ mit externen Validierungen gleich. Mit 393 Patienten ist die prospektive Kohorte verhältnismäßig groß. Zudem vergleicht diese Arbeit mehrere ANN anhand einer prospektiven und einer retrospektiven Kohorte. Damit unterscheidet sich diese Studie von üblichen externen Validierungen, die beispielsweise ein multivariates Modell mit diagnostischen Parametern wie tPSA und %fPSA oder mit den Ergebnissen für die Ursprungspopulation vergleichen. Erst der Vergleich mehrerer, auch externer Netzwerke anhand derselben Kohorte erlaubt es, vermeintliche Unterschiede zwischen den multivariaten Modellen für dieselbe Kohorte darzustellen.

Ein weiterer Vorteil dieser Arbeit liegt in der umfangreichen Bewertung der Qualität der ANN mittels verschiedener Methoden. Erst durch die Anwendung von Methoden zur Diskrimination und von Methoden zur Kalibrierung ist eine differenzierte, klinisch relevante Bewertung der ANN-Modelle möglich.

Die Arbeit vergleicht ausschließlich ANN und lässt folglich keine Einschätzung möglicher Kohorteneinflüsse auf die Ergebnisse von anderen multivariaten Modellen wie beispielsweise Nomogrammen zu. Ferner beschränkt sich diese Studie auf den Vergleich von nur fünf ANN-Modellen. Zukünftige Arbeiten sollten weitere Netzwerke einbeziehen und diese gegebenenfalls mit Nomogrammen anhand prospektiver Daten vergleichen. Schließlich könnten auch ANN-Modelle, die zusätzliche Eingangsparameter neben tPSA, %fPSA, Alter, Prostatavolumen und DRU-Status verwenden, in weitere Vergleichsstudien eingeschlossen werden. Nicht immer gelingt es aber,

erfolgreiche Parameter wie z.B. die PSA-Anstiegsgeschwindigkeit (PSA-Velocity) (64) erfolgreich in ANN-Modelle zu integrieren (150). Die ANN, die in der vorliegenden Arbeit verglichen werden, sind alle zur Vorhersage der PCa-Wahrscheinlichkeit bei Prostatabiopsie konzipiert. Es werden keine Modelle zur Verlaufsprognose des PCa einbezogen. In Zukunft sollten aber im Zuge der Einführung neuer ANN-Modelle zur Prognoseabschätzung ähnliche Studien mit diesen Modellen folgen. Für den Patienten ist mittlerweile weniger die Diagnose PCa von Bedeutung als die Prognose der vorliegenden Erkrankung. Es besteht daher besonders großer Bedarf Modelle zu entwickeln und zu validieren, die zwischen aggressiven und weniger aggressiven Formen des PCa unterscheiden. Erst eine solche Differenzierung ließe eine verbesserte individuelle Anpassung der Therapie zu, wodurch die Zahl der Überbehandlungen reduziert werden könnte.

Alle in dieser Studie verwendeten ANN-Modelle errechnen ein sehr differenziertes Karzinomrisiko, indem sie Werte zwischen 0 (geringstes Risiko) und 1 (höchstes Risiko) angeben. Hierbei bleibt der klinische Nutzen einer prozentgenauen Risikobewertung fraglich. Solange keine Grenzwerte für die Biopsieindikation etabliert sind, müssen Arzt und Patient anhand des Risikos eine Entscheidung zur Prostatastanzbiopsie und bei Anwendung ähnlicher prognostischer ANN-Modelle auch bezüglich weiterer therapeutischer Schritte treffen. Ein Vorteil von zwei der hier verwendeten ANN-Modelle (112;133) ist die praktische Anwendbarkeit von Grenzwerten, um damit die Entscheidung für oder gegen eine Biopsie zu erleichtern. Bei Nomogrammen (121;123;128) oder anderen nicht frei im Internet zugänglichen ANN-Modellen (110;111) wird oftmals nur ein relatives PCa-Risiko angegeben ohne die Angabe, ob ein bestimmter Grenzwert z.B. bei 95% Sensitivität überschritten wird.

Die vorliegende Arbeit zeigt den wesentlichen Einfluss der zu beurteilenden Kohorte auf das Abschneiden von ANN, welche tPSA, %fPSA, Alter, Prostatavolumen und DRU-Status als Eingangsparameter verwenden. Dies wirft erneut die Frage auf, in wie weit multivariate Modelle tatsächlich in unabhängigen Populationen angewandt werden können. Zudem zeigt das schwächere Abschneiden der ANN gegenüber %fPSA für die „Interne Validierung“ Kohorte, dass alle multivariaten Modelle anhand unabhängiger, möglichst prospektiver Daten evaluiert werden müssen.

Zusätzlich verdeutlicht diese Arbeit die Wichtigkeit einer differenzierten Bewertung

multivariater Modelle mittels Kalibrierung und Diskrimination. Es kann klar gezeigt werden, dass die Analyse der ROC-Kurven wie auch die Spezifitätsvergleiche keine Beurteilung des tatsächlichen klinischen Nutzens eines Modells ermöglichen. Zur differenzierten Betrachtung des Abschneidens verschiedener Modelle bedarf es der Gegenüberstellung von vorausgesagten PCa-Wahrscheinlichkeiten mit den histologisch gesicherten PCa-Raten.

6 Zusammenfassung

Multivariate Modelle wie ANN finden verstärkt Anwendung in der PCa-Diagnostik. Besonders im Bereich der Vorhersage von Biopsieergebnissen werden multivariate Modelle zur Steigerung der Spezifität der etablierten Marker tPSA und %fPSA eingesetzt. Fraglich ist allerdings, ob multivariate Modelle für Populationen, die nicht ihrer Trainingspopulation entsprechen, valide Voraussagen treffen können. Validierungsstudien, bei denen ANN-Modelle anhand unabhängiger Populationen bewertet werden, sind von besonderer Wichtigkeit.

Die vorliegende Arbeit validiert das online verfügbare ANN „ProstataClass“ (ANNpc) anhand prospektiver Daten von 393 Patienten und vergleicht die Ergebnisse des ANNpc mit vier weiteren ANN-Modellen.

Von allen Patienten wurden die Eingangsparameter tPSA, %fPSA, Alter, DRU-Status und Prostatavolumen erhoben. Die tPSA- und fPSA-Werte wurden mittels Elecsys 2010[®]-Test der Firma Roche (Roche Diagnostics GmbH, Penzberg, Deutschland) ermittelt. Zur Bewertung der diagnostischen Genauigkeit des ANNpc, des ANNiv, des ANNall und der externen ANN-Modelle (ANNmlp und ANNham) wurden die AUC-Werte und die Spezifitäten an den Sensitivitätsgrenzen von 90 % und 95 % verglichen. Zusätzlich wurden die vorausgesagten PCa-Wahrscheinlichkeiten den beobachteten PCa-Raten gegenübergestellt.

Von den 393 Patienten der „Interne Validierung“ Kohorte liegt bei 229 Männern (58,3 %) ein histologisch gesichertes PCa und bei 164 Männern (41,7 %) kein Anhalt für Malignität vor. Die tPSA-Werte liegen für diese Kohorte zwischen 1,0 µg/L und 22,8 µg/L und die tPSA- und %fPSA-Werte, sowie der Anteil der Patienten mit suspekter DRU weisen signifikante Unterschiede zwischen der prospektiven „Interne Validierung“ Kohorte und der Ursprungkohorte des ANNpc („ProstataClass“ Kohorte) auf. Bezüglich Alter der Patienten und Prostatavolumen bestehen keine signifikanten Unterschiede zwischen den Kohorten. Die AUC-Werte von tPSA und %fPSA sind 0,527 bzw. 0,726. Das ANNpc erreicht einen AUC-Wert von 0,747 und zeigt somit keine signifikante Steigerung der diagnostischen Genauigkeit verglichen mit %fPSA ($P = 0,085$). Für das ANNiv und das ANNall liegen die AUC-Werte mit 0,754 und 0,774 signifikant höher als für %fPSA ($P = 0,021$ bzw. $0,0008$). Die AUC-Werte für die

externen ANN-Modelle (0,726 und 0,729) zeigen indes keine signifikanten Unterschiede im Vergleich zu %fPSA. Letztlich lassen sich zwischen allen fünf ANN-Modellen keine signifikanten Unterschiede bezüglich der AUC-Werte feststellen. Der Vergleich der vorausgesagten PCa-Wahrscheinlichkeiten mit den histologisch gesicherten Karzinomraten zeigt aber deutliche Unterschiede zwischen allen ANN-Modellen.

Das ungünstigere Abschneiden des ANNpc in der hier getesteten prospektiven Kohorte ist wahrscheinlich auf Unterschiede zwischen der Ursprungkohorte des ANNpc und der prospektiven Validierungskohorte zurückzuführen. Der Anteil der Patienten mit suspekter DRU ist in der Validierungskohorte um etwa 30 % geringer. Dies führt zum Unterschätzen des PCa-Risikos. Zudem liegen die %fPSA Mittel- und Medianwerte der prospektiven Kohorte 3,7 % bzw. 3,6 % höher als die der „ProstataClass“ Kohorte, was vermutlich durch die verlängerte Lagerung der Serumproben der „ProstataClass“ Kohorte zustande kommt. Es können keine Unterschiede zwischen den ANN-Modellen festgestellt werden, obgleich den externen Modellen Daten zugrunde liegen, die mittels anderer tPSA und fPSA Assays gemessen wurden. Es kann klar gezeigt werden, dass das ANNpc aufgrund von Kohortenunterschieden die signifikante Steigerung der AUC verglichen mit %fPSA nicht wiederholen konnte. Außerdem wird deutlich, dass der Vergleich von AUC-Werten und Spezifitäten zur Validierung von multivariaten Modellen nicht ausreichend ist. Ergänzend zu den Methoden der Diskriminierung sollten zukünftig Methoden der Kalibrierung, wie der Vergleich von vorausgesagten PCa-Wahrscheinlichkeiten und bestätigten PCa-Raten, genutzt werden.

7 Literatur

1. Batzler WU, Giersiepen K, Hentschel S, et al. Krebs in Deutschland 2003 - 2004. Häufigkeiten und Trends. 6. überarbeitete Auflage. 2008. Berlin.
2. Jemal A, Siegel R, Ward E, Hao Y, Xu J, Thun MJ. Cancer statistics, 2009. *CA Cancer J Clin* 2009;59:225-49.
3. American Cancer Society. Cancer Facts & Figures 2009. Atlanta: American Cancer Society; 2009.
4. Bostwick DG, Burke HB, Djakiew D, et al. Human prostate cancer risk factors. *Cancer* 2004;101:2371-490.
5. Bestmann B, Loetters C, Diemer T, Weidner W, Kuchler T, Rohde V. Prostate-specific symptoms of prostate cancer in a German general population. *Prostate Cancer Prostatic Dis* 2007;10:52-9.
6. Lilja H, Ulmert D, Vickers AJ. Prostate-specific antigen and prostate cancer: prediction, detection and monitoring. *Nat Rev Cancer* 2008;8:268-78.
7. Han M, Partin AW, Piantadosi S, Epstein JI, Walsh PC. Era specific biochemical recurrence-free survival following radical prostatectomy for clinically localized prostate cancer. *J Urol* 2001;166:416-9.
8. Schroder FH, Hugosson J, Roobol MJ, et al. Screening and prostate-cancer mortality in a randomized European study. *N Engl J Med* 2009;360:1320-8.
9. Roobol MJ, Kerkhof M, Schroder FH, et al. Prostate cancer mortality reduction by prostate-specific antigen-based screening adjusted for nonattendance and contamination in the European Randomised Study of Screening for Prostate Cancer (ERSPC). *Eur Urol* 2009;56:584-91.
10. Andriole GL, Crawford ED, Grubb RL, III, et al. Mortality results from a randomized prostate-cancer screening trial. *N Engl J Med* 2009;360:1310-9.
11. Draisma G, Boer R, Otto SJ, et al. Lead times and overdiagnosis due to prostate-specific antigen screening: estimates from the European Randomized Study of Screening for Prostate Cancer. *J Natl Cancer Inst* 2003;95:868-78.
12. Smith RA, Cokkinides V, Brawley OW. Cancer screening in the United States, 2009: a review of current American Cancer Society guidelines and issues in cancer screening. *CA Cancer J Clin* 2009;59:27-41.
13. Bonkhoff K, Fornara P, Gleißner J, et al. Leitlinie PSA-Bestimmung in der Prostatakarzinomdiagnostik (Früherkennung des Prostatakarzinoms). Deutsche Krebsgesellschaft e. V. 2002. http://www.krebsgesellschaft.de/download/ebmleitlinie_frueherkennung_des_prostatakarzinoms.pdf (Stand 26.02.2010)

14. Finne P, Zhang WM, Auvinen A, et al. Use of the complex between prostate specific antigen and alpha 1-protease inhibitor for screening prostate cancer. *J Urol* 2000;164:1956-60.
15. Polascik TJ, Oesterling JE, Partin AW. Prostate specific antigen: a decade of discovery--what we have learned and where we are going. *J Urol* 1999;162:293-306.
16. Hara M, Koyanagi Y, Inoue T, Fukuyama T. Some physico-chemical characteristics of " γ -seminoprotein", an antigenic component specific for human seminal plasma. Forensic immunological study of body fluids and secretion. *Nippon Hoigaku Zasshi* 1971;25:322-4.
17. Wang MC, Valenzuela LA, Murphy GP, Chu TM. Purification of a human prostate specific antigen. *Invest Urol* 1979;17:159-63.
18. Papsidero LD, Wang MC, Valenzuela LA, Murphy GP, Chu TM. A prostate antigen in sera of prostatic cancer patients. *Cancer Res* 1980;40:2428-32.
19. Yousef GM, Diamandis EP. The new human tissue kallikrein gene family: structure, function, and association to disease. *Endocr Rev* 2001;22:184-204.
20. Lundwall A, Lilja H. Molecular cloning of human prostate specific antigen cDNA. *FEBS Lett* 1987;214:317-22.
21. Watt KW, Lee PJ, M'Timkulu T, Chan WP, Loo R. Human prostate-specific antigen: structural and functional similarity with serine proteases. *Proc Natl Acad Sci U S A* 1986;83:3166-70.
22. Lilja H, Oldbring J, Rannevik G, Laurell CB. Seminal vesicle-secreted proteins and their reactions during gelation and liquefaction of human semen. *J Clin Invest* 1987;80:281-5.
23. Stephan C, Jung K, Lein M, Diamandis EP. PSA and other tissue kallikreins for prostate cancer detection. *Eur J Cancer* 2007;43:1918-26.
24. Stenman UH, Leinonen J, Alfthan H, Rannikko S, Tuhkanen K, Alfthan O. A complex between prostate-specific antigen and alpha 1-antichymotrypsin is the major form of prostate-specific antigen in serum of patients with prostatic cancer: assay of the complex improves clinical sensitivity for cancer. *Cancer Res* 1991;51:222-6.
25. Lilja H, Christensson A, Dahlen U, et al. Prostate-specific antigen in serum occurs predominantly in complex with alpha 1-antichymotrypsin. *Clin Chem* 1991;37:1618-25.
26. Christensson A, Björk T, Nilsson O, et al. Serum prostate specific antigen complexed to alpha 1-antichymotrypsin as an indicator of prostate cancer. *J Urol* 1993;150:100-5.

27. Catalona WJ. Clinical utility of measurements of free and total prostate-specific antigen (PSA): a review. *Prostate Suppl* 1996;7:64-9.
28. Woodrum DL, Brawer MK, Partin AW, Catalona WJ, Southwick PC. Interpretation of free prostate specific antigen clinical research studies for the detection of prostate cancer. *J Urol* 1998;159:5-12.
29. Catalona WJ, Smith DS, Wolfert RL, et al. Evaluation of percentage of free serum prostate-specific antigen to improve specificity of prostate cancer screening. *JAMA* 1995;274:1214-20.
30. Catalona WJ, Partin AW, Slawin KM, et al. Use of the percentage of free prostate-specific antigen to enhance differentiation of prostate cancer from benign prostatic disease: a prospective multicenter clinical trial. *JAMA* 1998;279:1542-7.
31. Van Cangh PJ, De Nayer P, De Vischer L, et al. Free to total prostate-specific antigen (PSA) ratio improves the discrimination between prostate cancer and benign prostatic hyperplasia (BPH) in the diagnostic gray zone of 1.8 to 10 ng/mL total PSA. *Urology* 1996;48:67-70.
32. Lee R, Localio AR, Armstrong K, Malkowicz SB, Schwartz JS. A meta-analysis of the performance characteristics of the free prostate-specific antigen test. *Urology* 2006;67:762-8.
33. Catalona WJ, Partin AW, Finlay JA, et al. Use of percentage of free prostate-specific antigen to identify men at high risk of prostate cancer when PSA levels are 2.51 to 4 ng/mL and digital rectal examination is not suspicious for prostate cancer: an alternative model. *Urology* 1999;54:220-4.
34. Djavan B, Zlotta A, Kratzik C, et al. PSA, PSA density, PSA density of transition zone, free/total PSA ratio, and PSA velocity for early detection of prostate cancer in men with serum PSA 2.5 to 4.0 ng/mL. *Urology* 1999;54:517-22.
35. Jung K, Stephan C, Elgeti U, et al. Molecular forms of prostate-specific antigen in serum with concentrations of total prostate-specific antigen <4 µg/l - are they useful tools for early detection and screening of prostate cancer? *Int J Cancer* 2001;93:759-65.
36. Roehl KA, Antenor JA, Catalona WJ. Robustness of free prostate specific antigen measurements to reduce unnecessary biopsies in the 2.6 to 4.0 ng./ml. range. *J Urol* 2002;168:922-5.
37. Raaijmakers R, Blijenberg BG, Finlay JA, et al. Prostate cancer detection in the prostate specific antigen range of 2.0 to 3.9 ng/ml: value of percent free prostate specific antigen on tumor detection and tumor aggressiveness. *J Urol* 2004;171:2245-9.
38. Stephan C, Rittenhouse H, Cammann H, et al. New markers and multivariate

models for prostate cancer detection. *Anticancer Res* 2009;29:2589-600.

39. Zhang WM, Finne P, Leinonen J, et al. Characterization and immunological determination of the complex between prostate-specific antigen and alpha2-macroglobulin. *Clin Chem* 1998;44:2471-9.
40. Zhang WM, Finne P, Leinonen J, Vesalainen S, Nordling S, Stenman UH. Measurement of the complex between prostate-specific antigen and alpha1-protease inhibitor in serum. *Clin Chem* 1999;45:814-21.
41. Allard WJ, Zhou Z, Yeung KK. Novel immunoassay for the measurement of complexed prostate-specific antigen in serum. *Clin Chem* 1998;44:1216-23.
42. Brawer MK, Meyer GE, Letran JL, et al. Measurement of complexed PSA improves specificity for early detection of prostate cancer. *Urology* 1998;52:372-8.
43. Lein M, Kwiatkowski M, Semjonow A, et al. A multicenter clinical trial on the use of complexed prostate specific antigen in low prostate specific antigen concentrations. *J Urol* 2003;170:1175-9.
44. Roddam AW, Duffy MJ, Hamdy FC, et al. Use of Prostate-Specific Antigen (PSA) Isoforms for the Detection of Prostate Cancer in Men with a PSA Level of 2-10 ng/ml: Systematic Review and Meta-Analysis. *Eur Urol* 2005;48:386-99.
45. Mikolajczyk SD, Marks LS, Partin AW, Rittenhouse HG. Free prostate-specific antigen in serum is becoming more complex. *Urology* 2002;59:797-802.
46. Peter J, Unverzagt C, Krogh TN, Vorm O, Hoesel W. Identification of precursor forms of free prostate-specific antigen in serum of prostate cancer patients by immunosorption and mass spectrometry. *Cancer Res* 2001;61:957-62.
47. Mikolajczyk SD, Marker KM, Millar LS, et al. A truncated precursor form of prostate-specific antigen is a more specific serum marker of prostate cancer. *Cancer Res* 2001;61:6958-63.
48. Mikolajczyk SD, Rittenhouse HG. Pro PSA: a more cancer specific form of prostate specific antigen for the early detection of prostate cancer. *Keio J Med* 2003;52:86-91.
49. Catalona WJ, Bartsch G, Rittenhouse HG, et al. Serum pro prostate specific antigen improves cancer detection compared to free and complexed prostate specific antigen in men with prostate specific antigen 2 to 4 ng/ml. *J Urol* 2003;170:2181-5.
50. Sokoll LJ, Wang Y, Feng Z, et al. [-2]proenzyme prostate specific antigen for prostate cancer detection: a national cancer institute early detection research network validation study. *J Urol* 2008;180:539-43.

51. Stephan C, Kahrs AM, Cammann H, et al. A [-2]proPSA-based artificial neural network significantly improves differentiation between prostate cancer and benign prostatic diseases. *Prostate* 2009;69:198-207.
52. Lein M, Semjonow A, Graefen M, et al. A multicenter clinical trial on the use of (-5, -7) pro prostate specific antigen. *J Urol* 2005;174:2150-3.
53. Bangma CH, Wildhagen MF, Yurdakul G, Schroder FH, Blijenberg BG. The value of (-7, -5)pro-prostate-specific antigen and human kallikrein-2 as serum markers for grading prostate cancer. *BJU Int* 2004;93:720-4.
54. Mikolajczyk SD, Millar LS, Wang TJ, et al. "BPSA," a specific molecular form of free prostate-specific antigen, is found predominantly in the transition zone of patients with nodular benign prostatic hyperplasia. *Urology* 2000;55:41-5.
55. Stephan C, Cammann H, Deger S, et al. Benign prostatic hyperplasia-associated free prostate-specific antigen improves detection of prostate cancer in an artificial neural network. *Urology* 2009;74:873-7.
56. Haese A, Noldus J, Steuber T, Huland H, Lilja H. Isoformen des freien prostataspezifischen Antigens. *Urologe A* 2004;43:675-9.
57. Steuber T, Nurmikko P, Haese A, et al. Discrimination of benign from malignant prostatic disease by selective measurements of single chain, intact free prostate specific antigen. *J Urol* 2002;168:1917-22.
58. Stephan C, Jung K, Diamandis EP, Rittenhouse HG, Lein M, Loening SA. Prostate specific antigen, its molecular forms, and other kallikrein markers for detection of prostate cancer. *Urology* 2002;59:2-8.
59. Oesterling JE, Jacobsen SJ, Chute CG. Serum prostate-specific antigen in a community-based population of healthy men: establishment of age-specific reference ranges. *JAMA* 1993;270:860-6.
60. Oesterling JE, Jacobsen SJ, Cooner WH. The use of age-specific reference ranges for serum prostate specific antigen in men 60 years old or older. *J Urol* 1995;153:1160-3.
61. Carter HB, Pearson JD, Metter EJ, et al. Longitudinal evaluation of prostate-specific antigen levels in men with and without prostate disease. *JAMA* 1992;267:2215-20.
62. Schroder FH, Roobol MJ, van der Kwast TH, Kranse R, Bangma CH. Does PSA velocity predict prostate cancer in pre-screened populations? *Eur Urol* 2006;49:460-5.
63. Berger AP, Deibl M, Strasak A, et al. Large-scale study of clinical impact of PSA velocity: long-term PSA kinetics as method of differentiating men with from those without prostate cancer. *Urology* 2007;69:134-8.

64. Loeb S, Roehl KA, Catalona WJ, Nadler RB. Prostate specific antigen velocity threshold for predicting prostate cancer in young men. *J Urol* 2007;177:899-902.
65. Komatsu K, Wehner N, Prestigiacomo AF, Chen Z, Stamey TA. Physiologic (intraindividual) variation of serum prostate-specific antigen in 814 men from a screening population. *Urology* 1996;47:343-6.
66. Nixon RG, Wener MH, Smith KM, Parson RE, Strobel SA, Brawer MK. Biological variation of prostate specific antigen levels in serum: an evaluation of day-to-day physiological fluctuations in a well-defined cohort of 24 patients. *J Urol* 1997;157:2183-90.
67. Stephan C, Klaas M, Muller C, Schnorr D, Loening SA, Jung K. Interchangeability of measurements of total and free prostate-specific antigen in serum with 5 frequently used assay combinations: an update. *Clin Chem* 2006;52:59-64.
68. Benson MC, Whang IS, Pantuck A, et al. Prostate specific antigen density: a means of distinguishing benign prostatic hypertrophy and prostate cancer. *J Urol* 1992;147:815-6.
69. Stephan C, Stroebel G, Heinau M, et al. The ratio of prostate-specific antigen (PSA) to prostate volume (PSA density) as a parameter to improve the detection of prostate carcinoma in PSA values in the range of < 4 ng/mL. *Cancer* 2005;104:993-1003.
70. Veneziano S, Pavlica P, Querze R, Nanni G, Lalanne MG, Vecchi F. Correlation between prostate-specific antigen and prostate volume, evaluated by transrectal ultrasonography: usefulness in diagnosis of prostate cancer. *Eur Urol* 1990;18:112-6.
71. Thon WF, Gadban F, Truss MC, Kuczyk M, Hartmann U, Jonas U. Prostate-specific antigen density--a reliable parameter for the detection of prostate cancer? *World J Urol* 1996;14:53-8.
72. Stamey TA, Kabalin JN, McNeal JE, et al. Prostate specific antigen in the diagnosis and treatment of adenocarcinoma of the prostate. II. Radical prostatectomy treated patients. *J Urol* 1989;141:1076-83.
73. Lundwall A, Band V, Blaber M, et al. A comprehensive nomenclature for serine proteases with homology to tissue kallikreins. *Biol Chem* 2006;387:637-41.
74. Partin AW, Catalona WJ, Finlay JA, et al. Use of human glandular kallikrein 2 for the detection of prostate cancer: preliminary analysis. *Urology* 1999;54:839-45.
75. Kwiatkowski MK, Recker F, Piironen T, et al. In prostatism patients the ratio of human glandular kallikrein to free PSA improves the discrimination between prostate cancer and benign hyperplasia within the diagnostic "gray zone" of total PSA 4 to 10 ng/mL. *Urology* 1998;52:360-5.

76. Nakamura T, Scorilas A, Stephan C, Jung K, Soosaipillai AR, Diamandis EP. The usefulness of serum human kallikrein 11 for discriminating between prostate cancer and benign prostatic hyperplasia. *Cancer Res* 2003;63:6543-6.
77. Rabien A, Fritzsche F, Jung M, et al. High expression of KLK14 in prostatic adenocarcinoma is associated with elevated risk of prostate-specific antigen relapse. *Tumour Biol* 2008;29:1-8.
78. Sardana G, Dowell B, Diamandis EP. Emerging biomarkers for the diagnosis and prognosis of prostate cancer. *Clin Chem* 2008;54:1951-60.
79. Brown DA, Stephan C, Ward RL, et al. Measurement of serum levels of macrophage inhibitory cytokine 1 combined with prostate-specific antigen improves prostate cancer diagnosis. *Clin Cancer Res* 2006;12:89-96.
80. Stephan C, Xu C, Brown DA, et al. Three new serum markers for prostate cancer detection within a percent free PSA-based artificial neural network. *Prostate* 2006;66:651-9.
81. Meyer-Siegler KL, Bellino MA, Tannenbaum M. Macrophage migration inhibitory factor evaluation compared with prostate specific antigen as a biomarker in patients with prostate carcinoma. *Cancer* 2002;94:1449-56.
82. Michael A, Stephan C, Kristiansen G, et al. Diagnostic validity of macrophage migration inhibitory factor in serum of patients with prostate cancer: a re-evaluation. *Prostate* 2005;62:34-9.
83. Tahir SA, Ren C, Timme TL, et al. Development of an immunoassay for serum caveolin-1: a novel biomarker for prostate cancer. *Clin Cancer Res* 2003;9:3653-9.
84. Tahir SA, Frolov A, Hayes TG, et al. Preoperative serum caveolin-1 as a prognostic marker for recurrence in a radical prostatectomy cohort. *Clin Cancer Res* 2006;12:4872-5.
85. Stattin P, Rinaldi S, Biessy C, Stenman UH, Hallmans G, Kaaks R. High levels of circulating insulin-like growth factor-I increase prostate cancer risk: a prospective study in a population-based nonscreened cohort. *J Clin Oncol* 2004;22:3104-12.
86. Finne P, Auvinen A, Koistinen H, et al. Insulin-like growth factor I is not a useful marker of prostate cancer in men with elevated levels of prostate-specific antigen. *J Clin Endocrinol Metab* 2000;85:2744-7.
87. Oliver SE, Holly J, Peters TJ, et al. Measurement of insulin-like growth factor axis does not enhance specificity of PSA-based prostate cancer screening. *Urology* 2004;64:317-22.
88. Roddam AW, Allen NE, Appleby P, et al. Insulin-like growth factors, their binding proteins, and prostate cancer risk: analysis of individual patient data from 12

- prospective studies. *Ann Intern Med* 2008;149:461-8.
89. von der Kammer H, Jurincic-Winkler C, Horlbeck R, Klippel KF, Pixberg HU, Scheit KH. The potential use of prostatic secretory protein of 94 amino acid residues (PSP94) as a serum marker for prostatic tumor. *Urol Res* 1993;21:227-33.
 90. Nam RK, Reeves JR, Toi A, et al. A novel serum marker, total prostate secretory protein of 94 amino acids, improves prostate cancer detection and helps identify high grade cancers at diagnosis. *J Urol* 2006;175:1291-7.
 91. Reeves JR, Dulude H, Panchal C, Daigneault L, Ramnani DM. Prognostic value of prostate secretory protein of 94 amino acids and its binding protein after radical prostatectomy. *Clin Cancer Res* 2006;12:6018-22.
 92. Wozny W, Schroer K, Schwall GP, et al. Differential radioactive quantification of protein abundance ratios between benign and malignant prostate tissues: cancer association of annexin A3. *Proteomics* 2007;7:313-22.
 93. Kollermann J, Schlomm T, Bang H, et al. Expression and prognostic relevance of annexin A3 in prostate cancer. *Eur Urol* 2008;54:1314-23.
 94. Schostak M, Schwall GP, Poznanovic S, et al. Annexin A3 in urine: A highly specific noninvasive marker for prostate cancer early detection. *J Urol* 2009;181:343-53.
 95. Hessels D, Klein Gunnewiek JM, van Oort I, et al. DD3(PCA3)-based molecular urine analysis for the diagnosis of prostate cancer. *Eur Urol* 2003;44:8-15.
 96. de la Taille A, Irani J, de Reijke TM, Haese A. Can Prostate cancer gene (PCA3) predict initial biopsy outcome? [Abstract]. *J Urol [Suppl]* 2009;181:655.
 97. Haese A, de la Taille A, van Poppel H, et al. Clinical utility of the PCA3 urine assay in European men scheduled for repeat biopsy. *Eur Urol* 2008;54:1081-8.
 98. Nakanishi H, Groskopf J, Fritsche HA, et al. PCA3 molecular urine assay correlates with prostate cancer tumor volume: implication in selecting candidates for active surveillance. *J Urol* 2008;179:1804-9.
 99. Whitman EJ, Groskopf J, Ali A, et al. PCA3 score before radical prostatectomy predicts extracapsular extension and tumor volume. *J Urol* 2008;180:1975-8.
 100. Schilling DA, Hennenlotter J, von Weyhern CH, et al. Does the PCA3 score depend on tumor localization within the prostate? - A morphometric computer animated analysis. *J Urol [Suppl]* 2009;181:655.
 101. van Gils MP, Hessels D, Hulsbergen-van de Kaa CA, et al. Detailed analysis of histopathological parameters in radical prostatectomy specimens and PCA3 urine test results. *Prostate* 2008;68:1215-22.

102. Varambally S, Laxman B, Mehra R, et al. Golgi protein GOLM1 is a tissue and urine biomarker of prostate cancer. *Neoplasia* 2008;10:1285-94.
103. Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol Rev* 1958;65:386-408.
104. Widrow B and Hoff MEJr. Adaptive switching circuits. August IRE WESCON Convention Record 4, 96-104. 1960.
105. Widrow B, Stearns SD. Adaptive signal processing. Prentice Hall, Englewood Cliffs, NJ. 1985.
106. Werbos PJ. Beyond regression: New tools for prediction and analysis in the behavioral sciences. Harvard University. 1974.
107. Anagnostou T, Remzi M, Lykourinas M, Djavan B. Artificial neural networks for decision-making in urologic oncology. *Eur Urol* 2003;43:596-603.
108. Reckwitz T, Potter SR, Snow PB, Zhang Z, Veltri RW, Partin AW. Artificial neural networks in urology: Update 2000. *Prostate Cancer Prostatic Dis* 1999;2:222-6.
109. Snow PB, Smith DS, Catalona WJ. Artificial neural networks in the diagnosis and prognosis of prostate cancer: a pilot study. *J Urol* 1994;152:1923-6.
110. Babaian RJ, Fritsche H, Ayala A, et al. Performance of a neural network in detecting prostate cancer in the prostate-specific antigen reflex range of 2.5 to 4.0 ng/mL. *Urology* 2000;56:1000-6.
111. Djavan B, Remzi M, Zlotta A, Seitz C, Snow P, Marberger M. Novel artificial neural network for early detection of prostate cancer. *J Clin Oncol* 2002;20:921-9.
112. Finne P, Finne R, Auvinen A, et al. Predicting the outcome of prostate biopsy in screen-positive men by a multilayer perceptron network. *Urology* 2000;56:418-22.
113. Kalra P, Togami J, Bansal BSG, et al. A neurocomputational model for prostate carcinoma detection. *Cancer* 2003;98:1849-54.
114. Stephan C, Cammann H, Semjonow A, et al. Multicenter evaluation of an artificial neural network to increase the prostate cancer detection rate and reduce unnecessary biopsies. *Clin Chem* 2002;48:1279-87.
115. Stephan C, Jung K, Cammann H, et al. An artificial neural network considerably improves the diagnostic power of percent free prostate-specific antigen in prostate cancer diagnosis: results of a 5-year investigation. *Int J Cancer* 2002;99:466-73.
116. Schroder F, Kattan MW. The comparability of models for predicting the risk of a

positive prostate biopsy with prostate-specific antigen alone: a systematic review. *Eur Urol* 2008;54:274-90.

117. Stephan C, Cammann H, Meyer HA, Lein M, Jung K. PSA and new biomarkers within multivariate models to improve early detection of prostate cancer. *Cancer Lett* 2007;249:18-29.
118. Chun FK, Karakiewicz PI, Huland H, Graefen M. Role of nomograms for prostate cancer in 2007. *World J Urol* 2007;25:131-42.
119. Shariat SF, Capitanio U, Jeldres C, Karakiewicz PI. Can nomograms be superior to other prediction tools? *BJU Int* 2008; 103:492-5.
120. Eastham JA, May R, Robertson JL, Sartor O, Kattan MW. Development of a nomogram that predicts the probability of a positive prostate biopsy in men with an abnormal digital rectal examination and a prostate-specific antigen between 0 and 4 ng/mL. *Urology* 1999;54:709-13.
121. Chun FK, Briganti A, Graefen M, et al. Development and external validation of an extended 10-core biopsy nomogram. *Eur Urol* 2007;52:436-44.
122. Garzotto M, Hudson RG, Peters L, et al. Predictive modeling for the presence of prostate carcinoma using clinical, laboratory, and ultrasound parameters in patients with prostate specific antigen levels $<$ or $=$ 10 ng/mL. *Cancer* 2003;98:1417-22.
123. Karakiewicz PI, Benayoun S, Kattan MW, et al. Development and validation of a nomogram predicting the outcome of prostate biopsy based on patient age, digital rectal examination and serum prostate specific antigen. *J Urol* 2005;173:1930-4.
124. Chun FK, Briganti A, Graefen M, et al. Development and external validation of an extended repeat biopsy nomogram. *J Urol* 2007;177:510-5.
125. Kawakami S, Numao N, Okubo Y, et al. Development, validation, and head-to-head comparison of logistic regression-based nomograms and artificial neural network models predicting prostate cancer on initial extended biopsy. *Eur Urol* 2008;54:601-11.
126. Chun FK, Karakiewicz PI, Briganti A, et al. A critical appraisal of logistic regression-based nomograms, artificial neural networks, classification and regression-tree models, look-up tables and risk-group stratification models for prostate cancer. *BJU Int* 2007;99:794-800.
127. Finne P, Finne R, Bangma C, et al. Algorithms based on prostate-specific antigen (PSA), free PSA, digital rectal examination and prostate volume reduce false-positive PSA results in prostate cancer screening. *Int J Cancer* 2004;111:310-5.

128. Chun FK, Graefen M, Briganti A, et al. Initial biopsy outcome prediction--head-to-head comparison of a logistic regression-based nomogram versus artificial neural network. *Eur Urol* 2007;51:1236-40.
129. Sargent DJ. Comparison of artificial neural networks with other statistical approaches. *Cancer* 2001;91:1636-42.
130. Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review. *J Biomed Inform* 2002;35:352-9.
131. Stephan C, Meyer HA, Cammann H, Lein M, Loening SA, Jung K. Re: Felix K.-H. Chun, Markus Graefen, Alberto Briganti, Andrea Gallina, Julia Hopp, Michael W. Kattan, Hartwig Huland and Pierre I. Karakiewicz. Initial biopsy outcome prediction--head-to-head comparison of a logistic regression-based nomogram versus artificial neural network. *Eur Urol* 2007;51:1236-43. *Eur Urol* 2007;51:1446-7.
132. Stephan C, Vogel B, Cammann H, et al. Nutzung von artifiziellen neuronalen Netzwerken zur Risikoabschätzung eines Prostatakarzinoms: Biopsieindikationen im PSA-Bereich 2-20 µg/l. *Urologe A* 2003;42:1221-9.
133. Stephan C, Cammann H, Meyer HA, et al. An artificial neural network for five different assay systems of prostate-specific antigen in prostate cancer diagnostics. *BJU Int* 2008;102:799-805.
134. MacKay DJC. Bayesian Interpolation. *Neural Computation* 1992;4:415-47.
135. Stephan C, Jung K, Soosaipillai A, et al. Clinical utility of human glandular kallikrein 2 within a neural network for prostate cancer detection. *BJU Int* 2005;96:521-7.
136. Jung K, Stephan C, Lein M, et al. Receiver-operating characteristic as a tool for evaluating the diagnostic performance of prostate-specific antigen and its molecular forms--What has to be considered? *Prostate* 2001;46:307-10.
137. Kairisto V, Poola A. Software for illustrative presentation of basic clinical characteristics of laboratory tests - GraphROC for Windows. *Scand J Clin Lab Invest* 1995;55:43-60.
138. Lin LI. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 1989;45:255-68.
139. Guillonneau B. Ceteris paribus and nomograms in medicine. *Eur Urol* 2007;52:1287-9.
140. Stephan C, Xu C, Cammann H, et al. Assay-specific artificial neural networks for five different PSA assays and populations with PSA 2-10 ng/ml in 4,480 men. *World J Urol* 2007;25:95-103.

141. Stephan C, Xu C, Finne P, et al. Comparison of two different artificial neural networks for prostate biopsy indication in two different patient populations. *Urology* 2007;70:596-601.
142. Canto EI, Singh H, Shariat SF, et al. Effects of systematic 12-core biopsy on the performance of percent free prostate specific antigen for prostate cancer detection. *J Urol* 2004;172:900-4.
143. Scaramuzzino DA, Schulte K, Mack BN, Soriano TF, Fritsche HA. Five-year stability study of free and total prostate-specific antigen concentrations in serum specimens collected and stored at -70 degrees C or less. *Int J Biol Markers* 2007;22:206-13.
144. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* 2007;115:928-35.
145. Cook NR. Statistical evaluation of prognostic versus diagnostic models: beyond the ROC curve. *Clin Chem* 2008;54:17-23.
146. Pencina MJ, D'Agostino RB, Sr., D'Agostino RB, Jr., Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 2008;27:157-72.
147. Utsumi T, Kawamura K, Suzuki H, et al. External validation and head-to-head comparison of Japanese and Western prostate biopsy nomograms using Japanese data sets. *Int J Urol* 2009;16:416-9.
148. Stephan C, Lein M, Jung K, Schnorr D, Loening SA. The influence of prostate volume on the ratio of free to total prostate specific antigen in serum of patients with prostate carcinoma and benign prostate hyperplasia. *Cancer* 1997;79:104-9.
149. Jung K, Stephan C, Lein M, et al. Analytical performance and clinical validity of two free prostate-specific antigen assays compared. *Clin Chem* 1996;42:1026-33.
150. Stephan C, Buker N, Cammann H, Meyer HA, Lein M, Jung K. Artificial neural network (ANN) velocity better identifies benign prostatic hyperplasia but not prostate cancer compared with PSA velocity. *BMC Urol* 2008;8:10.

Eidstattliche Erklärung

Ich, Martin Bender, erkläre an Eides statt, dass ich die vorgelegte Dissertationsschrift mit dem Titel **„Validierung eines artifiziellen neuronalen Netzwerkes zur Vorhersage von Prostatabiopsieergebnissen anhand prospektiver Daten“** in der Klinik für Urologie der Medizinischen Fakultät Charité – Universitätsmedizin Berlin mit Unterstützung durch Herrn PD Dr. med. Carsten Stephan und Herrn Dr.-Ing. Henning Cammann ohne sonstige Hilfe selbst verfasst habe.

Bei der Erstellung der Dissertation habe ich keine anderen als die aufgeführten Quellen und Hilfsmittel benutzt.

Die vorliegende Dissertation beruht auf den Ergebnissen eigenständiger wissenschaftlicher Arbeit und stellt auch in Teilen keine Kopie anderer Arbeiten dar.

Ich habe an keiner anderen Stelle ein Promotionsverfahren beantragt.

Berlin, den 7. Dezember 2011

Martin Bender

Lebenslauf

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht.

Publikationsliste

Teile dieser Dissertation sind in folgenden Publikationen erschienen:

Stephan C, Cammann H, Bender M, Miller K, Lein M, Jung K, Meyer HA. Internal validation of an artificial neural network for prostate biopsy outcome. Int J Urol 2009. Epub ahead of print.

Vorstellung der Studienergebnisse im Rahmen der 20th European Students' Conference in Berlin, 4. bis 7. Oktober 2009, unter dem Titel: „External validation of an artificial neural network for prostate cancer detection with population dependent variability“ (1. Platz beim Wettbewerb der Vorträge im Fachbereich Urologie).

Danksagung

Ich danke Herrn PD Dr. med. Carsten Stephan für die Überlassung des Themas und das damit in mich gesetzte Vertrauen, sowie für seine Unterstützung und sein außergewöhnliches Engagement bei der Betreuung meiner Arbeit.

Herrn Dr.-Ing. Henning Cammann möchte ich für die Beratung in mathematisch-statistischen Fragen danken.

Meinen Eltern danke ich für ihre Verlässlichkeit.

Mein besonderer Dank gilt Yina für ihr Verständnis und ihre Anwesenheit.

Berlin, den 7. Dezember 2011

Martin Bender