

## ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ

## Анализ диагностики, проводимой дерматологами с использованием сверточных нейронных сетей, в проспективном клиническом исследовании взаимодействия человека и машины

Julia K. Winkler, MD; Andreas Blum, MD; Katharina Kommoss, MD; Alexander Enk, MD; Ferdinand Toberer, MD; Albert Rosenberger, MSc; Holger A. Haenssle, MD

**Актуальность.** Проведенные исследования свидетельствуют о том, что сверточные нейронные сети (CNN) выполняют классификацию новообразований кожи с эффективностью, характерной для квалифицированных дерматологов. Несмотря на имеющееся одобрение первых нейронных сетей для клинического применения, проспективные исследования, демонстрирующие преимущества диагностики, проводимой совместно человеком и машиной, отсутствуют.

**Цели.** Оценить, может ли использование одобренных к обращению на рынке сверточных нейронных сетей (CNN) быть полезным для дерматологов при классификации меланоцитарных новообразований кожи.

**Дизайн, параметры и участники исследования.** В этом проспективном диагностическом двухцентровом исследовании дерматологи проводили скрининг рака кожи невооруженным глазом и методом дерматоскопии. Дерматологи классифицировали подозрительные меланоцитарные новообразования по степени вероятности злокачественности (по шкале от 0 до 1, пороговое значение злокачественности 0,5) и предлагали план дальнейших действий (ничего не предпринимать, наблюдать, иссекать). Затем дерматоскопические изображения анализировали с помощью одобренной к обращению на рынке CNN – программы Moleanalyzer pro (FotoFinder Systems). Выданные CNN оценки злокачественности (от 0 до 1, пороговое значение злокачественности 0,5) передавали дерматологам с просьбой провести повторную оценку новообразований и пересмотреть первичные решения с учетом результатов диагностики CNN. Референсные диагнозы для 125 (54,8 %) были основаны на данных гистопатологического анализа новообразований, а в случае неиссеченных новообразований – на данных клинического наблюдения и экспертного консенсуса. Данные собирали в период с октября 2020 по октябрь 2021 года.

Аффилиации авторов: Department of Dermatology, University of Heidelberg, Heidelberg, Germany (Winkler, Kommoss, Enk, Toberer, Haenssle); Public, Private and Teaching Practice of Dermatology, Konstanz, Germany (Blum); Institute of Genetic Epidemiology, University Medical Center, Georg-August University of Goettingen, Goettingen, Germany (Rosenberger).

Автор, ответственный за переписку: Holger A. Haenssle, MD, Department of Dermatology, University of Heidelberg, Im Neuenheimer Feld 440, 69120 Heidelberg, Germany (holger.haenssle@med.uni-heidelberg.de).

**Основные результаты и показатели.** Первичными показателями оценки являлись чувствительность и специфичность анализа, проведенного дерматологами самостоятельно и с использованием CNN. Точность и характеристики площади под ROC-кривой (ROC AUC) считались дополнительными показателями.

**Результаты.** В общей сложности 22 дерматолога выявили 228 подозрительных меланоцитарных новообразований (190 невусов, 38 меланом) у 188 пациентов [средний возраст (диапазон) составлял 53,4 (19–91) лет; 97 (51,6 %) участников были мужчины]. Чувствительность и специфичность диагностики значительно улучшались, когда при принятии решений дерматологи учитывали результаты диагностики, проведенной CNN, [средняя чувствительность возрастала с 84,2 % (95%-й доверительный интервал (ДИ), 69,6–92,6 %) до 100,0 % (95%-й ДИ, 90,8–100,0 %);  $p = 0,03$ ; среднее значение специфичности с 72,1 % (95%-й ДИ, 65,3–78,0 %) до 83,7 % (95%-й ДИ, 77,8–88,3 %);  $p < 0,001$ ; среднее значение точности с 74,1 % (95%-й ДИ, 68,1–79,4 %) до 86,4 % (95%-й ДИ, 81,3–90,3 %)  $p < 0,001$ ; и средняя площадь под ROC-кривой (ROC AUC) с 0,895 (95%-й ДИ, 0,836–0,954) до 0,968 (95%-й ДИ, 0,948–0,988);  $p = 0,005$ ]. Кроме того, при самостоятельной классификации меланоцитарных новообразований CNN достигала сопоставимой с дерматологами чувствительности, более высоких специфичности и диагностической точности. Число неоправданных иссечений доброкачественных новообразований сокращалось на 19,2 %, с 104 (54,7 %) из 190 доброкачественных невусов до 84, когда дерматологи пользовались оценкой CNN ( $p < 0,001$ ). Большинство новообразований обследовали дерматологи с 2–5-летним (96, 42,1 %) или менее чем двухлетним опытом работы (78, 34,2 %); меньшую часть (54, 23,7 %) оценивали дерматологи с опытом свыше пяти лет. Дерматологи с меньшим опытом в дерматоскопии при использовании CNN имели более заметные диагностические улучшения по сравнению с опытными дерматологами.

**Выводы.** В данном проспективном диагностическом исследовании было продемонстрировано, что эффективность диагностики, проводимой дерматологом, можно улучшать, используя одобренные CNN, и что широкое применение такой коллаборации человека и машины может быть полезным и для врачей, и для пациентов.

## КЛЮЧЕВЫЕ ПОЛОЖЕНИЯ

**Вопрос.** Выигрывают ли дерматологи от использования одобренных сверточных нейронных сетей (CNN) при классификации меланоцитарных новообразований?

**Результат.** В данном проспективном диагностическом исследовании при анализе меланоцитарных новообразований CNN достигала сопоставимой с дерматологами чувствительности, более высокой специфичности и более высокой диагностической точности, а при использовании дерматологами оценок CNN эффективность диагностики значительно повышалась. Более высокая специфичность CNN помогала дерматологам иссекать гораздо меньше доброкачественных невусов, кроме того, у дерматологов с небольшим опытом работы эффективность диагностики при использовании CNN улучшалась наиболее заметно.

**Значение.** Необходимо продолжать изучать такое сотрудничество человека и машины, и в перспективе внедрять его в клиническую практику, в частности при классификации меланокитарных новообразований в дерматологии.

Частота заболеваемости раком кожи остается высокой во всем мире, и ранняя его диагностика крайне важна для пациента [1, 2]. При классификации рака кожи CNN достигали диагностической точности, характерной для квалифицированных дерматологов [3–6]. Большинство предыдущих исследований были ретроспективными и изучали изображения новообразований с подтвержденным диагнозом [7]. В одном из ранних исследований, сравнивавших эффективность диагностики человеком и машиной, дерматологи и CNN классифицировали дерматоскопические снимки новообразований кожи [4, 5]. Многочисленные исследования, проведенные с тех пор, подтвердили высокую эффективность обоих методов диагностики, но также разоблачили важные ограничения, особенно повышенное количество ошибочных диагнозов, поставленных на основании изображений, имевших артефакты, такие как элементы шкалы, рисунок кожи [8, 9], или в отношении редких новообразований, локализованных на слизистых или под ногтевой пластиной [10]. Несмотря на одобрение к обращению на европейском рынке ряда нейронных сетей для классификации новообразований кожи, не хватает проспективных клинических исследований, анализирующих интеграцию CNN-диагностики в повседневную клиническую практику после «живого» обследования пациентов. Большинство предыдущих исследований при оценке сотрудничества человека и машины ограничивались ретроспективным методом, когда дерматологов просили проанализировать изображения новообразований кожи с учетом результатов CNN-классификации или без доступа к ним [11, 12]. Однако имеются важные различия между ретроспективным и проспективным исследованиями, изучающими диагностику новообразований кожи дерматологами и CNN. Во-первых, в проспективном исследовании дерматологи могут непосредственно осматривать и обследовать пациентов (живое обследование), тогда как в большинстве ретроспективных исследований можно анализировать только отдельные дерматоскопические изображения. Во-вторых, клинические решения в проспективном исследовании имеют прямую связь с будущим состоянием здоровья пациента, тогда как в ретроспективных исследованиях тяжелые последствия пропуска каких-либо злокачественных новообразований отсутствуют. Таким образом, необходимо выяснить, могут ли и каким образом дерматологи внедрять CNN-рекомендации в практику принятия клинических решений.

Проводя данное исследование, мы стремились пролить свет на взаимодействие дерматологов с одобренными CNN в проспективных клинических условиях. Кроме того, мы использовали утвержденную анкету, измеряющую отношение и доверие пациентов к тестируемым CNN.

## МЕТОДЫ

Проспективное диагностическое исследование было одобрено комитетом по этике медицинского факультета Гейдельбергского университета (номер одобрения S-836/2020) и проводилось с соблюдением принципов Хельсинкской декларации. Все пациенты предоставили письменное информированное согласие до начала каких-либо процедур, связанных с иссле-

дованием. Также соблюдались рекомендации стандартов в отношении отчетов о диагностической точности (STARD).

## Условия исследования

Данное проспективное диагностическое обсервационное клиническое исследование было призвано изучить взаимодействие дерматологов и CNN, одобренной для клинического применения – программы Moleanalyzer pro (FotoFinder systems), и оценить возможное его влияние на эффективность диагностики. По статистическим соображениям это исследование рассматривало только меланоцитарные новообразования. Исследование проводили в Германии в двух центрах: в Департаменте дерматологии Гейдельбергского университета и в частной дерматологической клинике (Public, Private and Teaching Practice of Dermatology, Andreas Blum) города Костанц. Все процедуры описаны в схеме исследования (электронный Рисунок 1 в Приложении 1). Вкратце процедура заключалась в следующем: дерматологи с разным опытом применения дерматоскопии в скрининге рака кожи (5 лет) проводили полное обследование тела невооруженным глазом и с использованием дерматоскопии. Их просили определить вероятность злокачественности подозрительных меланоцитарных новообразований по визуальной аналоговой шкале от 0 до 1, используя пороговое значение как минимум 0,5 в случае подозрения, что новообразование является меланомой. Решения дерматологов относительно новообразования (ничего не предпринимать, наблюдать, иссекать) фиксировали. Затем пациентов направляли в другой кабинет для анализа при помощи CNN. Программа оценивала злокачественность новообразования по шкале от 0 до 1, минимальным пороговым значением было 0,5 [4, 5]. Оценки, присвоенные CNN, передавали обследовавшим новообразования дерматологам, которых просили пересмотреть диагноз и план дальнейших действий с учетом результатов CNN-диагностики. И наконец, дерматологов просили высказать мнение, была ли оценка CNN полезной и/или убедительной. Пациентов просили заполнить опросник, чтобы оценить степень их доверия и отношение к системам диагностики на основе CNN. Опросник включал 10 пунктов, основанных на утвержденной анкете «Доверие к медицинской технологии» [13, 14]. Гистопатологический анализ проводил опытный сертифицированный гистопатолог, который в сложных для диагностики случаях прибегал к консультациям с коллегами для достижения консенсуса.

## Статистический анализ

Выборку в объеме 183 пациентов сочли необходимой для выявления улучшенного показателя специфичности 90 % (минимальный порог > 75 %) у дерматологов, использовавших CNN, и у дерматологов, дававших самостоятельную оценку новообразований, со статистической мощностью 90 % и статистической значимостью 5 %. Первичными показателями оценки являлись чувствительность и специфичность выявления меланомы. Точность и площадь под ROC-кривой (ROC AUC) считались дополнительными критериями. Вторичными критериями были эффективность самой работы CNN, а также отношение и доверие к ней пациентов, согласно опроснику. Референсный диагноз исследованных новообразований (эталонные данные) ос-

новывался на результатах гистопатологического анализа иссеченных новообразований или данных клинического наблюдения и экспертного консенсуса в случае неиссеченных новообразований. Для статистического подсчета дихотомических диагностических классификаций меланоцитарных новообразований (доброкачественное/злокачественное) использовали решения, принятые дерматологами. Решения «иссекать» для меланом и «наблюдать» или «ничего не предпринимать» для меланоцитарных невусов считались истинно положительными или истинно отрицательными соответственно. Всякий раз, когда дерматологи рекомендовали иссечение, их просили обозначить причину/причины такого решения (клинические характеристики, дерматоскопические характеристики, анамнез/обеспокоенность пациента). Мы исследовали разницу результатов диагностики, проведенной только CNN, и диагностики, проведенной дерматологами, принимавшими решение с учетом мнения CNN или не зная его. Мы применяли тест МакНемара, чтобы выявить разницу пропорций категориальных переменных, и ранговый критерий Уилкоксона для оценки непрерывных данных. Кроме того, проводили парное статистическое сравнение ROC AUC [15]. Поправку Бонферрони использовали для подгонки множественного тестирования. Статистическая значимость составляла  $p < 0,05$ . Тесты на определение значимости были двухсторонними. Для всех видов анализа использовали статистическую программу SPSS, version 25 (IBM).

## РЕЗУЛЬТАТЫ

### *Пациенты и характеристики новообразований*

В данном исследовании 22 дерматолога обследовали 188 пациентов [средний возраст (диапазон) составлял 53,4 (19–91) лет; из них 97 (51,6 %) мужчин, Таблица 1] и выявили 228 подозрительных меланоцитарных новообразований. В общей сложности 166 (88,3 %) пациентов были вовлечены в исследование в Отделении дерматологии Гейдельбергского университета, Германия, и 22 (11,7 %) пациента в Public, Private and Teaching Practice of Dermatology в Констанце, Германия. Большинство пациентов имели типы кожи по Фицпатрику II (33,5 %) или III (56,4 %). Большое суммарное количество невусов на одного пациента ( $> 50$ ) было зафиксировано у 30,9 % участников, (51–100 у 33 пациентов,  $> 100$  у 25 пациентов). В исследовании участвовал 51 (27,1 %) пациент, имевший меланому в прошлом, и 25 (13,3 %) пациентов со множественными ( $> 5$ ) атипичными невусами. Семейная история меланомы имела у 13 (6,9 %) пациентов. Немеланоцитарный рак кожи ранее был диагностирован у 29 (15,4 %) участников. В общей сложности у 111 (59,0 %) пациентов не было личной или семейной истории рака кожи. Из 228 подозрительных новообразований 190 (83,3 %) были невусами и 38 (16,7 %) меланомами (Таблица 1). Большинство новообразований были локализованы на торсе ( $n = 148$ , 64,9 %), нижних ( $n = 35$ , 15,4 %) и верхних конечностях ( $n = 22$ , 9,6 %). Важно, что наше исследование также рассматривало новообразования особой локализации, в том числе 18 на голове и шее, 3 – на коже акральных участков тела, 2 – под ногтевой пластиной. Для 125 (54,8 %) новообразований референсный диагноз ставили на основе отчетов о гистопатологическом анализе. Гистопатологическое исследование выявило 44 диспластических невуса, не предполагавших меланомы или рекомендаций к повторному иссечению. Диагноз для оставшихся неиссеченных новообразований был подтвержден путем клинического наблюдения и/или экспертным консенсусом.

**Таблица 1.** Характеристики пациентов и новообразований, включенных в исследование

Характеристика	Количество (%)
Пациенты (188 человек)	
Пол	
Женский	91 (48,4)
Мужской	97 (51,6)
Тип кожи по Фицпатрику	
I	5 (2,7)
II	63 (33,5)
III	106 (56,4)
IV	12 (6,4)
V	2 (1,1)
VI	0 (0)
Количество невусов	
0–15	42 (22,3)
16–50	88 (46,8)
51–100	33 (17,6)
> 100	25 (13,3)
Количество атипичных невусов	
0–5	163 (86,7)
> 5	25 (13,3)
Личная/семейная история меланомы	
Личная история меланомы	51 (27,1)
Личная история немеланоцитарного рака кожи	29 (15,4)
Семейная история меланомы	13 (6,9)
Скрининг рака кожи в прошлом	74 (39,6)
Новообразования (количество 228)	
Подтвержденный диагноз	
Меланоцитарный невус	190 (83,3)
Меланома	38 (16,7)
Меланома in situ	12 (5,3)
Инвазивная меланома (медианная толщина 1,0 мм)	26 (11,4)
Локализация	
Голова и шея	18 (7,9)
Торс	148 (64,9)
Верхние конечности	22 (9,6)
Нижние конечности	35 (15,4)
Кожа акральных участков	3 (1,3)
Ногти	2 (0,9)
Вид подтверждающей диагностики	
Отчет о гистопатологическом исследовании	125 (54,8)
Контрольное наблюдение/экспертное мнение	103 (45,2)



## Диагностическая классификация

Дерматологи, которые проспективно обследовали пациентов и новообразования (живое обследование, без доступа к результатам CNN-диагностики), достигали средних показателей диагностической чувствительности 84,2 % (95%-й ДИ, 69,6–92,6 %) и специфичности 72,1 % (95%-й ДИ, 65,3–78,0 %) (Таблица 2). После получения и учета результатов CNN-оценки дерматологи значительно улучшали степень средних чувствительности и специфичности до 100 % (95%-й ДИ, 90,8–100,0%;  $p=0,03$ ) и 83,7% (95%-й ДИ, 77,8–88,3%;  $p<0,001$ ) соответственно. CNN-диагностика сама по себе, при которой оценивался только один дерматоскопический снимок одного новообразования, достигала средней чувствительности 81,6 % (95%-й ДИ, 66,6–90,8 %) и специфичности 88,9 % (95%-й ДИ, 83,7–92,7 %). Не имея доступа к результатам CNN, дерматологи достигали среднего значения диагностической точности 74,1 % (95%-й ДИ, 68,1–79,4 %), которое существенно улучшалось до 86,4 % (95%-й ДИ, 81,3–90,3 %) при использовании оценок CNN ( $p < 0,001$ ). Средний показатель ROC AUC при оценке только дерматологом составлял 0,895 (95%-й ДИ, 0,836–0,954) и повышался до 0,968 (95%-й ДИ, 0,948–0,988), когда дерматологи использовали результаты CNN-оценки ( $p = 0,005$ , Рисунок 1). Важно, что средняя чувствительность самостоятельной диагностики дерматологом и CNN были сопоставимы [84,2 % (95%-й ДИ, 69,6–92,6 %) против 81,6 % (95%-й ДИ, 66,6–90,8 %);  $p > 0,99$ ]; специфичность CNN была значительно выше, чем дерматологов [72,1 % (95%-й ДИ, 65,3–78,0 %) против 88,9 % (95%-й ДИ, 83,7–92,7 %);  $p < 0,001$ ]. В результате средний процент корректных диагнозов (точность) был значительно выше у CNN, чем у дерматологов [87,7 % (95%-й ДИ, 82,8–91,4 %) против 74,1 % (95%-й ДИ, 68,1–79,4 %);  $p < 0,001$ ]. Среднее значение ROC AUC CNN-диагностики [0,904 (95%-й ДИ, 0,856–0,951)] было не существенно, но выше по сравнению с дерматологами [0,895 (95%-й ДИ, 0,836–0,954);  $p = 0,82$ ] (Рисунок 1).

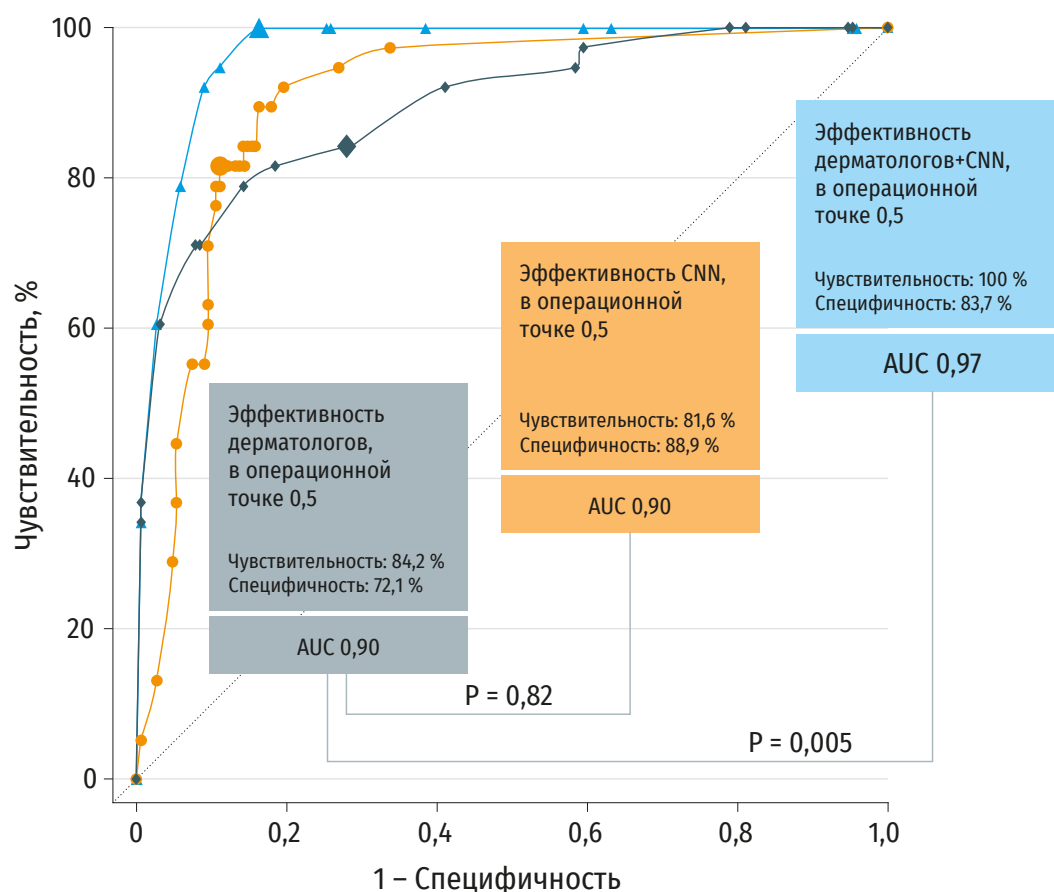
**Таблица 2.** Чувствительность, специфичность, точность диагностической классификации и клинических решений дерматологов, CNN и взаимодействия дерматологов и CNN<sup>a</sup>

	Чувствительность	Специфичность	Точность
Классификация			
Только дерматолог	84,2 (69,9–92,6)	72,1 (65,3–78,0)	74,1 (68,1–79,4)
Только CNN	81,6 (66,6–90,8)	88,9 (83,7–92,7)	87,7 (82,8–91,4)
Дерматолог + CNN	100,0 (90,8–100,0)	83,7 (77,8–88,3)	86,4 (81,3–90,3)
План действий			
Только дерматолог	97,4 (86,5–99,5)	45,3 (38,3–52,4)	53,9 (47,5–60,3)
Дерматолог + CNN	100,0 (90,8–100,0)	55,8 (48,7–62,7)	63,2 (56,7–69,2)

<sup>a</sup>Все данные выражены как среднее значение в процентах (95%-й доверительный интервал (ДИ)).

## Управление решениями

Кроме диагностической классификации, фиксировали план дальнейших действий, предложенный дерматологами (Таблица 2). Среднее значение чувствительности решения, пред-



**Рисунок 1.** ROC-кривые для бинарной классификации невусов/меланомы только дерматологами, только CNN, дерматологами с применением CNN

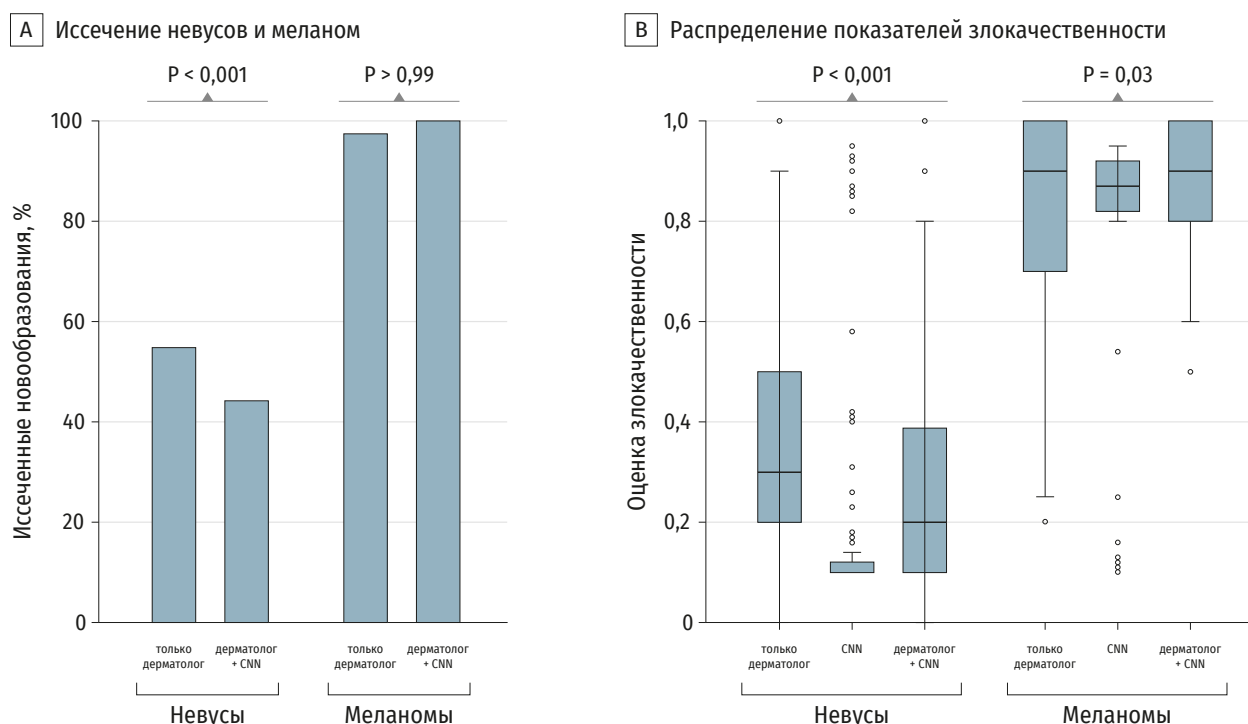
ложенного дерматологами, составляло 97,4 % (95%-й ДИ, 86,5–99,5 %) и возрастало до 100 % (95%-й ДИ, 90,8–100,0 %), если дерматолог использовал CNN ( $p > 0,99$ ). Имея доступ к результатам CNN-диагностики, дерматологи существенно повышали среднее значение специфичности с 45,3 % (95%-й ДИ, 38,3–52,4 %) до 55,8 % (95%-й ДИ, 48,7–62,7 %;  $p < 0,001$ ). Дерматологи изначально рекомендовали иссечение 104 из 190 (54,7 %) доброкачественных невусов. После просмотра результатов CNN-диагностики и их учета, количество неоправданных иссечений значительно сокращалось на 19,2 % с 104 до 84 невусов ( $p < 0,001$ ; Рисунок 2 А). В то же время частота иссечения злокачественных новообразований при использовании результатов CNN-диагностики существенно не менялась ( $p > 0,99$ ). Процент невусов, выявленных в ходе контрольного наблюдения, немного увеличивался после CNN-диагностики (от 37,9 до 44,7 %), но разница не достигала статистической значимости ( $p = 0,053$ ).

## Распределение показателей злокачественности

Оценку злокачественности для подозрительных новообразований давали дерматологи, CNN независимо и дерматологи с использованием CNN (Рисунок 2 В). Средний показатель злокачественности в отношении невусов составлял 0,35 (95%-й ДИ, 0,31–0,38) при диагностике только дерматологом, 0,20 (95%-й ДИ, 0,16–0,23) только CNN и 0,27 (95%-й ДИ, 0,24–0,30) дерматологами, использовавшими оценку CNN. У дерматологов наблюдались значительно более



низкие (следовательно, улучшенные) оценки злокачественности невусов после ознакомления с результатами CNN-диагностики ( $p < 0,001$ ). Средние показатели злокачественности для меланом составляли 0,80 (95%-й ДИ, 0,72–0,88) при диагностике только дерматологом, 0,74 (95%-й ДИ, 0,64–0,84) при диагностике только CNN и 0,86 (95%-й ДИ, 0,81–0,91) при анализе дерматологом с использованием результатов CNN-диагностики. То есть показатели дерматологов, учитывавших оценку сверточных нейронных сетей, опять же были значительно выше (и следовательно, лучше), чем при диагностике только дерматологом ( $p = 0,03$ ). В целом диаграмма оценок злокачественности, выданных CNN, имела более маленькие блоки из-за меньшего разброса (различия) результатов по сравнению с самостоятельной оценкой дерматологов (Рисунок 2 В).



**Рисунок 2.** Столбиковая и блочная диаграммы, иллюстрирующие иссеченные невусы и меланомы, а также распределение оценок злокачественности

А. Эта столбиковая диаграмма изображает количество невусов и меланом, иссеченных по решению только дерматологов и дерматологов при помощи CNN.

В. Эта блочная диаграмма показывает распределение оценок злокачественности (диапазон 0–1) невусов и меланом, согласно оценке только дерматологов и дерматологов при помощи CNN. Верхняя и нижняя границы блоков представляют собой 25-й и 75-й процентиля, тогда как медианы изображены горизонтальными линиями внутри блоков. Усики блоков показывают самые высокие и самые низкие значения, не превышающие 1,5 межквартильных размахов вне блоков.

### Зависимость эффективности диагностики от клинического опыта

Все 22 дерматолога, принимавших участие в исследовании, указывали свой опыт в дерматоскопии. Большую часть новообразований обследовали дерматологи с опытом работы от 2 до 5 лет (96, 42,1 %) или менее двух лет (78, 34,2 %). Диагностику остальных новообразований проводили дерматологи с более чем пятилетним опытом работы (54, 23,7 %).

Из 78 обследованных новообразований у дерматологов с опытом менее двух лет наблюдалось существенное повышение процента постановки корректного диагноза с 70,5 до 87,2 % после изучения оценок злокачественности, полученных при помощи CNN-диагностики ( $p < 0,01$ ) (электронная таблица 1 в Приложении 1). Аналогичным образом из 96 обследованных новообразований точность диагноза дерматологов с 2–5-летним опытом работы повышалась с 77,1 до 91,7 % ( $p < 0,01$ ). И наоборот, из 54 оцененных новообразований точность диагноза дерматологов с более чем 5-летним опытом работы, показывало лишь незначительное повышение, не достигавшее статистической значимости (74,1 % против 75,9 %,  $p > 0,99$ ).

### **Перспективы для дерматологов и пациентов при использовании CNN**

Дерматологов просили высказать их личное мнение о CNN-диагностике в отношении каждого из 228 обследованных новообразований. Для большинства новообразований дерматологи соглашались, что оценки CNN были убедительными (77,6 %, 159 из 205 ответов) и/или помогли (84,4 %, 173 из 205 ответов). Кроме того, в этом исследовании от 152 пациентов были собраны заполненные опросники, в которых они оценивали отношение и доверие к CNN-диагностике (электронный рисунок 2 в Приложении 1). Большинство пациентов согласилось, что CNN могут повышать достоверность диагностики, проводимой врачом [85 (56,3 %) абсолютно согласен, 48 (31,8 %) согласен]. Большая часть пациентов была готова принять более продолжительное обследование в связи с дополнительной диагностикой при помощи CNN [50 (33,3 %) полностью согласен, 60 (40 %) согласен]. Но очень немногие считали, что классификация CNN может полностью заменить мнение врача [8 (5,3 %) полностью согласен, 20 (13,2 %) согласен]. Большинству пациентов требуется экспертное мнение врача для интерпретации результатов CNN-диагностики [120 (79,5 %) полностью согласен, 28 (18,5 %) согласен].

## **ОБСУЖДЕНИЕ**

Ранняя диагностика меланомы является важнейшим фактором для прогнозирования течения заболевания. Однако врачи могут иметь разный уровень квалификации и опыта, что напрямую связано с успехом диагностики. Поэтому важно разработать инструменты поддержки врачей, чтобы (1) не пропустить меланому, (2) ограничить неоправданные иссечения и (3) сократить количество диагностически спорных новообразований, а также уменьшить количество ненужных повторных обследований. В этой связи нейронная сеть, обученная распознавать рак кожи, показала равную с дерматологами, а иногда и превосходящую степень диагностической точности [3–6]. Для решения различных задач в ходе исследований изучали подходы к взаимодействию человека и машины [16–18] и обсуждали стратегии оптимизированного взаимодействия при скрининге рака кожи [11]. На данный момент, насколько нам известно, это первое проспективное исследование в клинических условиях для оценки того, выигрывают ли дерматологи от использования CNN-диагностики при принятии решений. Это проспективное диагностическое исследование проводилось в университетской больнице и

в амбулаторной клинике, чтобы обеспечить показательную популяцию пациентов и спектр новообразований. Важно отметить, что новообразования специфической локализации (например, на коже акральных участков, под ногтевой пластиной) также были включены в исследование, обычно их диагностика затруднительна для врачей и CNN, вероятнее всего из-за недостаточного количества практических случаев, необходимых для приобретения опыта и обучения [10]. Многие участники ожидаемо имели повышенный риск меланомы (история меланомы, множественный атипичный невус), поскольку пациенты с высоким риском чаще и регулярнее обращаются за обследованием в специализированные центры [19, 20]. В данном исследовании дерматологи проводили живые обследования, что позволяло основывать выбор плана дальнейших действий на клиническом и дерматоскопическом описании меланотических новообразований, а также анамнезе и профиле риска каждого пациента. CNN-классификация, наоборот, была основана только на анализе дерматоскопических снимков. Тем не менее, как и в предыдущих ретроспективных исследованиях, CNN достигала высокой диагностической эффективности, показывая аналогичный дерматологам уровень чувствительности, но значительно более высокий уровень специфичности (88,9 % и 72,1 % соответственно) [4, 5]. Независимо от таких убедительных результатов CNN, основными критериями оценки в данном исследовании были разница чувствительности, специфичности, точности и ROC AUC до и после ознакомления дерматологов с результатами CNN-диагностики. Насколько нам известно, на сегодняшний день отсутствуют данные исследований, показывающие, в какой мере дерматологи могли бы использовать рекомендации CNN и пересматривать свои первоначальные решения. Интересно, что в данном исследовании все ранее упомянутые основные показатели оценки существенно улучшались после того, как дерматолог получал доступ к результатам CNN-диагностики (примерно на 10–15 %). Результаты данного проспективного исследования полностью подтверждают данные ретроспективных исследований, в которых анализировались изображения новообразований вместо живых обследований. При изучении научной литературы Tschandl и соавторы [11] обнаружили, что точность диагностики человеком с использованием мультиклассовых вероятностей на основе искусственного интеллекта увеличивалась с 63,6 до 77 % в отношении большого спектра новообразований кожи, что соответствует 10–15%-му улучшению, обнаруженному в нашем исследовании. Nekler и соавторы [12] сообщали об удивительных диагностических улучшениях у врачей, использующих поддержку CNN (точность от 43 до 83 %, чувствительность с 66 до 89 %, специфичность с 62 до 84 %), в то время как эффективность диагностики дерматологами без поддержки CNN была довольно низкой. Кроме того, в ретроспективном исследовании Maron и соавторов [21] дерматологи с поддержкой CNN показывали существенное повышение средних показателей чувствительности (с 59,4 до 74,6 %) и точности (с 65 до 73,6 %) при почти неизменной специфичности (70,6 против 72,4 %). В проспективных условиях данного исследования дерматологи с CNN-поддержкой не пропускали ни одной меланомы, хотя чувствительность и CNN, и дерматологов при независимой диагностике была относительно низкой. Поэтому сочетание диагностики дерматологом и CNN-оценки превосходит каждый из методов по отдельности, что является главным выводом данного исследования. Очевидно, что, когда CNN давали оценку злокачественности, отличную от первичного диагноза дерматолога, это служило поводом критического пересмотра решений. Общий результат настоящего исследования лучше всего иллюстрируют ROC-кривые, графически изображающие диагностику дерматологами, CNN и совместную их работу. Здесь ROC AUC для дерматологов (0,895) и CNN (0,904) существенно не различались и были близки

к ранее опубликованным результатам (и в отношении CNN-диагностики меланоцитарных новообразований: 0,86, и в отношении CNN-диагностики широкого спектра новообразований: 0,918) [4, 5]. То есть CNN помогали существенно увеличивать AUC до заметно более высокого уровня 0,968. Неудивительно, что менее опытные дерматологи больше выигрывали от поддержки CNN [11]. Это наблюдение укрепляет позицию о необходимости внедрения CNN в ежедневную практику менее специализированных медучреждений. В повседневной практике решение о плане дальнейших действий даже важнее диагностических решений. Дерматологи без поддержки CNN в нашем исследовании показывали высокую чувствительность 97,4 %; но низкую специфичность, достигавшую лишь 45,3 %. Это подчеркивает, что в клинических условиях дерматологи проявляют осторожность и иссекают больше новообразований, чтобы не пропустить меланому. В нашем исследовании внедрение CNN-диагностики сокращало число неоправданных иссечений доброкачественных новообразований на 19,2 % и тем самым существенно повышала специфичность до 55,8 %. Эти результаты соответствуют выводам исследования Tschandl и соавторов [11], которые просили дерматологов пересмотреть свои решения после получения доступа к результатам CNN-диагностики, и обнаружили, что дерматологи переключались с иссечений на наблюдение в отношении 15,5 % доброкачественных новообразований, при этом не допуская пропуска иссечений злокачественных новообразований. Для более глубокого изучения взаимодействия дерматологов и CNN при диагностике мы оценивали распределение оценок злокачественности. Столбики диаграммы показывали, что оценки CNN в основном располагались в верхней (для меланом) или нижней (для невусов) частях шкалы, тогда как дерматологи скорее показывали средние оценки, ближе к предельному значению злокачественности (0,4–0,6), особенно когда не чувствовали уверенности в своем диагнозе. В случаях, когда характер невусов был спорным для дерматологов, низкие оценки злокачественности от CNN часто убеждали не иссекать новообразования и в некоторых случаях рекомендовать наблюдение и контрольное обследование. Интересно, что ранее считалось, что обеспечение доверительного уровня, свойственного CNN в отношении результата классификации (например, в байесовской модели глубокого обучения с известным риском), может продолжать улучшать взаимодействие человека и компьютера [22]. Кроме основных показателей, мы также оценивали отношение врачей и пациентов к применению в диагностике CNN. Дерматологи соглашались, что для классификации большинства новообразований поддержка CNN была убедительной и/или полезной. В целом эти данные подтверждают общее оптимистичное отношение дерматологов к CNN-диагностике [23]. Также оптимистичными по отношению к CNN были и результаты использованного в исследовании опросника [24]. Тем не менее большинство пациентов все же хотели получить интерпретацию результатов экспертом и отвергали полную замену врача нейронной сетью [25].

## Ограничения

Во-первых, мы включали в исследование только меланоцитарные новообразования, чтобы лучше осветить взаимодействие человека и машины в проспективных, но хорошо контролируемых условиях, вместо того чтобы исследовать полностью обобщенные данные об эффективности. Поэтому результаты данного исследования не репрезентативны для условий с

широким спектром новообразований и менее распространенными диагнозами. Во-вторых, проспективный характер исследования сделал возможным подсчет истинной чувствительности в отношении всех новообразований пациентов. Статистическая оценка в нашем исследовании применялась только к тем 228 новообразованиям, которые участвовавшие в исследовании дерматологи проспективно считали подозрительными. В-третьих, у большинства пациентов исследования тип кожи был II или III. Поэтому эффективность данных в отношении человека, CNN и их взаимодействия может сильно отличаться применительно к более светлой (тип I) или более темной коже (тип IV и >), что требует дальнейшего изучения [26]. И наконец, многие пациенты в данном исследовании имели повышенный риск меланомы, что не позволяет транслировать результаты на всю популяцию пациентов.

## ВЫВОДЫ

Насколько нам известно, мы в данной статье впервые представляем проспективное диагностическое исследование, изучавшее коллаборацию дерматологов и одобренной к обращению CNN при скрининге меланомы. В данном исследовании дерматологи существенно повышали эффективность диагностики при использовании CNN. Эти результаты показывают, что широкое применение такого подхода совместной работы человека и машины, особенно в неспециализированных медучреждениях, может быть выгодным для врачей и пациентов.

## ИНФОРМАЦИЯ О СТАТЬЕ

Принята к публикации: 5 марта 2023 года

Опубликована онлайн: 3 мая 2023 года

doi:10.1001/jamadermatol.2023.0905

**Вклад авторов:** Dr Winkler и Prof Haenssle имели полный доступ ко всем данным исследования и принимают на себя ответственность за целостность данных и точность их анализа.

**Концепция и дизайн:** Winkler, Kommoss, Enk, Haenssle.

**Сбор, анализ или интерпретация данных:** Winkler, Blum, Toberer, Rosenberger, Haenssle.

**Написание проекта текста данной статьи:** Winkler, Haenssle.

**Критическая проверка текста статьи на важный интеллектуальный контент:** все авторы.

**Статистический анализ:** Winkler, Rosenberger, Haenssle.

**Административная, техническая или материальная поддержка:** Toberer, Haenssle.



**Руководство Supervision:** Enk, Haenssle.

**Конфликт интересов.** Dr Winkler сообщил о персональных выплатах от FotoFinder Systems, Amgen, Bristol Myers Squibb, MSD, Philochem и Roche вне этой работы.

Prof Blum сообщил о персональных выплатах, получаемых от HEINE Optotechnik и FotoFinder Systems.

Prof Haenssle сообщил о персональных выплатах, получаемых от SciBase, FotoFinder Systems, HEINE Optotechnik и Magnosco за пределами данной работы. О раскрытии иной информации не сообщалось.

**Дополнительный вклад:** Мы хотим поблагодарить и выразить признательность всем дерматологам, чье участие не оплачивалось, включая Ines Bertlich, Olivia Bochnig, Therezia Bokor-Billmann, Jana Burghaus-Chang, Christiane Cussigh, Carmen Dietrich, Lara-Elena Hakim-Meibodi, Susanne Hanner, Julia Hartmann, Katharina Kälber, Aric Keller, Kai Linse, Pawel Majenka, Silvia Mihalceanu, Christina Naoum, Alexander Wald, Priscila Wölbing и других, кто остался не названным, за активное и добровольное участие в вовлечении пациентов в исследование.

**Заявление о распространении данных:** см. Приложение 2.

## СПИСОК ЛИТЕРАТУРЫ:

1. Arnold M, Singh D, Laversanne M, et al. Global burden of cutaneous melanoma in 2020 and projections to 2040. *JAMA Dermatol*. 2022;158(5):495-503. doi:10.1001/jamadermatol.2022.0160
2. Barreiro-Capurro A, Andrés-Lencina JJ, Podlipnik S, et al. Differences in cutaneous melanoma survival between the 7th and 8th edition of the American Joint Committee on Cancer (AJCC): a multicentric population-based study. *Eur J Cancer*. 2021;145:29-37. doi:10.1016/j.ejca.2020.11.036
3. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542(7639):115-118. doi:10.1038/nature21056
4. Haenssle HA, Fink C, Schneiderbauer R, et al. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Ann Oncol*. 2018;29(8):1836-1842. doi:10.1093/annonc/mdy166
5. Haenssle HA, Fink C, Toberer F, et al. Man against machine reloaded: performance of a market-approved convolutional neural network in classifying a broad spectrum of skin lesions in comparison with 96 dermatologists working under less artificial conditions. *Ann Oncol*. 2020;31(1):137-143. doi:10.1016/j.annonc.2019.10.013
6. Tschandl P, Codella N, Akay BN, et al. Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study. *Lancet Oncol*. 2019;20(7):938-947. doi:10.1016/S1470-2045(19)30333-X
7. Haggemüller S, Maron RC, Hekler A, et al. Skin cancer classification via convolutional neural networks: systematic review of studies involving human experts. *Eur J Cancer*. 2021;156:202-216. doi:10.1016/j.ejca.2021.06.049
8. Winkler JK, Sies K, Fink C, et al. Association between different scale bars in dermoscopic images and diagnostic performance of a market-approved deep learning convolutional neural network for melanoma recognition. *Eur J Cancer*. 2021;145:146-154. doi:10.1016/j.ejca.2020.12.010



9. Winkler JK, Fink C, Toberer F, et al. Association between surgical skin markings in dermoscopic images and diagnostic performance of a deep learning convolutional neural network for melanoma recognition. *JAMA Dermatol*. 2019;155(10):1135-1141. doi:10.1001/jamadermatol.2019.1735
10. Winkler JK, Sies K, Fink C, et al. Melanoma recognition by a deep learning convolutional neural network—performance in different melanoma subtypes and localisations. *Eur J Cancer*. 2020;127:21-29. doi:10.1016/j.ejca.2019.11.020
11. Tschandl P, Rinner C, Apalla Z, et al. Human-computer collaboration for skin cancer recognition. *Nat Med*. 2020;26(8):1229-1234. doi:10.1038/s41591-020-0942-0
12. Hekler A, Utikal JS, Enk AH, et al; Collaborators. Superior skin cancer classification by the combination of human and artificial intelligence. *Eur J Cancer*. 2019;120:114-121. doi:10.1016/j.ejca.2019.07.019
13. Montague E. Validation of a trust in medical technology instrument. *Appl Ergon*. 2010;41(6):812-821. doi:10.1016/j.apergo.2010.01.009
14. Fink C, Uhlmann L, Hofmann M, et al. Patient acceptance and trust in automated computer-assisted diagnosis of melanoma with dermatofluoroscopy. *J Dtsch Dermatol Ges*. 2018;16(7):854-859. doi:10.1111/ddg.13562
15. Vergara IA, Norambuena T, Ferrada E, Slater AW, Melo F. StAR: a simple tool for the statistical comparison of ROC curves. *BMC Bioinformatics*. 2008;9(1):265. doi:10.1186/1471-2105-9-265
16. Garg AX, Adhikari NK, McDonald H, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. *JAMA*. 2005;293(10):1223-1238. doi:10.1001/jama.293.10.1223
17. Codella NC, Lin C-C, Halpern A, Hind M, Feris R, Smith JR. Collaborative Human-AI (CHAI): evidence-based interpretable melanoma classification in dermoscopic images. In: Stoyanov D, Taylor Z, Kia SM, et al, eds. *Understanding and Interpreting Machine Learning in Medical Image Computing Applications*. Springer; 2018:97-105.
18. Bien N, Rajpurkar P, Ball RL, et al. Deep-learning-assisted diagnosis for knee magnetic resonance imaging: development and retrospective validation of MRNet. *PLoS Med*. 2018;15(11): e1002699. doi:10.1371/journal.pmed.1002699
19. Winkler JK, Tschandl P, Toberer F, et al. Monitoring patients at risk for melanoma: may convolutional neural networks replace the strategy of sequential digital dermoscopy? *Eur J Cancer*. 2022;160:180-188. doi:10.1016/j.ejca.2021.10.030
20. Tschandl P, Hofmann L, Fink C, Kittler H, Haenssle HA. Melanomas vs nevi in high-risk patients under long-term monitoring with digital dermatoscopy: do melanomas and nevi already differ at baseline? *J Eur Acad Dermatol Venereol*. 2017;31(6):972-977. doi:10.1111/jdv.14065
21. Maron RC, Utikal JS, Hekler A, et al. Artificial intelligence and its effect on dermatologists' accuracy in dermoscopic melanoma image classification: web-based survey study. *J Med Internet Res*. 2020;22(9):e18091. doi:10.2196/18091
22. Mobiny A, Singh A, Van Nguyen H. Risk-aware machine learning classifier for skin lesion diagnosis. *J Clin Med*. 2019;8(8):1241. doi:10.3390/jcm8081241
23. Polesie S, Gillstedt M, Kittler H, et al. Attitudes towards artificial intelligence within dermatology: an international online survey. *Br J Dermatol*. 2020; 183(1):159-161. doi:10.1111/bjd.18875
24. Jutzi TB, Kriehoff-Henning EI, Holland-Letz T, et al. Artificial intelligence in skin cancer diagnostics: the patients' perspective. *Front Med (Lausanne)*. 2020;7:233. doi:10.3389/fmed.2020.00233
25. Nelson CA, Pérez-Chada LM, Creadore A, et al. Patient perspectives on the use of artificial intelligence for skin cancer screening: a qualitative study. *JAMA Dermatol*. 2020;156(5):501-512. doi:10.1001/jamadermatol.2019.5014
26. Goyal M, Knackstedt T, Yan S, Hassanpour S. Artificial intelligence-based image classification methods for diagnosis of skin cancer: challenges and opportunities. *Comput Biol Med*. 2020;127:104065. doi:10.1016/j.combiomed.2020.10406

## Приложение 1.

## Дополнительный онлайн-контент

Winkler JK, Blum A, Kommos K и соавторы «Анализ диагностики, проводимой дерматологами с использованием сверточных нейронных сетей, в проспективном клиническом исследовании взаимодействия человека и машины». JAMA Dermatol. Published online May 3, 2023. doi:10.1001/jamadermatol.2023.0905

**Электронная таблица 1.** Процент корректных диагнозов, поставленных дерматологами самостоятельно и при использовании CNN-анализа, в зависимости от опыта дерматолога

**Электронный рисунок 1.** Схема исследования

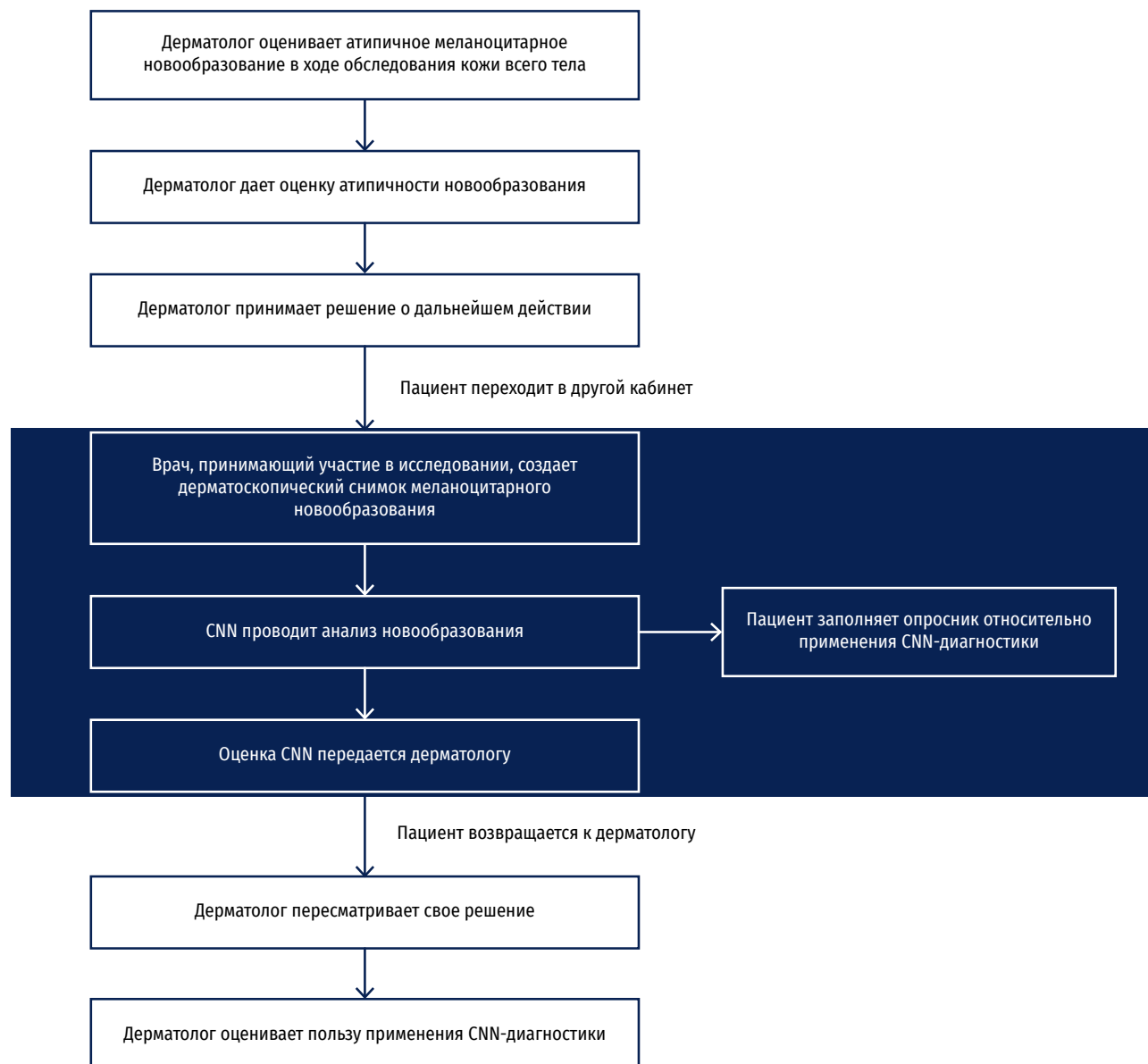
**Электронный рисунок 2.** Анализ 152 заполненных пациентами опросников

Эти дополнительные материалы были переданы авторами, чтобы предоставить читателям больше информации об их работе.

**Дополнительная таблица 1.** Процент корректных диагнозов, поставленных дерматологами самостоятельно и при использовании CNN-анализа, в зависимости от опыта дерматолога

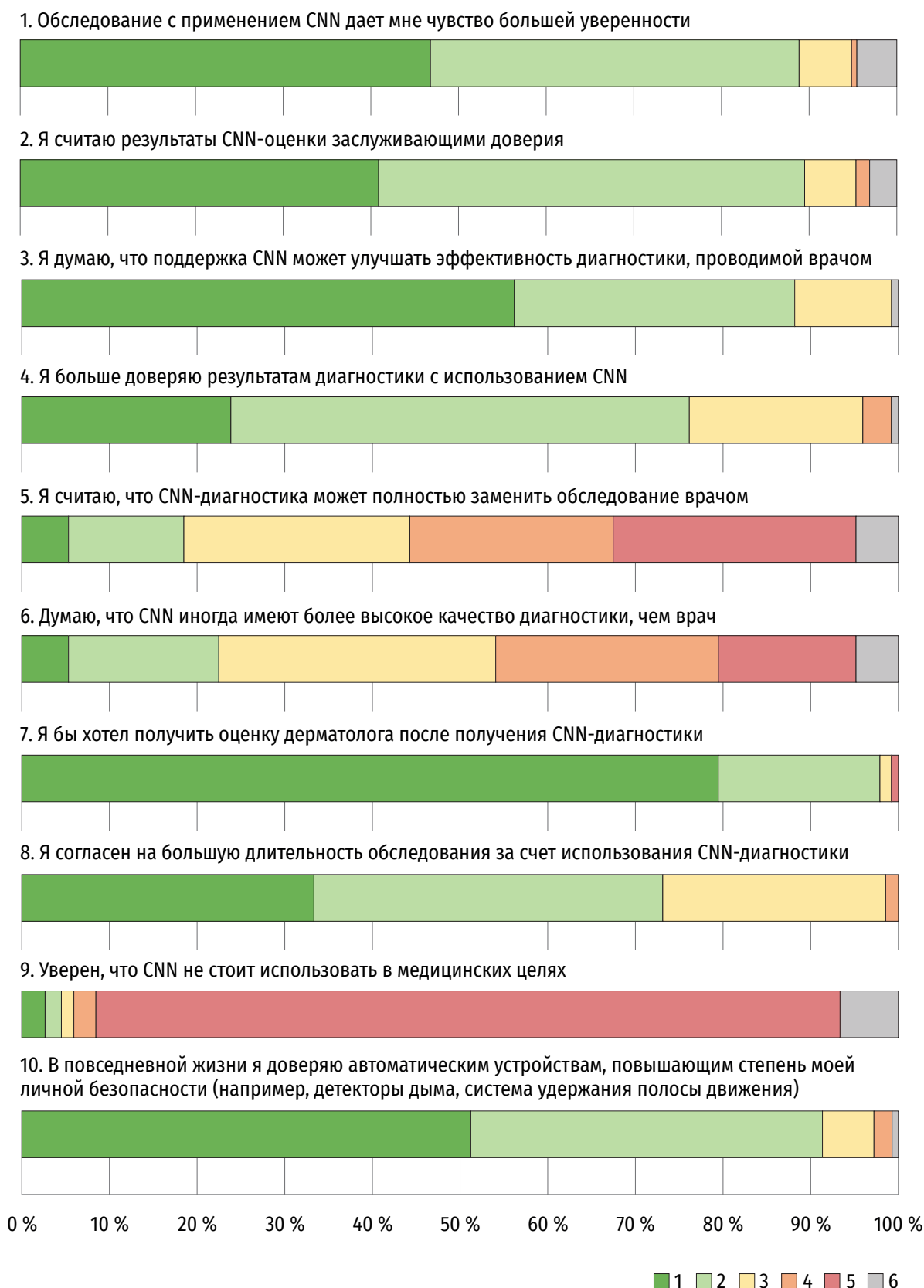
Опыт дерматолога в дерматоскопии (гг)	< 2	2–5	> 5
Количество оцененных новообразований (n = 228)	78	96	54
Только дерматологи: процент корректных диагнозов	70,5 %	77,1 %	74,1 %
Дерматологи + CNN <sup>^</sup> процент корректных диагнозов	87,2 %	91,7 %	75,9 %
P-значение	< 0,01*	< 0,01*	1,0*

\* Использована поправка Бонферрони

**Дополнительный рисунок 1. Схема исследования**

**Дополнительный рисунок 2.** Анализ 152 заполненных пациентами опросников

Степень согласия в отношении 10 пунктов опросника измеряли по 6-балльной визуальной аналоговой шкале (1 = очень высокая степень согласия, 2 = высокая степень согласия, 3 = умеренная степень согласия, 4 = низкая степень согласия, 5 = несогласие, 6 = решение не принято).



**Приложение 2.****Заявление о распространении данных**

Winkler. «Анализ диагностики, проводимой дерматологами с использованием сверточных нейронных сетей, в проспективном клиническом исследовании взаимодействия человека и машины». JAMA Dermatol. Published May 03, 2023. doi:10.1001/jamadermatol.2023.0905

**Данные**

Доступные данные: Нет