

Цели. Для демонстрации точности диагностики ПНК с помощью экспертной системы DANAOS было проведено многоцентровое исследование на представителях различных национальностей.

Методы. Для этого многоцентрового исследования, проведённого в 13 дерматологических центрах девяти европейских стран, была создана и использована система с калиброванной камерой, позволившая получить снимки ПНК. Этот комплект данных был использован для обучения экспертной системы ИНС в целях компьютерной диагностики меланомы. Мы проанализировали различные аспекты комплекта данных и его влияния на эффективность экспертной системы. Экспертная система ИНС прошла тренировку на комплекте данных из 2218 дерматоскопических снимков ПНК.

Результаты. Экспертная система показала в результате эффективность, сравнимую с эффективностью опытных дерматологов, как об этом сообщалось в литературе. Эффективность зависела от размера и качества базы данных, и от её состава.

Корреспонденция: Клаус Хоффманн.
E-mail: k.hoffmann@derma.de

© 2003 Британская Ассоциация дерматологов

801

802 К. Хоффманн и др.

Заключение. В данном исследовании были продемонстрированы необходимость большой базы данных, польза мультицентрового сбора данных, а также выгода репрезентативной подборки случаев из клинической практики. Снимки, которые было трудно классифицировать с помощью экспертной системы ИНС, отличались от тех снимков, которые были сложны для классификации клиницистами. В этой связи мы предполагаем, что сочетание клинициста и компьютера может потенциально улучшить точность диагностики ПНК. Это может привести к улучшению диагностики меланомы и уменьшению числа ненужных биопсий.

Ключевые слова: компьютерный анализ, DANAOS, дерматоскопия, меланома.

Количество случаев злокачественной меланомы и показатели смертности от неё возрастают уже много десятилетий подряд. Поскольку развитая меланома кожи всё ещё не поддаётся лечению, раннее обнаружение средствами точного мониторинга является важным шагом в сторону уменьшения смертности.¹ Исследования в ходе мониторинга обычно просты и требуют только краткого осмотра кожи опытным врачом. Правило ABCD является одним из наиболее широко используемых методов оценки пигментных новообразований кожи (ПНК)

невооружённым глазом. Однако в отношении трудных или пограничных случаев ПНК точность диагностики может быть лишь немного выше 60% даже в специализированных центрах.² Дерматоскопия – это неинвазивный метод, который позволяет клиницистам оценивать многие морфологические особенности с помощью диаскопии, использующей увеличительные линзы и иммерсионную жидкость. В нескольких исследованиях было показано, что этот метод улучшает точность диагностики на 20-30% по сравнению с простым клиническим осмотром.³⁻⁸ Во многих исследованиях использовались специальные дерматоскопические критерии, такие как рисунок, цвет, интенсивность пигментации, конфигурация, правильность форм, характер границ и поверхности.^{6,9,10} В помощь неопытным клиницистам были разработаны чётко определённые дерматоскопические методы, а также составлены подробные атласы. Это помогает проводить правильное обучение дерматоскопии, которая необходима для улучшения точности диагностики.^{7,8,11,12} Однако качественная оценка многочисленных морфологических особенностей ПНК, наблюдаемых в дерматоскопии, часто является очень сложной и субъективной. Чтобы преодолеть эти проблемы качественной интерпретации и избежать необходимости долгого обучения, недавно были разработаны методы, основанные на математическом анализе. Эти методы базируются на компьютерном анализе цифровых изображений, полученных при дерматоскопии. Компьютерный анализ ПНК потенциально пригоден как для предварительного скрининга пациентов, выполняемого неопытным персоналом (например, медсёстрами), так и для оказания помощи клиницистам в принятии решений по классификации ПНК. Более того, компьютерный анализ дерматоскопических изображений может служить как дополнительный инструмент для улучшения повторных осмотров, особенно у пациентов с многочисленными атипичными невусами.¹³⁻²⁴

В клинической практике диагноз обычно основан на мнемоническом правиле ABCD, которое представляет собой линейную комбинацию четырёх наиболее характерных особенностей (асимметрия, граница, цвет и диаметр).^{1,2,25} Осуществляются некоторые попытки преобразовать критерии ABCD, а также и другие критерии – например, текстуру – в автоматически вычисляемые параметры.^{24,26-30} Мы решили использовать комплект данных для обучения экспертной системы, основанной на искусственной нейросети (ИНС). Искусственные нейросети являются популярным методом обработки информации в компьютерной науке и системах искусственного интеллекта. Главной особенностью нейросетей является параллельная обработка информации большими группами сравнительно простых, но тесно взаимосвязанных процессоров, и способность самоорганизации или адаптации с помощью так называемых "алгоритмов обучения", которые управляют связью между отдельными процессорами.²⁰

Так как меланомы и некоторые формы доброкачественных новообразований кожи очень мало отличаются по своим физическим характеристикам, необходимо большое количество сочетаний различных особенностей для принятия надёжных диагностических решений. При осмотре новообразований кожи, дерматологи основывают свои решения на опыте, а также на сложных логических выводах и больших знаниях, которые невозможно сузить до четырёх параметров правила ABCD.³¹ В отличие от этого, нейросеть способна к обучению на основе опыта, получаемого из сложных связей, образующихся многократно повторяющимся исследованием примеров – то есть, таким же путём, как её биологические партнёры учатся проводить разнообразную дифференциальную диагностику. Для того, чтобы нейросеть могла хорошо "натренироваться", необходимо большое количество образцов, взятых из разных групп населения с различным типом кожи.

АО "Центр нейроинформатики" (Бохум, Германия) в тесном сотрудничестве с отделением дерматологии Университета Рура в Бохуме (Бохум, Германия) недавно разработал прототип компьютерной диагностики ПНК: диагностический и нейросетевой анализ рака кожи (DANAOS).^{23,24}

Этот прототип выполняет сложный анализ изображений, полученных методом цифровой дерматоскопии, выделяет характерные особенности и производит классификацию структуры. Чтобы продемонстрировать точность диагностики ПНК с помощью экспертной

© 2003 Британская Ассоциация дерматологов. Британский журнал дерматологии, **149**, 801-809

ДИАГНОСТИЧЕСКИЙ И НЕЙРОСЕТЕВОЙ АНАЛИЗ РАКА КОЖИ 803

системы ИНС, было проведено мультицентровое исследование различного многонационального населения в девяти европейских странах.

Материалы и методы

План и темы исследования

С июля 1997 г. по декабрь 2000 г. группа исследователей проекта DANAOS, включающая 14 членов из разных стран Европы (один из Дании, двое из Франции, четверо из Германии, один из Греции, один из Италии, один из Нидерландов, один из Испании, один из Швеции и двое из Великобритании) провели мультицентровое исследование. Пациенты осматривались опытными исследователями в ходе планового мониторинга на рак кожи в 13 дерматологических отделениях университетов и клиниках дерматологического профиля. Приём пациента состоял из клинического осмотра и цифровой дерматоскопии ПНК. Они затем классифицировались как доброкачественные, или как подозрительные на злокачественность. ПНК, которые были отмечены как подозрительные, были подвергнуты биопсии и изучению двумя независимыми гистопатологами. По этическим соображениям, безусловно доброкачественные новообразования биопсии не подвергались; вместо этого, они были ещё раз проверены независимым дерматологом. Их также включили в процесс тренировки экспертной системы, поскольку ставилась цель создать такую систему, которая была бы пригодна для анализа широкого спектра различных новообразований, а не только пограничных случаев.

Получение дерматоскопических снимков и выборка данных

Дерматоскопия проводилась во всех центрах одной и той же видеокамерой и одной и той же программой анализа изображений, которые были разработаны компанией АО "Визиомед" (Бохум, Германия). Новообразования были засняты при стандартизованных условиях с увеличением x10, x20 и x30 как с использованием иммерсионных жидкостей, так и без них, и сохранены на компьютере. Снимки с данными анамнеза и типа кожи по Фитцпатрику³² были анонимизированы и отосланы в Бохум, где их проверили на пригодность и затем присоединили к глобальной базе данных. Разработка и обучение экспертной системы ИНС проводились в АО "Визиомед". Так как было не совсем ясно, будут ли все снимки годиться для автоматизированного компьютерного анализа, был разработан инструмент для верификации оригиналов снимков и отбора соответствующих требованиям "чистых" снимков. Критерии отбора снимков были следующими: новообразование должно быть хорошо

видимым и резким, и на нём не должно быть никаких посторонних цветных артефактов. Снимки с низкой контрастностью изображения (вероятно, из-за того, что новообразования были очень светлыми) имели большое значение для тестирования допустимых пределов нашего алгоритма, поэтому такие снимки не отсортировывались.

Описание камеры

Дерматоскопия проводилась с помощью видеокамеры и системы обработки изображений, использующих компьютер Pentium II 200 МГц. Компьютер содержал устройство захвата изображений Matrox и магнито-оптический диск ёмкостью 1 Гб, который использовался для обмена данными. Видеокамера была вмонтирована в небольшой по размеру (длиной 16 см) и весу (750 г) ручной корпус и обладала способностью делать обзорные и микроскопические снимки с увеличением в 10, 20 и 30 раз. Находившаяся в этом ручном устройстве миниатюрная видеокамера имела 3 ССD-матрицы, разрешение 768 x 576 пикселей (TU-1K48P, Toshiba, Токио, Япония) и специально разработанную систему оптических линз³³ с источником света.

Оптическая система линз обладала тремя фиксированными значениями увеличения, при которых объект увеличивался в 10, 20 и 30 раз во время просмотра на 17-дюймовом мониторе. Эффективный размер пикселя на коже был равен 104 микрометра при увеличении в 30 раз. Система линз позволяла производить также обзорные снимки больших участков тела. Подсветка состояла из светодиодов (LED) белого спектра, находящихся в оптимизированных прозрачных корпусах из материала Perspex (ПММА, полиметилметакрилат), которые были закреплены в головке ручного устройства с

Таблица 1. Количество снимков, сделанных по каждому диагнозу

Диагноз	Гистологический	Окончательный
Диспластический невус	149	149
Актинический лентиго	6	20
Лентиго простое	10	39
Лентиго возрастное	0	1
"Кофейное пятно"	0	5
Себорейный кератоз	48	66
Доброкачественное образование, НК*	23	97
Невоидный лентиго/ интрадермальный невус	127	383
Смешанный невус	262	633
Дермальный невус	62	164
Рассыпанный невус	1	8
Голубой невус	21	26
Невус Сеттона	2	3
Врождённый невус	28	66
Ангиома	9	15
Невус Шпитца	17	29
Меланоцитарный невус, НК*	0	311
Злокачественный лентиго*	24	24
Поверхностно-распространяющаяся меланома	115	115
Акролентигинозная меланома	2	2
Злокачественная узелковая меланома	22	22

Неклассифицирующаяся меланома	24	24
Базально-клеточная карцинома (пигментная)	16	16
Всего:	968	2218

* а также злокачественная лентиго-меланома; НК - неклассифицированное

© 2003 Британская Ассоциация дерматологов. Британский журнал дерматологии, **149**, 801-809

804 К. Хоффманн и др.

определённым углом наклона. Все светодиоды контролировались независимо программой. Это позволяло использовать или прямой, или рассеянный свет, когда было необходимо.

По окончании первоначального исследования в рамках DANAOS камера была усовершенствована с учётом требовавшихся по результатам исследования улучшений и в настоящее время присутствует в продаже под маркой microDERM.

Экспертная нейросетевая система

При обучении экспертной нейросетевой системы было важно, чтобы образцы представляли самые различные диагнозы (Таблица 1). После получения снимка ПНК сначала определялась его внешняя граница. Этот процесс называется "определением границы" или "сегментацией". Мы использовали гибридный метод, который соединял статистическую кластеризацию цветного пространства и иерархический алгоритм, основанный на методе "роста области". Детали техники сегментации ранее публиковались.²⁶⁻²⁹ В тех случаях, когда волосы покрывали часть ПНК, для уменьшения влияния волос на сегментацию использовался специальный фильтр. Как только новообразование было успешно сегментировано, процесс определения особенностей мог быть сконцентрирован строго на образовании, его краях и прилегающих к ним участках. Определение *процента асимметрии* новообразования начиналось с проведения главных осей инерции новообразования, затем изображение "складывалось" вдоль этих осей, и подсчитывалась величина участков, выступающих за оси, а в конце этот результат делился на общую площадь новообразования.³⁰ *Нерегулярность границ* выражалась индексом нерегулярности³¹ и фрактальными особенностями границы.³⁰ Вариации цвета выражалась вариациями красных, зелёных и голубых компонентов цвета.²⁹ Текстура также является важной особенностью, которая была учтена.³⁰ Так как в локализации, ориентации и степени увеличения новообразований неизбежны вариации, данный фактор не должен оказывать влияние на выбранные особенности. Некоторые предложенные особенности не изменяются при трансформации³⁴ или вращении.³⁵ Мы использовали особенности, основанные на специализированных импульсах Габора, которые не зависят от масштаба, трансформации или вращения. Комплект данных, собранный во время

исследования, был разделён на 10 частей, которые были затем использованы для обучения и тестирования системы классификации ИНС. Девять из десяти частей были использованы для обучения и одна для тестирования. Для статистического анализа и построения диаграмм был использован статистический инструмент "R".^{36,37}

Результаты

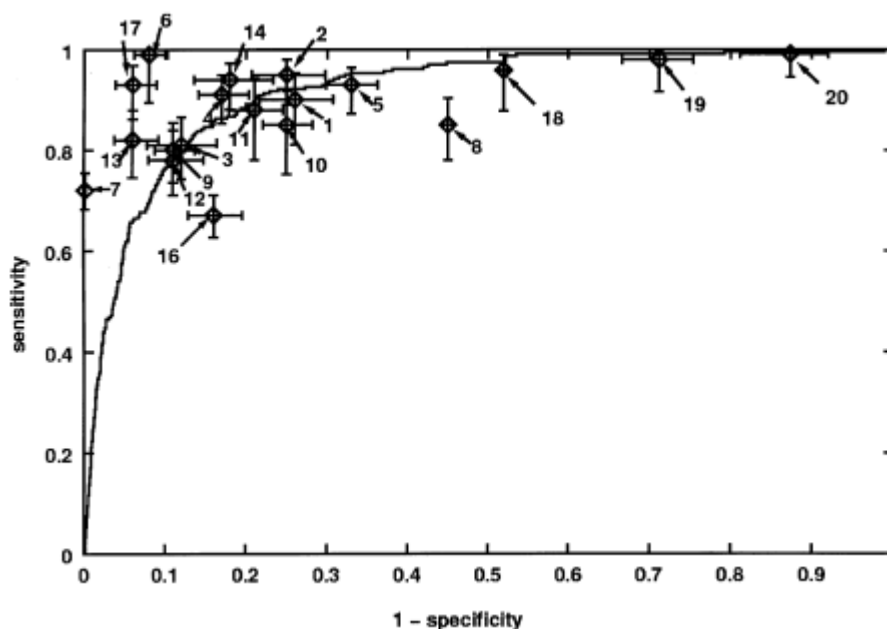
Тринадцать членов DANAOS сделали в общей сложности 4280 клинических и 3519 дерматоскопических снимков. Гистологические и окончательные диагнозы ПНК даны в Таблице 1. Только 2465 (70%) дерматоскопических снимков были достаточно хорошего

Таблица 2. Список опубликованных данных по чувствительности и специфичности данных для дифференцирования доброкачественных и злокачественных ПНК

Авторы	Метод			К-во случаев			ROC	
	ЦА	ЭЛМ	К	Д	З	В	Чувствит.	Специфич.
1. Биндер ²⁰	х			81	39	120	90	74
2. Ардженциано ⁴⁶		х		86	60	146	95	75
3. Андреасси ³	х			90	57	147	81	88
4. Шиндевольф ¹⁶	х			229	80	309	91	83
5. Мензис ³⁸	х			95	75	170	93	67
6. Зайденари ²¹	х			385	18	403	99	92
7. Ганстер ³⁹	х			5281	108	5389	72	99,8
8. Бенелли ⁴⁰			х	341	60	401	85	55
9.		х		341	60	401	80	89
10. Кристофолини ⁴¹			х	187	33	220	85	75
11.		х		187	33	220	88	79
12. Лоренцен ⁴²			х	183	49	232	78	89
13.		х		183	49	232	82	94
14. Соьер ⁴³			х	94	65	159	94	82
15.		х		94	65	159	94	82
16. Станганелли ⁴⁴			х	207	55	262	67	84
17.		х		210	58	268	96	97
18. Соьер ⁴⁵		х		73	35	108	95,93	48,1
19.		х		73	35	108	98,02	28,8
20.		х		73	35	108	99,17	12,6

ROC – характеристика эффективности получателя; ЦА – цифровой анализ; ЭЛМ – эпифлуоресцентная микроскопия (эксперты); К – клиническое исследование врачами; Д – доброкачественное; З – злокачественное; В – всего.

Рис. 1. Кривая характеристики эффективности получателя при дифференцировании доброкачественных и злокачественных новообразований с использованием нашей нейросети в сравнении с опубликованными данными по чувствительности и специфичности, полученными клиницистами и другими компьютерными системами классификации. Цифры соответствуют номерам строк в Таблице 2.



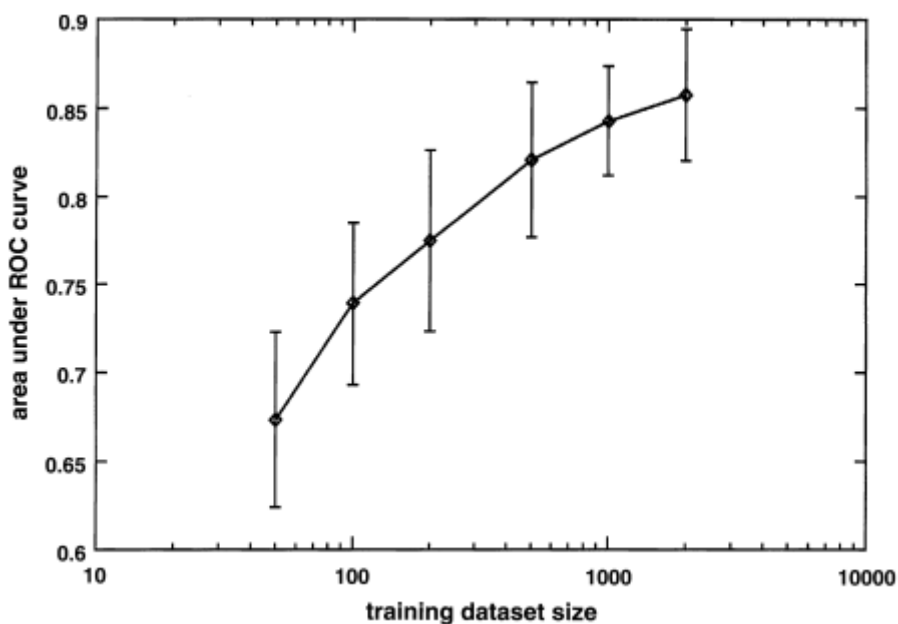
качества. Сегментация была сделана успешно в 2218 (90%) из этих снимков, и поэтому анализ в целом стал возможен.

Эти 2218 ПНК были получены от 1173 пациентов (694 женщин и 479 мужчин) со средним возрастом (\pm SD) $36,6 \pm 16,4$ года. Сорок четыре пациента имели фототип кожи I, 418 фототип II, 430 фототип III, 114 фототип IV, 17 фототип V и один – фототип VI; у 139 пациентов тип кожи определить было невозможно.

Эффективность нейросети

В порядке сравнения эффективности ИНС и других систем классификации, мы извлекли данные о чувствительности и специфичности, относящиеся к клинической и компьютерной диагностике, из литературы (Таблица 2)^{3,16,20,21,38-46} и нанесли их на тот же график (Рис. 1), на котором изображена кривая ROC.³⁶ Кривая ROC показывает все возможные сочетания чувствительности и специфичности в нашей нейросети. Мы взяли только статьи, в которых описывалось более 30 злокачественных новообразований, а общее число всех ПНК было как минимум 100. Линии ошибок чувствительности и специфичности были вычислены на основе числа случаев, описанных в публикациях, а также с использованием предположительной бинормальной кривой ROC.³⁶ Линии 18-20 взяты из исследования⁴⁵ по эффективности 40 дерматологов, которые проводили оценку 108 ПНК. Эти показатели чувствительности и специфичности были получены с использованием

Рис 2. Пространство под кривой характеристики эффективности получателя (ROC) как функция объёма обучающих данных



© 2003 Британская Ассоциация дерматологов. Британский журнал дерматологии, 149, 801-809

806 К. Хоффманн и др.

в качестве диагноза решения, предложенного системой управления. Например, значение в строке 18 может быть получено, если только образования, назначенные к биопсии, рассматриваются как злокачественные, а все остальные образования считаются доброкачественными. Строка 19 может быть получена, если образования, назначенные к биопсии или цифровому мониторингу, рассматриваются как злокачественные, а все остальные считаются доброкачественными. Полученная в результате экспертная система ИНС показала эффективность, равную эффективности клиницистов.

Влияние размера базы данных на эффективность нейросети

Для определения размера обучающей базы данных, необходимой для "тренировки" хорошей ИНС, мощностью проводить дифференцирование доброкачественных и злокачественных ПНК, фактически было создано несколько нейросетей, которые прошли тренировку на обучающей информации разного размера – от 50 до 2000 случаев. Поле под кривой ROC³⁶ на Рис 2. является функцией количества обучающих образцов. Эффективность экспертной системы обычно улучшается при росте размера обучающего комплекта данных.

Влияние метода сбора данных: единственный центр против множества центров

Для анализа надёжности классификации в отношении различного населения были рассмотрены три сценария. (1) Обучение и тестирование в единственном центре: для тренировки классифицирующей системы нейросети были выбраны данные из единственного центра. Эта система была затем протестирована на данных из того же центра (за исключением 10%). (2) Более широкая задача для обучения в единственном центре: для тренировки классифицирующей системы нейросети были выбраны данные из единственного центра. Эта система была затем протестирована на данных из всех остальных центров. (3) Мультицентровое обучение и тестирование: классифицирующая система ИНС была обучена на случайно выбранных образцах, полученных из всех центров, а затем была протестирована на всех остальных данных (за исключением 10%). Размер обучающего комплекта данных был одинаков во всех трёх экспериментах.

Эффективность системы в отношении населения, относящегося к одному месту, была немного выше, чем эффективность по отношению к смешанному населению. Результаты были наихудшими в случае (2), где обучение проводилось на населении из единственного места, но система классификации была затем использована для населения из других мест. Результирующие области под кривой ROC³⁶ даны в Таблице 3.

Влияние обучения на "чистых" данных против обучения на всём комплекте данных

Важным вопросом во время обучения системы классификации были критерии выбора обучающих данных. Одним вариантом было использование данных только с высоким

Таблица 3. Эффективность обучения и тестирования в единственном центре против множества центров

Сценарий	Комплект данных	Комплект для теста	Область ROC	± SD
(1)	Место 1	Место 1	0,847	0,046
(2)	Место 1	Все кроме (1)	0,807	0,014
(3)	Все места	Все места	0,826	0,021

ROC – характеристика эффективности получателя

Таблица 4. Влияние сортировки обучающих данных на эффективность классификации

Сценарий	Комплект данных	Комплект для теста	Область ROC
(1)	Чистые	Чистые	0,848
(2)	Чистые	С включениями	0,803
(3)	Все	Все	0,844

ROC – характеристика эффективности получателя

качеством изображения и выбраковка снимков с волосами или воздушными пузырьками при тренировке экспертной системы. Другим вариантом было использование всех снимков, полученных в клинической практике, невзирая на присутствующие в них артефакты. Мы оценили влияние выбора данных, выполнив три различных сценария тренировок и тестирования. Снимки были разделены на две группы – "чистые" снимки и снимки, содержащие волосы или пузырьки воздуха. Эти три сценария прошли испытания: (1) обучение на чистых снимках, тестирование на чистых снимках; (2) обучение на чистых снимках, тестирование на снимках с посторонними включениями; (3) обучение на всех снимках и тестирование на всех снимках.

Сценарии (1) и (3) показали очень похожие результаты, а в сценарии (2), где система была натренирована на чистых снимках и протестирована на снимках с включениями, эффективность была ниже. Результирующие области под кривой ROC³⁶ даны в Таблице 4.

Дискуссия

Что является наиболее важным в исследовании DANAOS – это его объем и широта: используя доступное стандартизированное съёмочное оборудование, было получено большое количество дерматоскопических и клинических снимков ПНК из 13 дерматологических центров, расположенных в девяти европейских странах. Собранный база данных включала широкий спектр ПНК в репрезентативной выборке (Таблица 1). Категоризация тех доброкачественных меланоцитарных невусов, для которых биопсия не была показана, может быть произвольной, но это не должно влиять на результаты, так как мы старались не выделять эти случаи в нашем анализе.

Этот комплект данных может использоваться как основа для обучения в компьютерных системах классификации ПНК. Эффективность ИНС, показанная на Рис. 1, подобна средней эффективности, о которой сообщалось в литературе.

© 2003 Британская Ассоциация дерматологов. Британский журнал дерматологии, 149, 801-809

ДИАГНОСТИЧЕСКИЙ И НЕЙРОСЕТЕВОЙ АНАЛИЗ РАКА КОЖИ 807

Суммировав большую коллекцию опубликованных данных по чувствительности и специфичности как разновидность мета-исследования, мы пришли к заключению, что эффективность нашей экспертной системы можно сравнить с эффективностью клиницистов. Однако это сравнение не следует чересчур переоценивать, так как базы данных различны, а результаты сильно зависят от сложности случаев. Для достижения сравнимых результатов, комплект данных должен быть репрезентативной выборкой случаев, типичных для каждой конкретной клинической ситуации – например, такой, когда люди хотят пройти мониторинг на рак кожи. Размер комплекта данных также имеет значение. Для улучшения обобщающих

способностей экспертной системы, данные должны собираться с различных мест, находящихся в разных странах.⁴⁷

Анализ влияния размера комплекта данных, использованного для тренировки системы, показал, что эффективность увеличивается с ростом размера комплекта. Для небольшого числа случаев это увеличение более важно, чем для большого комплекта данных, но даже при наличии 2000 случаев кривая обучения не достигла насыщения. Этот результат, кажется, подтверждает то, что уже хорошо известно при обучении клиницистов, - это то, что они должны строить свой опыт на значительном количестве случаев. Анализ влияния комплекта данных, собранных из множества мест, показал, что экспертная система, обученная и протестированная на комплекте данных из единственного места, демонстрирует неплохую эффективность классификации, но когда она тестируется на данных, полученных из других мест, её эффективность оказывается значительно более низкой. При обучении экспертной системы на комплекте данных из множества центров, эффективность почти также высока, как в случае обучения и тестирования на образцах из единственного места. Тем не менее, результаты являются актуальными для большого разнообразия людей, включённых в данное исследование.

Так как все эксперименты проводились на обучающих комплектах равного размера, можно сделать заключение, что выгода мультицентрового исследования состоит не только в возможности собрать большее число случаев в сжатые сроки. Анализ влияния предварительного отбора данных для обучения показал такой же результат. При обучении на предварительно отобранных "чистых" снимках, когда исключались снимки с плохой резкостью или с такими артефактами, как волосы и воздушные пузырьки, результаты тестирования на других чистых снимках были очень хорошими, но эффективность снижалась очень сильно, когда система классификации использовалась для анализа снимков из клинической практики, в которых не всегда можно было обеспечить отсутствие артефактов. Когда для обучения используется репрезентативный комплект снимков, включающий как чистые снимки, так и снимки с артефактами, эффективность экспертной системы примерно одинакова, и система мало восприимчива к наличию артефактов.

Суммируя эти наблюдения, мы предлагаем проводить обучение экспертной системы на репрезентативном комплекте данных, собранных из множества центров без предварительной сортировки, исключающей снимки с артефактами. Комплект данных должен быть как можно более объёмным. Рассматривая эффективность экспертной системы более детально, мы обнаружили, что некоторые снимки были классифицированы экспертной системой корректно, но те же снимки были трудны для визуальной классификации, а некоторые снимки классифицировались экспертной системой неверно, однако легко классифицировались специалистами. Естественно, что диагностической точности в 100% достигнуть практически невозможно. Это касается также и очень опытных дерматологов.² Во многих случаях было бы полезно получить быстрое "второе мнение" с помощью экспертной системы – например, в клинической обстановке, где не всегда легко получить быстрое "второе мнение" от другого клинициста. Альтернативным подходом является получение быстрого "второго мнения" с помощью теледерматоскопии, которая недавно прошла оценку в одном мультицентровом исследовании.⁴⁸ Описанная выше экспертная система, которая была обучена на обширной базе данных, не может заменить собой простое клиническое обследование, но может использоваться как "второе мнение" для поддержки врачебной диагностики. Основываясь на наших данных и на опыте предыдущих исследований, мы продемонстрировали потенциал таких экспертных систем.^{47,49-53}

В заключение скажем, что прототип компьютерной диагностики ПНК с использованием ИНС был разработан на основе европейского мультицентрового исследования (DANAOS).⁵² Опыт данного исследования указывает на необходимость наличия широкой базы данных, полезность сбора данных из множества мест и выгоду репрезентативной выборки случаев из клинической практики, а также говорит против выбора только самых лучших снимков. Полученная экспертная система показала эффективность, сравнимую с эффективностью клиницистов, хотя диагностические трудности, которые испытывают эта система и клиницисты, отличаются друг от друга.^{45,47,51}

Таким образом, мы предполагаем, что использование комбинации "клиницист – экспертная система" обладает потенциалом улучшения диагностики ПНК. Результатом такой комбинации экспертов может стать улучшение ранней диагностики меланомы и уменьшение случаев неоправданных биопсий. Мы планируем проведение дальнейших исследований для математической оценки этих выгод.

Признательность

Проект DANAOS был частично поддержан грантом IN 10483 I DANAOS Европейского Союза. Мы благодарим всех коллег, которые принимали участие в исследовании DANAOS.

Литература

1. П. Альтмайер, К. Хоффманн, М. Штюкер. Рак кожи и УФ-радиация. Нью-Йорк: Изд-во Шпрингер, 1997 г.
2. К. Э. Мортон, Р. М. МакКи. Клиническая точность диагностики злокачественной меланомы кожи. Британский журнал дерматологии, 1998 г. 138: 283-287.

© 2003 Британская Ассоциация дерматологов. Британский журнал дерматологии, **149**, 801-809

808 К. Хоффманн и др.

3. Л. Андреасси, Р. Перотти, П. Рубеньи и др. Цифровой дерматоскопический анализ для дифференцирования атипичных невусов и ранней меланомы. Журнал "Архивы дерматологии", 1999 г. **135**: 1459-1465.
4. М.Л. Банофута, Э. Ботчер, П. Эгертер, П. Сайяк. Является ли дерматоскопия (эпилюминесцентная микроскопия) полезной при диагностике меланомы? Результаты мета-анализа с использованием техники, приспособленной для оценки диагностических тестов. Журнал "Архивы дерматологии", 2002 г. **137**: 1343-1350.
5. П. Карли, В. ДеДжиорджи, Х.П. Соьер и др. Дерматоскопия в диагностике пигментных новообразований кожи: новая семиология для дерматолога. Журнал Европейской Академии дерматологии и венерологии, 2002 г. **14**: 353-369.

6. Х.П. Соьер, Дж. Ардженциано, В. Руокко, С. Кименти. Дерматоскопия в диагностике пигментных новообразований кожи (часть II).
Европейский журнал дерматологии, 2002 г. **137**: 483-498.
7. Дж. Ардженциано, Х.П. Соьер. Дерматоскопия пигментных новообразований – ценный инструмент для ранней диагностики меланомы.
Журнал "Ланцет – Онкология", 2001 г. **2**: 443-449.
8. Х. Киттлер, Х. Пехамбергер, К. Вольф, М. Биндер. Диагностическая точность дерматоскопии.
Журнал "Ланцет – Онкология" 2002 г. **3**: 159-165.
9. Р.М. МакКи. В помощь предоперативной оценке пигментных новообразований кожи.
Британский журнал дерматологии, 1971 г. **85**: 232-238.
10. А. Штайнер, М. Биндер, М. Шемпер и др. Статистическая оценка критериев эпилюминесценции для меланоцитарных пигментных новообразований кожи.
Журнал Американской Академии дерматологии, 1993 г. **29**: 581-588.
11. В. Штольц, О. Браун-Фалько, П. Билек и др. Цветной атлас дерматоскопии.
Оксфорд: Изд-во "Блэкуэлл Сайенс", 1994 г.
12. С.В. Мензис. Автоматизированная эпилюминесцентная микроскопия: человек против машины в диагностике меланомы.
Журнал "Архивы дерматологии", 1999 г. **135**: 1538-1540.
13. Р.О. Кенет, С. Канг, Б.Дж. Кенет и др. Клинический диагноз пигментного новообразования с использованием цифровой эпилюминесцентной дерматоскопии.
Журнал "Архивы дерматологии", 1993 г. **129**: 157-174.
14. Э. Грин, Н. Мартин, Й. Пфицнер и др. Компьютерный анализ снимков меланомы в диагностике меланомы.
Журнал Американской Академии дерматологии, 1994 г. **31**: 958-964.
15. Р. Перотти, М. Бьяджиоли, М. Буррони и др. Анализ снимков пигментных новообразований: клиничко-гистопатологическая корреляция.
Американский журнал дерматопатологии, 1994 г. **16**: 103-104.
16. Т. Шиндевольф, Р. Шиффнер, В. Штольц и др. Оценка различных техник получения снимков для визуальных компьютерных систем в диагностике злокачественной меланомы.
Журнал Американской Академии дерматологии, 1994 г. **31**: 33-41.
17. Т. Шиндевольф. Процесс классификации меланоцитарных опухолей с помощью цифрового анализа цветных снимков.
Аахен: Изд-во "Шакер", 1997 г.
18. Э. Собер, Дж. Бурстайн. Компьютерный анализ цифровых изображений: помощь в диагностике злокачественной меланомы.
Журнал дерматологии, 1994 г. **21**: 163-170.
19. П. Холл, Э. Клэридж. Компьютерный мониторинг для раннего обнаружения меланомы: есть ли будущее?
Британский журнал дерматологии, 1995 г. **132**: 325-338.
20. М. Биндер, Х. Киттлер, А. Зибер и др. Классификация пигментных новообразований кожи на основе эпилюминесцентной микроскопии с использованием компьютерного анализа снимков и искусственной нейросети.
Журнал "Исследование меланомы", 1998 г. **8**: 261-266.
21. С. Сейденари, Дж. Пелликани, А. Джанетти. Цифровая видеодерматоскопия и анализ снимков с автоматической классификацией для обнаружения тонких меланом.
Журнал "Исследование меланомы", 1999 г. **9**: 163-171.
22. Р.О. Дебей, К. Декестекер, Дж.Л. Пастелс и др. Компьютерный анализ снимков эпилюминесцентной микроскопии пигментных новообразований кожи.
Журнал "Цитометрия", 1999 г., **37**: 255-266.

23. Р. Хуземанн, С. Тёльг, В. Фон Зеелен и др. Компьютерная диагностика рака кожи с использованием нейросетей.
В: "Рак кожи и УФ-радиация" (П. Альтмайер, К. Хоффманн, М. Штюкер и др.). Нью-Йорк: Изд-во "Шпрингер", 1997 г. 1052-1063.
24. К. Хоффманн, К. Каспар, М. Хаппе и др. Новые техники получения изображений как рутинная? Взгляд в будущее: DANAOS. Дерматология.
В: "Основные направления и обеспечение качества в диагностике и терапии" (К. Гарбе, Г. Расснер и др.). Берлин: Изд-во "Шпрингер", 1999 г., 20-26.
25. Ф. Нахбар, В. Шотльц, Т. Меркле и др. Правило ABCD в дерматоскопии. Потенциально высокая ценность в диагностике сомнительных меланоцитарных новообразований кожи.
Журнал Американской Академии дерматологии, 1994 г. **30**: 551-559.
26. С.Е. Умбаух, Р.Х. Мосс, В.В. Штекер. Алгоритм автоматической цветовой сегментации для применения в идентификации границ опухолей кожи.
Журнал "Компьютерная обработка медицинских снимков и графики", 1992 г. **16**: 227-235.
27. Ф. Эрвал, М. Моньяти, В.В. Штекер, Р.Х. Мосс. Установление границ опухолей кожи на цветных изображениях.
IEEE Журнал "Операции в обработке медицинских изображений", 1993 г. **12**: 624-628.
28. Л. Цзу, М. Яковски, А. Гоштасби и др. Сегментация изображений рака кожи.
Журнал "Компьютерная обработка изображений", 1999 г. **17**: 65-74.
29. П. Шмид. Сегментация цифровых дерматоскопических изображений с помощью цветной кластеризации.
IEEE Журнал "Операции в обработке медицинских изображений", 1999 г. **18**: 164-171.
30. В.В. Штекер, Ч.С. Чанг, Р.Х. Мосс. Текстура на снимках кожи: сравнение трёх методов определения гладкости.
Журнал "Компьютерная обработка изображений", 1992 г. **16**: 179-90.
31. Ф. Эрвал, А. Чавла, В.В. Штекер и др. Нейросетевая диагностика злокачественной меланомы по цветным изображениям.
IEEE Журнал "Операции в биоинженерных разработках", 1994 г. **41**: 837-845.
32. Т.В. Фитцпатрик. Солнце и кожа.
Журнал медицинской эстетики, 1975 г. **2**: 33-34.
33. ГмбХ "Центр нейроинформатики". Патент DE № 197 533 92.3 (12.12.1997 г.). Настройка оптической системы для отображения поверхностей, особенно в целях обработки медицинских изображений.
34. Дж.Ф. Лью, Дж.Ч.М. Ли. Действенный и эффективный подход к классификации текстуры с использованием новых знаний в теории волновых импульсов.
Том "Б" документов Международной конференции по классификации структур, 1996 г. 820-824.
35. Р. Портер, Н. Канагарая. Надёжная классификация текстур, независимых от вращения. Схемы, основанные на волновых импульсах, фильтре Габора и случайных полях GMRF.
IEEE Журнал "Процедуры обработки изображений и сигналов", 1997 г. **144**: 188.
36. К.Е. Метц. Статистический анализ данных ROC в оценке диагностической эффективности.
В: Множественный регрессивный анализ: применение в медицинских науках (Д.Е. Херберт, Р.Х. Майерс и др.). Нью-Йорк: Американский Институт физики, 1986 г. 365-384.
37. Д. Нолан, Т. Спид. Статистическая лаборатория: Математическая статистика через приложения.
Берлин: Изд-во "Шпрингер", 2000.
38. С.В. Мензис, Л.М. Бишоф, Г. Педен и др. Автоматизированные инструменты для диагностики инвазивной меланомы: анализ снимков, сделанных с помощью иммерсионной микроскопии.
В: "Рак кожи и УФ-радиация" (П. Альтмайер, К. Хоффманн, М. Штюкер и др.). Нью-Йорк: Изд-во "Шпрингер", 1997 г. 1064-1070.
39. Х. Ганстер, А. Пинц, Р. Рёерер и др. Автоматизированное обнаружение меланомы.
IEEE Журнал "Операции в обработке медицинских изображений", 2001 г. **3**: 233-239.

40. К. Бенелли, Е. Росетти, В. Даль Поццо. Воспроизводимость метода дерматоскопии (7FFM) при диагностике злокачественной меланомы.
Европейский журнал дерматологии, 2000 г. **10**: 110-114.
41. М. Кристофолини, Дж. Цумияни, П. Бауэр и др. Дерматоскопия: полезность для дифференцированной диагностики пигментных кожных новообразований.
Журнал "Исследование меланомы", 1994 г. **4**: 391-394.
42. Х. Лоренцен, К. Вайсманн, К.С. Петерсен и др. Клиническая и дерматоскопическая диагностика злокачественной меланомы: оценка группой экспертов и группой не имеющих опыта.
Журнал "Акта дермато-венерология" (Стокгольм), 1999 г. **79**: 301-304.
43. Х.П. Соьер, Й. Смолле, Г. Ляйтингер и др. Диагностическая надёжность дерматоскопических критериев для обнаружения злокачественной меланомы.
Журнал "Дерматология", 1995 г. **190**: 25-30.

© 2003 Британская Ассоциация дерматологов. Британский журнал дерматологии, **149**, 801-809

ДИАГНОСТИЧЕСКИЙ И НЕЙРОСЕТЕВОЙ АНАЛИЗ РАКА КОЖИ 809

44. И. Станджанелли, М. Серафини, Л.А. Букки. Оценка точности цифровой эпилюминесцентной микроскопии, связанная с клиническим исследованием пигментных новообразований кожи, с помощью регистра случаев рака кожи.
Журнал "Дерматология", 2000 г. **200**: 11-16.
45. Х.П. Соьер, Дж. Ардженциано, С. Кименти и др. Дерматоскопия пигментных новообразований кожи.
В: Атлас, основанный на Консенсусной сетевой конференции по дерматоскопии в 2000 г.
Милан: Изд-во "Эдра медикал пабблишинг и Нью Медиа", 2001 г.
46. Дж. Ардженциано, Дж. Фаброчини, П. Карли и др. Эпилюминесцентная дерматоскопия в диагностике сомнительных меланоцитарных новообразований кожи: сравнение Правила ABCD для дерматоскопии с новым Списком из 7 пунктов, основанным на анализе структур.
Журнал "Архивы дерматологии", 1998 г. **134**: 1563-1570.
47. Г. Шварцер, В. Вах, М. Шумахер. Касательно неправильного использования искусственных нейросетей для прогнозирования и диагностической классификации в онкологии.
Журнал "Статистика в медицине", 2000 г. **19**: 541-561.
48. Д. Пикколо, Й. Смолле, Дж. Ардженциано и др. Теледерматоскопия – результаты мультицентрового исследования 43 пигментных новообразований кожи.
Журнал "Телемедицина и удалённое лечение", 2000 г. **6**: 132-137.
49. В. Штольц, Р. Шиффнер, Л. Пиллет и др. Улучшение мониторинга меланоцитарных новообразований кожи с использованием компьютерного устройства получения и наблюдения, оборудованного видеокамерой для микроскопии поверхности кожи.
Журнал Американской Академии дерматологии, 1996 г. **35**: 202-207.
50. Г. Потт, К. Шульц, М. Андрес и др. Системы анализа изображений и данных для автоматизированного обнаружения рака кожи.
В: Документы 1. Конференция Академии дерматологической науки и повышения квалификации земли Северный Рейн-Вестфалия, 2-4 октября 1998 г., Бонн.

51. Х. Хандельс, Т. Росс, Й. Кройш и др. Выбор особенностей для оптимизированного обнаружения рака кожи с использованием генетического алгоритма.
Журнал "Искусственный разум в медицине", 1999 г. **16**: 283-297.
52. М. Кройц, М. Аншютц, С. Гелен и др. Автоматическая диагностика рака кожи, использующая цифровую обработку изображений и участие экспертов.
В: "Обработка изображений для медицины – алгоритмы, системы, программы".
"Актуальная информатика" (Х. Хандельс, А. Хорш, Т. Леманн, Х.П. Майнцер и др.). Берлин: Изд-во "Шпрингер", 2001 г. 357-361.
53. А. Мурали, В.жВ. Штекер, Р.Х. Мосс. Обнаружение плотного пигмента в дерматоскопических изображениях, используя анализ текстуры.
Журнал "Исследование кожи и технология", 2000 г. **6**: 193-198.

Копия оригинала любезно предоставлена компанией
Visiomed AG, Билефельд, Германия.

Перевод на русский язык выполнен в ООО "МАКСмедикал"

Разрешается использование данного документа только в целях личного ознакомления.

Разрешается устное цитирование документа со ссылкой на его авторов.

Не разрешается копирование и любые другие способы воспроизведения
с целью дальнейшего распространения.

©2018 All rights reserved / Все права защищены

Epidemiology and Health Services Research

Diagnostic and neural analysis of skin cancer (DANAOS). A multicentre study for collection and computer-aided analysis of data from pigmented skin lesions using digital dermoscopy

K. HOFFMANN,* T. GAMBICHLER,* A. RICK,† M. KREUTZ,† M. ANSCHUETZ,†
T. GRÜNENDICK,† A. ORLIKOV,*† S. GEHLEN,† R. PEROTTI,‡ L. ANDREASSI,‡
J. NEWTON BISHOP,§ J.-P. CÉSARINI,¶ T. FISCHER,** P. J. FROSCHE,††
R. LINDSKOV,‡‡ R. MACKIE,§§ D. NASHAN,¶¶ A. SOMMER,*** M. NEUMANN,***
J. P. ORTONNE,††† P. BAHADORAN,††† P. F. PENAS,‡‡‡ U. ZORAS§§§
AND P. ALTMAYER*

*Department of Dermatology, Ruhr-University Bochum, Gudrunstraße 56, D-44791 Bochum, Germany

†Center of Neuroinformatics AG, Bochum, Germany

‡Department of Dermatology, University of Siena, Siena, Italy

§Genetic Epidemiology Division of the Cancer Research UK Clinical Centre at St James's Hospital, Leeds, U.K.

¶INSERM-LRTPH and Department of Dermatology, Fondation A.de Rothschild, Paris, France

**Department of Dermatology, Karolinska Sjukhuset, Stockholm, Sweden

††Department of Dermatology, Städtische Kliniken, Dortmund, Germany

‡‡Institute of Preventive Medicine, University of Copenhagen, Copenhagen, Denmark

§§Department of Dermatology, Glasgow University, Glasgow, U.K.

¶¶Department of Dermatology, University of Münster, Münster, Germany

***Department of Dermatology, University Hospital of Maastricht, Maastricht, the Netherlands

†††Department of Dermatology, University of Nice, Nice, France

‡‡‡Department of Dermatology, Hospital Universitario de la Princesa, Madrid, Spain

§§§Medical School, University of Crete, Crete, Greece

Accepted for publication 3 May 2003

Summary

Background Early detection of melanomas by means of diverse screening campaigns is an important step towards a reduction in mortality. Computer-aided analysis of digital images obtained by dermoscopy has been reported to be an accurate, practical and time-saving tool for the evaluation of pigmented skin lesions (PSLs). A prototype for the computer-aided diagnosis of PSLs using artificial neural networks (NNs) has recently been developed: diagnostic and neural analysis of skin cancer (DANAOS).

Objectives To demonstrate the accuracy of PSL diagnosis by the DANAOS expert system, a multicentre study on a diverse multinational population was conducted.

Methods A calibrated camera system was developed and used to collect images of PSLs in a multicentre study in 13 dermatology centres in nine European countries. The dataset was used to train an NN expert system for the computer-aided diagnosis of melanoma. We analysed different aspects of the data collection and its influence on the performance of the expert system. The NN expert system was trained with a dataset of 2218 dermoscopic images of PSLs.

Results The resulting expert system showed a performance similar to that of dermatologists as published in the literature. The performance depended on the size and quality of the database and its selection.

Conclusions The need for a large database, the usefulness of multicentre data collection, as well as the benefit of a representative collection of cases from clinical practice, were demonstrated in this trial. Images that were difficult to classify using the NN expert system were not identical to those found difficult to classify by clinicians. We suggest therefore that the combination of clinician and computer may potentially increase the accuracy of PSL diagnosis. This may result in improved detection of melanoma and a reduction in unnecessary excisions.

Key words: computer-aided analysis, DANAOS, dermoscopy, melanoma

The incidence and mortality rates of malignant melanoma have risen for many decades. Because advanced cutaneous melanoma is still incurable, early detection by means of accurate screening is an important step towards a reduction in mortality.¹ The screening tests are usually simple and require only a brief skin examination by an experienced investigator. The ABCD rule is one of the most widely used methods for evaluating pigmented skin lesions (PSLs) with the naked eye. For the many difficult or borderline PSLs, however, diagnostic accuracy is only slightly above 60%, even in specialized centres.² Dermoscopy is a noninvasive method that enables clinicians to evaluate numerous morphological features by means of magnified oil immersion diascopy. Several studies have shown that this method improves diagnostic accuracy by 20–30% compared with simple clinical observation.^{3–8} Many studies have used specific dermoscopic criteria such as pattern, colour, intensity of pigmentation, configuration, regularity, and margin and surface characteristics.^{6,9,10} Well-defined dermoscopic methods have also been developed for use by inexperienced clinicians, and detailed atlases are available. This allows formal training in dermoscopy, which is necessary to improve diagnostic accuracy.^{7,8,11,12} However, qualitative evaluation of the numerous morphological features of PSLs observable by dermoscopy is often extremely complex and subjective. To obviate these problems of qualitative interpretation and to avoid the need for time-consuming training, methods based on mathematical analysis have recently been developed. These methods are based on computer-aided analysis of digital images obtained by dermoscopy. Computer-based analysis of PSLs is potentially suitable both for prescreening of patients performed by non-experienced operators (e.g. nurses) and for aiding clinicians' decision-making in the evaluation of PSLs. Furthermore, computer-aided analysis of dermoscopic images may serve as an additional tool to improve follow-up, particularly in patients with multiple atypical naevi.^{13–24}

In clinical practice, the diagnosis is usually based on the mnemonic ABCD rule which is a linear combination of the four characterizing attributes (asymmetry, border, colour, diameter).^{1,2,25} Several approaches aim at the transfer of the ABCD criteria, as well as other criteria such as texture, into automatically computed quantities.^{24,26–30} We decided to use a dataset to train an expert system based on artificial neural networks (NNs). Artificial NNs are a popular approach to information processing in computer science and artificial intelligence. The main feature of NNs is parallel processing in a large group of relatively simple but highly interconnected processors and the ability to self-organize or adapt through so-called learning algorithms that change the connectivity between the units.²⁰

As melanomas and certain forms of benign skin tumours differ slightly in their physical characteristics, a large combination of features is required for reliable diagnostic decisions. When examining skin lesions, dermatologists base their decisions on experience, as well as on complex inferences and extensive knowledge which cannot be condensed into the four parameters of the ABCD rule.³¹ In contrast, NNs are capable of experience-based learning of complex relationships by repeated exposure to examples in much the same way as their biological counterparts learn to make their differential diagnoses. In order to train an NN well, however, a large sample size from a diverse population with different skin types is required.

The Center of Neuroinformatics AG (Bochum, Germany), in close cooperation with the Department of Dermatology of the Ruhr-University Bochum (Bochum, Germany), has recently developed a prototype for the computer-aided diagnosis of PSLs: diagnostic and neural analysis of skin cancer (DANAOS).^{23,24}

Based on digital dermoscopy, this prototype performs sophisticated image processing, feature extraction and pattern classification. To demonstrate the accuracy of PSL diagnosis by an NN expert system, a multicentre

study on a diverse multinational population was conducted in nine European countries.

Materials and methods

Study design and subjects

From July 1997 to December 2000, the DANAOS study group, including 14 members from various European countries (one in Denmark, two in France, four in Germany, one in Greece, one in Italy, one in the Netherlands, one in Spain, one in Sweden and two in the U.K.), performed a multicentre trial. As part of routine skin cancer screening, patients were examined by experienced investigators in 13 university dermatology departments and dermatology referral clinics. The examination included clinical inspection and digital dermoscopy of PSLs. These were classified as either benign or as suspicious of being malignant. PSLs considered suspicious by the investigators were excised and subjected to examination by two independent histopathologists. For ethical reasons, clearly benign lesions were not excised; instead they were double checked by an independent dermatologist. They were also included in the training of the expert system as the goal was to develop a system suitable for the whole range of lesions, not only the borderline cases.

Dermoscopic image acquisition and data selection

In all centres, dermoscopy was performed using exactly the same video camera and image analysis software developed by Visiomed AG (Bochum, Germany). The lesions were recorded and stored in a computer archive under standardized conditions at magnifications of $\times 10$, $\times 20$ or $\times 30$, both with and without immersion. The images, with anamnestic data and skin type according to Fitzpatrick,³² were anonymized and sent to Bochum where they were tested for plausibility and finally combined in a global database. The development and training of the NN expert system took place at Visiomed AG. As it was not clear if all the images were suitable for automated computer-aided analysis, a tool was developed for examining the original images to select adequate, 'clean', images. The criteria for image selection were that the lesion needed to be visible and fairly sharp, and that no artificial colour markings were allowed. Images with a low contrast (possibly due to very light coloured lesions) were of great importance for testing the limits of our algorithms. Therefore, such images were not rejected.

Camera description

Dermoscopy was performed with a video camera and image acquisition system based on a Pentium II 200 MHz PC. The PC included a frame-grabber board (Matrox) and a 1 GB magneto-optical disk which was used for exchanging of data. The video camera was built into a small (16 cm length), lightweight (750 g) hand-held unit and had the ability to take overview images and microscopic images at $\times 10$, $\times 20$ or $\times 30$ magnification. This hand-held unit contained a miniaturized 3 CCD 768×576 pixels camera (TU-IK48P; Toshiba, Tokyo, Japan), a specially developed optical lens system³³ and a light source.

The optical lens system was designed with three fixed magnification settings so that the skin was magnified 10, 20 or 30 times when the images were viewed on the 17-inch monitor. The effective pixel size on the skin was $10.4 \mu\text{m}$ for the $\times 30$ magnification. The lens system also allowed the acquisition of overview images of body parts. Illumination was by white light-emitting diodes (LEDs), fitted in optimized Perspex bodies which

Table 1. Number of images given for each diagnosis

Diagnosis	Histological	Final
Dysplastic naevus	149	149
Actinic lentigo	6	20
Lentigo simplex	10	39
Lentigo senilis	0	1
Café-au-lait spot	0	5
Seborrhoeic keratoses	48	66
Benign lesion n.o.s.	23	97
Naevoid lentigo/ junctional naevus	127	383
Compound naevus	262	633
Dermal naevus	62	164
Naevus spilus	1	8
Naevus coeruleus	21	26
Halo naevus	2	3
Congenital naevus	28	66
Angioma	9	15
Spitz naevus	17	29
Melanocytic naevus n.o.s.	0	311
Lentigo maligna ^a	24	24
Superficial spreading melanoma	115	115
Acrolentiginous melanoma	2	2
Nodular malignant melanoma	22	22
Unclassifiable melanoma	24	24
Basal cell carcinoma (pigmented)	16	16
Total	968	2218

^aAnd lentigo maligna melanoma; n.o.s., not otherwise specified.

were placed inside the head of the hand-held unit at a predetermined angle. All LEDs were independently software controlled. This arrangement permits the use of either direct or diffuse lighting, as needed.

After termination of the initial DANAOS study, the camera was redesigned to include improvements identified during the study and is now commercially available (microDERM).

Neural network expert system

For training the NN expert system, it was important that the samples represented a variety of diagnoses (Table 1). After capturing an image of a PSL, its outer border was determined. This process was referred to as 'boundary detection' or 'segmentation'. We employed a hybrid method that combined statistical clustering of the colour space and a hierarchical region-growing algorithm. Details of segmentation techniques have been previously published.^{26–29} In cases where hair covered parts of the PSL, a special filter was applied to reduce the influence of the hair on the segmentation. Once the lesion was successfully segmented, the feature extraction process could concentrate strictly on the lesion, its edge, and a margin area around it. The *asymmetry percentage* of a lesion was estimated by first finding the principal axes of inertia of the lesion, then 'folding' the image along these axes, computing the

magnitude of the nonoverlapping areas and finally dividing this by the total area of the lesion.³⁰ The *border irregularity* was expressed by the irregularity index³¹ and by fractal features of the boundary.³⁰ The *variation in colour* was expressed by the variance in red, green and blue colour components.²⁹ *Texture* also represents an important feature that was evaluated.³⁰ Given the inherent variation in position, orientation and magnification of skin lesions, the selected features needed to be unaffected by this factor. Some proposed features are invariant under either translation³⁴ or rotation.³⁵ We used features based on specialized Gabor wavelets which are scale, translation and rotation independent. The dataset collected during the study was separated into 10 packages which were used for the training and testing of an NN classification system. Nine of the 10 packages were used for training and one for testing. The training was repeated 10 times with different test sets. For the statistical analysis and graphing, the statistics tool 'R' was used.^{36,37}

Results

Thirteen DANAOS members made a total of 4280 clinical and 3519 dermoscopic digital images. The histological and final diagnoses of the PSLs are given in Table 1. Only 2465 (70%) of the dermoscopic images were of sufficiently high quality. Of these images,

Reference	Method			No. of cases			ROC	
	DA	ELM	C	B	M	T	Sensitivity	Specificity
1 Binder ²⁰	x			81	39	120	90	74
2 Argenziano ⁴⁶		x		86	60	146	95	75
3 Andreassi ³	x			90	57	147	81	88
4 Schindewolf ¹⁶	x			229	80	309	91	83
5 Menzies ³⁸	x			95	75	170	93	67
6 Seidenari ²¹	x			385	18	403	99	92
7 Ganster ³⁹	x			5281	108	5389	72	99.8
8 Benelli ⁴⁰			x	341	60	401	85	55
9		x		341	60	401	80	89
10 Cristofolini ⁴¹			x	187	33	220	85	75
11	x			187	33	220	88	79
12 Lorentzen ⁴²			x	183	49	232	78	89
13		x		183	49	232	82	94
14 Soyer ⁴³			x	94	65	159	94	82
15	x			94	65	159	94	82
16 Stanganelli ⁴⁴			x	207	55	262	67	84
17	x			210	58	268	96	97
18 Soyer ⁴⁵	x			73	35	108	95.93	48.1
19	x			73	35	108	98.02	28.8
20	x			73	35	108	99.17	12.6

Table 2. List of published sensitivity and specificity data for distinguishing benign and malignant pigmented skin lesions

ROC, receiver operating characteristic; DA, digital analysis; ELM, epiluminescence microscopy by human experts; C, clinical investigation by physicians; B, benign; M, malignant; T, total.

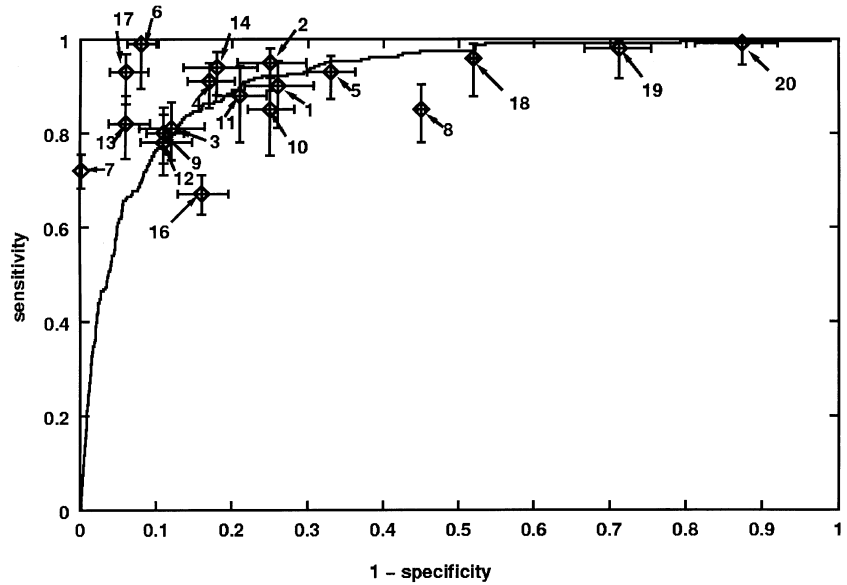


Figure 1. Receiver operating characteristic curve for the separation of benign and malignant skin lesions using our neural network compared with published sensitivity and specificity data for clinicians and other computer-aided classification systems. The numbers are references to line numbers in Table 2.

segmentation was successful in 2218 (90%), and hence analysis was made possible.

These 2218 PSLs were obtained from 1173 patients (694 females, 479 males) with a mean \pm SD age of 36.6 ± 16.4 years. Fifty-four patients were of skin phototype I, 418 type II, 430 type III, 114 type IV, 17 type V and one type VI; the skin type could not be defined in 139 patients.

Neural network performance

In order to compare the performance of the NN with that of other classification systems, sensitivity and

specificity data relating to clinical and computer-assisted diagnosis were retrieved from the literature (Table 2),^{3,16,20,21,38-46} and plotted on the same graph (Fig. 1) as the receiver operating characteristic (ROC) curve.³⁶ The ROC curve shows all possible compromises between sensitivity and specificity for the NN. Only papers with more than 30 malignant lesions and at least 100 PSLs in total were included. The error bars for the sensitivity and specificity were estimated using the number of cases given in the publication and using the assumption of a binormal ROC curve.³⁶ Lines 18-20 are taken from a study⁴⁵ on the performance of 40 dermatologists evaluating 108 PSLs. These sensitivity

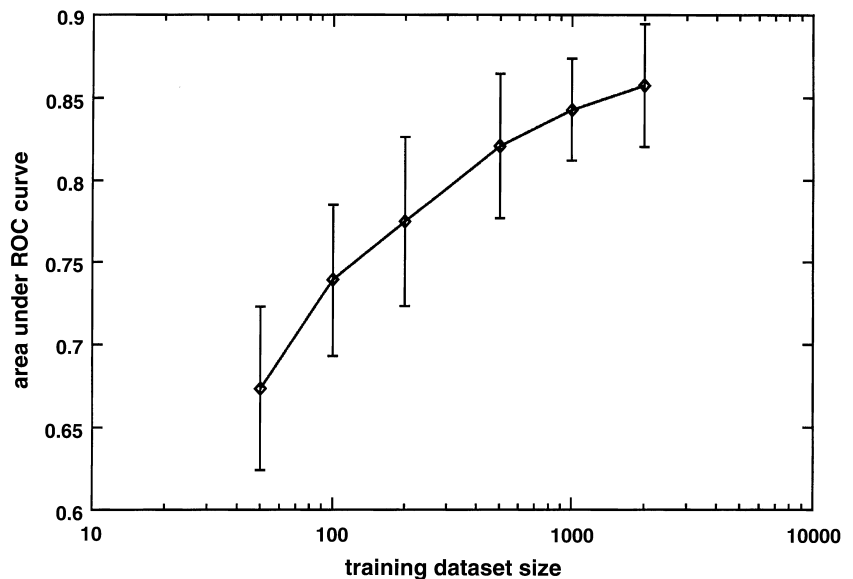


Figure 2. Area under the receiver operating characteristic (ROC) curve as function of training set size.

and specificity values were obtained by using the suggested management decision as the diagnosis. For instance, the value in line 18 is obtained if only lesions suggested for 'excision' are considered malignant and all other lesions are considered benign. Line 19 is obtained if lesions for suggested 'excision' as well as 'digital follow-up' are considered malignant and all other lesions are considered benign. The resulting NN expert system showed a performance similar to that of the clinicians.

Influence of database size on neural network performance

To estimate the size of the training database needed to train a good NN for the distinction of benign and malignant PSLs, several NNs were trained with different sized training sets between 50 and 2000 cases. In Figure 2, the area under the ROC curve³⁶ was plotted as a function of the number of training samples. The performance of the expert system generally improves with the size of the training dataset.

Influence of single-centre vs. multicentre data collection

To analyse the robustness of the classification with respect to different populations, three scenarios were investigated. (i) Single-centre learning and testing: data from only one site were selected to train an NN classification system. This system was then tested on data from the same site (omitting 10%). (ii) Generalization from single-centre learning: data from only one site were selected to train an NN classification system. This system was then tested on the data from the remaining sites. (iii) Multicentre learning and testing: an NN classification system was trained on a random selection of cases from all sites and was then tested on the remaining data (omitting 10%). The size of the training dataset was kept constant for all experiments.

The performance within the population of a single centre was slightly better than the performance on a mixed population. Results were worst for case (ii), where the training was performed using the population of a single site, but the classification system was then applied to a population from the other sites. The resulting areas under the ROC curve³⁶ are given in Table 3.

Influence of training on 'clean' data vs. training on all data

An important question during the training of the classification system was the selection criteria for the training data. One option was to use only data with

Table 3. Performance of single-centre vs. multicentre training and testing

Scenario	Training set	Test set	ROC area	± SD
(i)	Site 1	Site 1	0.847	0.046
(ii)	Site 1	All except 1	0.807	0.014
(iii)	All sites	All sites	0.826	0.021

ROC, receiver operating characteristic.

Table 4. Influence of the sorting of training data on classification performance

Scenario	Training set	Test set	ROC area
(i)	Clean	Clean	0.848
(ii)	Clean	'Noisy'	0.803
(iii)	All	All	0.844

ROC, receiver operating characteristic.

perfect image quality and to remove images with hair or air bubbles when training the expert system. The other option was to use all images acquired during the clinical practice regardless of any artefacts. We evaluated the influence of the data selection by performing three different learning and testing scenarios. The images were sorted into two groups, the *clean* ones and the *noisy* ones with hair or bubbles. Three scenarios were tested: (i) training on *clean* dataset, testing on *clean* dataset; (ii) training on *clean* dataset, testing on *noisy* dataset; and (iii) training on all data, testing on all data.

Scenarios (i) and (iii) showed very similar results but the performance declined in scenario (ii), when the system was trained on *clean* images and was tested on *noisy* images. The resulting areas under the ROC curve³⁶ are given in Table 4.

Discussion

What stands out in the DANAOS study is its size and breadth: a substantial number of dermoscopic and clinical images of PSLs from 13 dermatological centres located in nine European countries was collected using a common standardized image acquisition technique. The collected database includes a wide spectrum of PSLs in a representative distribution (Table 1). For those benign melanocytic naevi where no excision was indicated the lesion category may be considered arbitrary, but this should not have any influence on the results as we did not try to separate these cases in the analysis.

This dataset can be used as a basis for training in a computer-aided classification system for PSLs. The performance of the NN shown in Figure 1 is similar to the average performance published in the literature.

Combining the whole collection of published sensitivity and specificity data as a kind of meta-study, we conclude that the performance of our expert system is comparable with that of clinicians. This comparison should not be over-interpreted though, as the databases are different and the results are clearly dependent on the difficulty of the cases. In order to achieve comparable results, the dataset must be a representative selection of cases typical for a given clinical setting, such as the population seeking skin cancer screening. The size of the dataset is also important. To assure the generalization capabilities of the expert system, the data should be collected at different sites in different countries.⁴⁷

Analysis of the influence of the database size used for training has shown that performance increases with the size of the dataset. For a small number of cases, the increase is more important than for larger datasets, but even with 2000 cases the learning curve had not reached saturation. This result seems to confirm what is well known for the training of clinicians, that is, they must build their experience from a significant number of cases. Analysis of the influence of the multisite data collection shows that an expert system trained and tested on a single-site dataset displays a good classification performance, but when it is tested on data from other sites the performance is significantly lower. When training an expert system on the multicentre dataset, the performance is almost as good as in the case of single-site training and testing. Nevertheless, the results are valid for the whole diversity of the population included in the study.

As all experiments were done on equal-sized training sets it can be concluded that the benefit of a multicentre study is more than just the ability to collect more cases during the limited time of a study. Analysis of the influence of data preselection for training showed very similar results. When training on preselected 'clean' images which exclude all unsharp images or images with artefacts such as hair or air bubbles, the test results on other preselected 'clean' images are very good, but the performance is significantly degraded when the classification system is applied to images from the clinical practice where artefacts cannot always be avoided. When a representative set of images, including clean ones and some with artefacts, is used for training, the performance of the expert system is comparable, and the system is robust to artefacts.

Combining these results, we suggest training the expert system on a representative dataset from a multicentre collection of data without preselection to remove images with artefacts. The dataset should be as

large as possible. Looking at the performance of the expert system in more detail, we found that some images were correctly classified by the expert system but were difficult to classify manually, and some images were classified incorrectly by the expert system but were classified easily by human observers. Naturally, a diagnostic accuracy of 100% is unlikely to be attained. This is also true for experienced dermatologists.² In many cases it may be helpful to get a fast second opinion from an expert system, e.g. in a clinical environment where it is not always easy to get a quick second opinion from a clinician. An alternative approach is to obtain a quick second expert opinion by using teledermoscopic diagnosis, which was recently evaluated in a multicentre study.⁴⁸ The expert system described above, which was trained on a broad database, cannot replace routine clinical investigation, but it may be used as a second opinion to support the decision of an investigator. Based on our data and experience in earlier studies, we have demonstrated the potential of such an expert system.^{47,49-53}

In conclusion, a prototype for the computer-aided diagnosis of PSLs using NNs has been developed based on a Europe-wide multicentre study (DANAOS).⁵² The experience from this study demonstrates the need for a large database size, the usefulness of multicentre data collection, and the benefit of a representative collection of cases from clinical practice, and argues against the selection of only the best images. The resulting expert system has shown a performance similar to that of clinicians, while the difficulties of the expert system in diagnosing are not the same as those of clinicians.^{45,47,51} Therefore, we suggest that using the two in combination has the potential to increase the accuracy of PSL diagnosis. Thus improvement of early melanoma detection and reduction of unnecessary excisions may result from such a combination of expertise. Future studies are planned to quantify this benefit.

Acknowledgments

The DANAOS project was supported in part by the grant IN 10483 I DANAOS of the EC. We thank all coworkers who participated in the DANAOS study.

References

- 1 Altmeyer P, Hoffmann K, Stücker M. *Skin Cancer and UV Radiation*. New York: Springer-Verlag, 1997.
- 2 Morton CA, MacKie RM. Clinical accuracy of the diagnosis of cutaneous malignant melanoma. *Br J Dermatol* 1998; **138**: 283-7.

- 3 Andreassi L, Perotti R, Rubegni P *et al.* Digital dermoscopy analysis for the differentiation of atypical nevi and early melanoma. *Arch Dermatol* 1999; **135**: 1459–65.
- 4 Bafonuta ML, Beutcher A, Aegerter P, Saiag P. Is dermoscopy (epiluminescence microscopy) useful for the diagnosis of melanoma? Results of a meta-analysis using techniques adapted to the evaluation of diagnostic tests. *Arch Dermatol* 2002; **137**: 1343–50.
- 5 Carli P, De Giorgi V, Soyer HP *et al.* Dermoscopy in the diagnosis of pigmented skin lesions: a new semiology for the dermatologist. *J Eur Acad Dermatol Venereol* 2002; **14**: 353–69.
- 6 Soyer HP, Argenziano G, Ruocco V, Chimenti S. Dermoscopy of pigmented skin lesions (part II). *Eur J Dermatol* 2001; **11**: 483–98.
- 7 Argenziano G, Soyer HP. Dermoscopy of pigmented skin lesions—a valuable tool for early diagnosis of melanoma. *Lancet Oncol* 2001; **2**: 443–9.
- 8 Kittler H, Pehamberger H, Wolff K, Binder M. Diagnostic accuracy of dermoscopy. *Lancet Oncol* 2002; **3**: 159–65.
- 9 MacKie RM. An aid to the preoperative assessment of pigmented lesions of skin. *Br J Dermatol* 1971; **85**: 232–8.
- 10 Steiner A, Binder M, Schemper M *et al.* Statistical evaluation of epiluminescence criteria for melanocytic pigmented skin lesions. *J Am Acad Dermatol* 1993; **29**: 581–8.
- 11 Stolz W, Braun-Falco O, Bilek P *et al.* *Colour Atlas of Dermatoscopy*. Oxford: Blackwell Science, 1994.
- 12 Menzies SW. Automated epiluminescence microscopy: human vs machine in the diagnosis of melanoma. *Arch Dermatol* 1999; **135**: 1538–40.
- 13 Kenet RO, Kang S, Kenet BJ *et al.* Clinical diagnosis of pigmented lesions using digital epiluminescence microscopy. *Arch Dermatol* 1993; **129**: 157–74.
- 14 Green A, Martin N, Pfitzner J *et al.* Computer image analysis of melanoma in the diagnosis of melanoma. *J Am Acad Dermatol* 1994; **31**: 958–64.
- 15 Perotti R, Biagioli M, Burrioni M *et al.* Image analysis of pigmented lesions: clinico-histopathological correlation. *Am J Dermatopathol* 1994; **16**: 103–4.
- 16 Schindewolf T, Schiffner R, Stolz W *et al.* Evaluation of different image acquisition techniques for a computer vision system in the diagnosis of malignant melanoma. *J Am Acad Dermatol* 1994; **31**: 33–41.
- 17 Schindewolf T. *Verfahren zur Klassifikation melanozytärer Tumoren mit Hilfe der digitalen Farbbildanalyse*. Aachen: Shaker, 1997.
- 18 Sober A, Burstein J. Computerised digital image analysis: an aid for malignant melanoma diagnosis. *J Dermatol* 1994; **21**: 163–70.
- 19 Hall P, Claridge E. Computer screening for early detection of melanoma: is there a future? *Br J Dermatol* 1995; **132**: 325–38.
- 20 Binder M, Kittler H, Seeber A *et al.* Epiluminescence microscopy-based classification of pigmented skin lesions using computerized image analysis and an artificial neural network. *Melanoma Res* 1998; **8**: 261–6.
- 21 Seidenari S, Pellicani G, Gianetti A. Digital videomicroscopy and image analysis with automatic classification for detection of thin melanomas. *Melanoma Res* 1999; **9**: 163–71.
- 22 Debei RO, Decaestecker C, Pasteels JL *et al.* Computer-assisted analysis of epiluminescence microscopy images of pigmented skin lesions. *Cytometry* 1999; **37**: 255–66.
- 23 Husemann R, Tölg S, von Seelen W *et al.* Computerised diagnosis of skin cancer using neural networks. In: *Skin Cancer and UV Radiation* (Altmeyer P, Hoffmann K, Stücker M, eds). New York: Springer-Verlag, 1997; 1052–63.
- 24 Hoffmann K, Kaspar K, Happe M *et al.* Neue bildgebende Techniken für die Routine? Ein Blick in die Zukunft: DANAOS. Dermatologie. In: *Leitlinien und Qualitätssicherung für Diagnostik und Therapie* (Garbe C, Rassner G, eds). Berlin: Springer-Verlag, 1999; 20–6.
- 25 Nachbar F, Stoltz W, Merkle T *et al.* The ABCD rule of dermoscopy. High prospective value in diagnosis of doubtful melanocytic skin lesions. *J Am Acad Dermatol* 1994; **30**: 551–9.
- 26 Umbaugh SE, Moss RH, Stoecker WV. An automatic colour segmentation algorithm with application to identification of skin tumour borders. *Comput Med Imaging Graph* 1992; **16**: 227–35.
- 27 Ercal F, Mognati M, Stoecker WV, Moss RH. Detection of skin tumour boundaries in colour images. *IEEE Trans Med Imaging* 1993; **12**: 624–8.
- 28 Xu L, Jackowski M, Goshtasby A *et al.* Segmentation of skin cancer images. *Image Vision Computing* 1999; **17**: 65–74.
- 29 Schmid P. Segmentation of digitised dermatoscopic images by [2D] colour clustering. *IEEE Trans Med Imaging* 1999; **18**: 164–71.
- 30 Stoecker WV, Chiang CS, Moss RH. Texture in skin images: comparison of three methods to determine smoothness. *Comput Med Imaging Graph* 1992; **16**: 179–90.
- 31 Ercal F, Chawla A, Stoecker WV *et al.* Neural network diagnosis of malignant melanoma from colour images. *IEEE Trans Biomed Eng* 1994; **41**: 837–45.
- 32 Fitzpatrick TB. Soleil et peau. *J Med Esthet* 1975; **2**: 33–4.
- 33 Zentrum für Neuroinformatik GmbH. German Patent DE Nr. 197 533 92.3 (12.12.1997). Einstellvorrichtung für ein optisches System zur Abbildung von Oberflächen, insbesondere in der medizinischen Bildverarbeitung.
- 34 Liu JF, Lee JCM. An efficient and effective texture classification approach using a new notion in wavelet theory. Volume B. *Proc ICPR* 1996: 820–4.
- 35 Porter R, Canagarajah N. Robust rotation-invariant texture classification: [W]avelet, [G]abor filter and [GMRF] based schemes. *IEEE Proc Vision Image Signal Processing* 1997; **144**: 188.
- 36 Metz CE. Statistical analysis of ROC data in evaluating diagnostic performance. In: *Multiple Regression Analysis: Applications in the Health Sciences* (Herbert DE, Myers RH, eds). New York: American Institute of Physics; 1986; 365–84.
- 37 Nolan D, Speed T. *Stat Labs: Mathematical Statistics Through Applications*. Berlin: Springer-Verlag, 2000.
- 38 Menzies SW, Bischof LM, Peden G *et al.* Automated instrumentation for the diagnosis of invasive melanoma: image analysis of oil epiluminescence microscopy. In: *Skin Cancer and UV Radiation* (Altmeyer P, Hoffmann K, Stücker M, eds). New York: Springer-Verlag, 1997; 1064–70.
- 39 Ganster H, Pinz A, Röhner R *et al.* Automated melanoma recognition. *IEEE Trans Med Imaging* 2001; **3**: 233–239.
- 40 Benelli C, Roscetti E, Dal Pozzo V. Reproducibility of a dermoscopic method (7FFM) for the diagnosis of malignant melanoma. *Eur J Dermatol* 2000; **10**: 110–14.
- 41 Cristofolini M, Zumiani G, Bauer P *et al.* Dermatoscopy: usefulness in the differential diagnosis of cutaneous pigmentary lesions. *Melanoma Res* 1994; **4**: 391–4.
- 42 Lorentzen H, Weismann K, Petersen CS *et al.* Clinical and dermoscopic diagnosis of malignant melanoma: assessed by expert and non-expert groups. *Acta Derm Venereol (Stockh)* 1999; **79**: 301–4.
- 43 Soyer HP, Smolle J, Leitinger G *et al.* Diagnostic reliability of dermoscopic criteria for detecting malignant melanoma. *Dermatology* 1995; **190**: 25–30.

- 44 Stanganelli I, Serafini M, Bucchi LA. A cancer-registry-assisted evaluation of the accuracy of digital epiluminescence microscopy associated with clinical examination of pigmented skin lesions. *Dermatology* 2000; **200**: 11–16.
- 45 Soyer HP, Argenziano G, Chimenti S *et al.* Dermatoscopy of pigmented skin lesions. In: *An Atlas Based on the Consensus Net Meeting on Dermatoscopy 2000*. Milan: Edra Medical Publishing and New Media, 2001.
- 46 Argenziano G, Fabbrocini G, Carli P *et al.* Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions: comparison of the ABCD rule of dermatoscopy and a new 7-point checklist based on pattern analysis. *Arch Dermatol* 1998; **134**: 1563–70.
- 47 Schwarzer G, Vach W, Schuhmacher M. On the misuses of artificial neural networks for prognostic and diagnostic classification in oncology. *Stat Med* 2000; **19**: 541–61.
- 48 Piccolo D, Smolle J, Argenziano G *et al.* Teledermoscopy—results of a multicentre study on 43 pigmented skin lesions. *J Telemed Telecare* 2000; **6**: 132–7.
- 49 Stolz W, Schiffner R, Pillet L *et al.* Improvement of monitoring of melanocytic skin lesions with the use of a computerised acquisition and surveillance unit with a skin surface microscopic television camera. *J Am Acad Dermatol* 1996; **35**: 202–7.
- 50 Pott G, Schulz C, Andres M *et al.* Bild- und Datenanalysesystem für die automatisierte Hautkrebserkennung. In: *Proceedings 1. Tagung der Dermatologischen Wissenschafts- und Fortbildungsakademie NRW, 2–4 Oktober 1998, Bonn*.
- 51 Handels H, Ross T, Kreusch J *et al.* Feature selection for optimized skin tumour recognition using genetic algorithms. *Artif Intell Med* 1999; **16**: 283–97.
- 52 Kreutz M, Anschutz M, Gehlen S *et al.* Automated diagnosis of skin cancer using digital image processing and mixture-of-experts. In: *Bildverarbeitung für Medizin—Algorithmen, Systeme, Anwendungen Informatik Aktuell* (Handels H, Horsch A, Lehmann T, Meinzer HP, eds). Berlin: Springer-Verlag, 2001; 357–61.
- 53 Murali A, Stoecker WV, Moss RH. Detection of solid pigment in dermatoscopy images using texture analysis. *Skin Res Technol* 2000; **6**: 193–8.