

КЛИНИЧЕСКАЯ МЕДИЦИНА**КОЖНЫЕ И ВЕНЕРИЧЕСКИЕ БОЛЕЗНИ**

DOI - 10.32743/UniMed.2022.85.2.12993

**ДИНАМИКА ЧУВСТВИТЕЛЬНОСТИ И СПЕЦИФИЧНОСТИ НЕЙРОСЕТИ
НА ФОНЕ МОДЕРНИЗАЦИИ АЛГОРИТМОВ РАСПОЗНАВАНИЯ КОЖНОЙ ПАТОЛОГИИ***Соколов Кирилл Анатольевич**CEO «Skinive», Skinive Holding BV,
Нидерланды, Амстердам**Шпудейко Виктор Александрович**врач онколог-хирург, Пинский онкологический диспансер,
Республика Беларусь, Брестская обл., г. Пинск,
медицинский эксперт Skinive. ООО «Вайс ЭйАй»,
Республика Беларусь, г. Минск
E-mail: v.shpudeiko@gmail.com***DYNAMICS OF SENSITIVITY AND SPECIFICITY OF A NEURAL NETWORK DURING
THE MODERNIZATION OF THE ALGORITHMS DESIGNED
FOR RECOGNIZING SKIN PATHOLOGY***Sokolov Kirill**CEO «Skinive». Skinive Holding BV,
The Netherlands, Amsterdam**Shpudeiko Viktor**Oncologist. Oncological Health Center,
Belarus, Pinsk
Medical expert Skinive. «Wise AI» LLC,
Belarus, Minsk***АННОТАЦИЯ**

Проведено исследование точности алгоритмов искусственного интеллекта, предназначенных для распознавания кожной патологии. В исследовании применены валидационные датасеты, включающие фотографии доброкачественных и злокачественных новообразований кожи, акне и ВПЧ-патологии кожи. Изображения из валидационных датасетов никогда не использовались для обучения исследуемой нейросети. Полученные результаты чувствительности и специфичности нейросети свидетельствуют о высокой точности алгоритма, которая, согласно литературным данным, сопоставима с точностью врачей-специалистов.

ABSTRACT

A study was made of the accuracy of artificial intelligence algorithms designed to recognize skin pathology. The study used validation datasets, including photographs of benign and malignant skin tumors, acne, and HPV skin pathologies. Images from validation datasets have never been used to train the neural network under study. The obtained results of the sensitivity and specificity of the neural network indicate the high accuracy of the algorithm, which, according to the literature, is comparable to the accuracy of medical specialists.

Ключевые слова: нейронная сеть, искусственный интеллект, машинное обучение, обнаружение кожной патологии, кожная патология.

Keywords: Neural Network, Artificial Intelligence, Machine Learning, Skin Pathology Detection, Skin Diseases.

1. Введение

В связи с более частым появлением доброкачественных новообразований, предраковых состояний, ростом количества злокачественных новообразований и неопухолевых поражений кожи соответственно увеличивается потребность обращений как к врачам общей практики, так и к узким специалистам. Обращение за медицинской помощью в поздних случаях может привести к негативным последствиям в виде прогрессирования опухолевого процесса, более низкой эффективности лечения и плохому прогнозу.

В то же время, меры, направленные на раннее выявление патологии кожи, имеют ряд проблем: 1) чрезмерная бдительность, которая часто переходит в мнительность; 2) длительное ожидание консультации из-за большого количества необоснованных консультаций; 3) неоправданное увеличение нагрузки на врача и др. При этом, статистика говорит о значимом сокращении как коечного фонда по профилю дерматологии, так и уменьшению числа физических лиц врачей-дерматовенерологов в РФ [1]. Сложившаяся ситуация не может не повлиять на ухудшение качества и доступности медицинской помощи для населения.

Решение этих проблем стало возможным благодаря появлению и внедрению в медицинскую практику технологий машинного обучения. Особое значение в развитии такой автоматизации имеет разработка мобильных приложений.

Потенциал применения мобильных приложений, работающих на основе алгоритмов машинного обучения для выявления кожных заболеваний, становится особенно актуальным в периоды неблагоприятных эпидемиологических обстановок (например, ситуация с COVID-19, как это отражено в последних публикациях [2]), когда особенную актуальность приобретает дистанционное оказание медицинских услуг.

Возможности глубокого машинного обучения в проведении дифференциальной диагностики кожных заболеваний воодушевляет и показывает свой потенциал в оказании помощи клиницистам в их рутинной практике [3]. Например, такие решения могут помочь в случаях сортировки для определения приоритетности клинической помощи или помочь не-дерматологам более точно начать дерматологическую помощь и потенциально улучшить доступ к получению необходимой медицинской помощи.

Точность диагностики заболеваний кожи среди врачей общей практики (ВОП) и дерматологами существенно отличается. Например, в исследовании показано, что точность распознавания меланомы среди ВОП составляет всего 0,49-0,80, в то время как среди дерматологов этот показатель достигает 0,85-0,89 [4].

В другом исследовании авторы сделали попытку сравнить точность распознавания кожной патологии между алгоритмом машинного обучения и медицинскими специалистами. Результаты показали, что нейронная сеть достигает диагностической точности 90% (точность для первых трех результатов, выданных нейросетью) [5].

Эти исследования указывают на то, что точность алгоритмов машинного обучения может быть сопоставима с показателями точности дерматологов или даже значительно выше [6, 7]. Также, точность таких алгоритмов значительно выше, чем у врачей общей практики.

Несмотря на перечисленные преимущества, применение программных продуктов для решения таких ответственных задач, как раннее выявление онкопатологии кожи, подвергается различной критике. Основным аргументом критиков является сомнительная чувствительность и специфичность, а также отсутствие у производителей единого подхода к испытаниям, необходимого для объективной оценки.

Внедрение подобных решений в клиническую практику возможно только при наличии качественной, хорошо обученной нейросети. Это подразумевает целый ряд мероприятий, включающий: программное улучшение нейросети, работу над созданием и усовершенствованием тренировочных и валидационных датасетов, регулярные переобучения нейросети, улучшение внешней оболочки для удобного использования мобильного приложения конечным пользователем и другие.

В качестве примера работы над усовершенствованием сети и полученными результатами может быть нейросеть Skinive.

2. Материалы и методы

2.1. Характеристика нейросети Skinive

Алгоритм Skinive - это модели предобученных полносверточных дерматологических нейросетей и прикладное ПО нейросетевого окружения, которые расположены в защищенном облаке и интегрируются с приложениями (веб, мобильные и иное ПО) при помощи API-сервиса.

С момента своего создания и до мая 2020 года в архитектуре нейросети Skinive реализованы такие слои для повышения точности распознавания кожной патологии, как drop out layer, local response normalization layer и т.д.

2.2. Оптимизация работы нейросети Skinive в 2020-2021 годах

За период с мая 2020 года по август 2021 года нейросеть Skinive претерпела целый ряд как технических, так и клинических усовершенствований.

Технические улучшения за период 2020-2021 включали внедрение новых слоев распознавания, использован улучшенный алгоритм, выполнена задача по переходу на большее разрешение, внедрена функция аугментации (усовершенствование распознавания путем генерирования новых данных на основе имеющихся) и другое.

Клиническое усовершенствование базы данных (тренировочного датасета) с медицинскими изображениями включало: 1) количественные улучшения; 2) качественные улучшения; 3) собственные методики обработки медицинских изображений перед отправкой на обучение нейросети.

Под количественным улучшением датасета подразумевается увеличение как количества распознаваемых патологий, так и общее количество фотографий

для обучения нейросети. На август 2021 года число распознаваемых патологий увеличилось до 45 (в 2020 – 31 патология). Количество фотографий для обучения нейросети (объем тренировочного датасета) в 2021 году увеличилось до 115 000 изображений (в 2020 – 64 000). Всего к августу 2021 нейросетью Skinive проанализировано более 250 000 изображений (на май 2020 года всего проанализировано 150 000 изображений).

Особое внимание команда Skinive уделяла качеству тренировочных датасетов. Основным принципом работы в 2020-2021 годах было не стремление за увеличением количества размеченных фотографий, а качественное улучшение данных. Все данные, которые использовались в обучении нейросети, прошли многоуровневую проверку на соответствие качества изображений и однозначную трактовку медицинскими экспертами.

Собственная методика обработки медицинских изображений перед отправкой на обучение нейросети включала: 1) предварительное машинное распределение массива данных обученной ранее нейросетью; 2) ручная коррекция ошибочного распределения данных медицинским экспертом; 3) привлечение сторонних врачей-специалистов из нашей когорты доверенных врачей для верификации сложных случаев кожной патологии на фотографии; 4) использование специальных программных решений для разметки данных при множественной патологии на одном снимке или при сложной анатомической локализации.

Переобучение нейросети Skinive было инициировано в июле 2021 года и завершено в августе 2021 года. Сразу после этого, в августе 2021 проведена масштабная работа, которые позволила получить новые данные по результатам точности распознавания кожной патологии нейросетью Skinive 2021.

2.3. Формирование валидационного датасета

Валидационные датасеты были составлены в 2020 году. Эти наборы данных использовались при анализе точности как в 2020, так и в 2021 году, что позволило определить динамику роста показателей чувствительности и специфичности.

Для определения чувствительности алгоритма сформированы валидационные датасеты из пользовательской базы данных Skinive, состоящие из 285 изображений в каждом из классов: 1) Acne (включая Acne Vulgaris, Acne Pustular, Acne Cystic, Comedone Closed, Comedone Open, Miliun, Rosacea); 2) Viral diseases (включая Papilloma, Wart Vulgaris, Wart Plane, Wart Plantar, Molluscum Contagiosum); 3) Precancer + cancer (включая Actinic Keratosis, Bowen, Basal Cell Carcinoma, Squamous Cell Carcinoma, Melanoma).

Для определения специфичности алгоритма сформирован отдельный набор валидационных данных (класс Benign), состоящий из 6000 изображений с доброкачественными новообразованиями кожи (включая Benign Nevus, Papilloma Nevus, Hemangioma, Dermatofibroma, Halo Nevus, Spitz Nevus, Pyogenic Granuloma).

2.4. Статистическая обработка данных

Для статистической обработки данных использовалось программное обеспечение QlikView версии 11.2 с дополнительной статистической обработкой полученных данных в режиме онлайн по url-адресу: <https://statpages.info/ctab2x2.html/>. Статистическая значимость полученных результатов оценивалась с помощью доверительных интервалов. Уровень значимости: $p \leq 0,05$.

3. Результаты

После отправки сформированных валидационных датасетов на анализ алгоритму Skinive, получены значения чувствительности и специфичности по каждому из анализируемых классов (Таблица 1).

Таблица 1.

Результаты точности нейросети Skinive в 2020 и 2021 году, %

Классы кожных заболеваний	Чувствительность (CI 95%)		Специфичность (CI 95%)	
	2020	2021	2020	2021
Acne	86,3 (82,8-89,2)	91,2 (88,7-93,0)	99,3 (99,2-99,4)	99,8 (99,7-99,9)
Viral diseases	79,6 (75,1-83,6)	78,6 (74,5-82,2)	97,7 (97,6-97,9)	99,1 (98,9-99,2)
Precancer + cancer	89,1(85,0-92,3)	95,4 (92,4-97,4)	95,6 (95,4-95,8)	98,5 (98,4-98,6)
Benign	93,5 (93,2-93,6)	97,1 (96,9-97,2)	95,3 (93,7-96,6)	97,9 (97,2-98,5)

Для визуального понимания ошибок алгоритма и выявления тенденции неправильных распознаваний нейросетью, составлена матрица ошибок (Таблица 2),

которая отражает распределение верных ответов (по чувствительности) и ошибок между классами нозологий.

Таблица 2.

Матрица ошибок по группам кожных заболеваний, %

Классы кожных заболеваний	Acne		Viral diseases		Precancer + cancer		Benign		**Другая патология
	2020	2021	2020	2021	2020	2021	2020	2021	2021
Acne	86,3*	91,2*	4,9	0,7	6,7	0,4	2,1	1,1	6,6
Viral diseases	1,5	0,4	79,6*	78,6*	9,1	1,1	9,8	7,0	12,9
Precancer + cancer	7,4	0	1,4	0,4	89,1*	95,4*	2,1	2,8	1,4
Benign	0,3	0,1	2,2	1,1	4,0	1,7	93,5*	97,1*	0

Примечание: *) правильное распознавание уровня риска; **) кожная патология, распознавание которой начато только в 2021 году

В целом по алгоритму чувствительность нейросети Skinive в 2020 году составила 95,3% (CI 95% 93,7-96,6%) при специфичности 93,5% (CI 95% 93,2-93,6%). В 2021 году - 97,9% (CI 95% 97,2-98,5%) и 97,1% (CI 95% 96,9-97,2%) соответственно.

4. Обсуждение

Благодаря целому комплексу проведенных работ над нейросетью Skinive, удалось получить рост показателей чувствительности и специфичности практически по всем исследуемым классам новообразований и заболеваний кожи.

Анализ доверительных интервалов позволяет установить, что чувствительность алгоритма Skinive при распознавании предопухолевых состояний и злокачественных новообразований имеет достоверный позитивный рост за период 2020-2021 год. Также статистически достоверно увеличилась специфичность распознавания доброкачественных новообразований.

Полученные результаты позволяют утверждать, что комплекс мер, направленных на усовершенствование нейросети Skinive, нашел особое отражение на улучшении точности работы алгоритма с онкологической патологией кожи. Очевидно, что именно эта патология имеет наибольшую актуальность и значимость с точки зрения сохранения жизни и здоровья населения.

Детальный анализ матрицы ошибок показывает, среди каких заболеваний и с какой частотой возникают наибольшие трудности у алгоритма Skinive при дифференциальной диагностике новообразований и заболеваний кожи. Эти значения можно считать ключевым результатом для планирования дальнейшей работы по усовершенствованию нейросети Skinive и увеличению точности алгоритма по отдельным нозологиям.

Из матрицы ошибок очевидно, что наибольший процент погрешностей при распознавании кожной патологии отмечается у пары «Кожная патология, вызванная вирусом папилломы человека» - «Доброкачественные новообразования»: 9,8% и 7,0% за 2020 и 2021 год соответственно.

Анализ показал, что чаще всего нейросетью неверно классифицируются вирусные папилломы, вызванные ВПЧ. Алгоритм ошибочно отнес вирусные папилломы к классу доброкачественных новообразований из-за большого визуального сходства вирусных папиллом с папилломатозным невусом (группа доброкачественных новообразований).

Действительно, данные пары новообразований (вирусная папиллома - папилломатозный невус) часто выглядят очень похоже (Рисунок 1) и могут трактоваться неоднозначно.

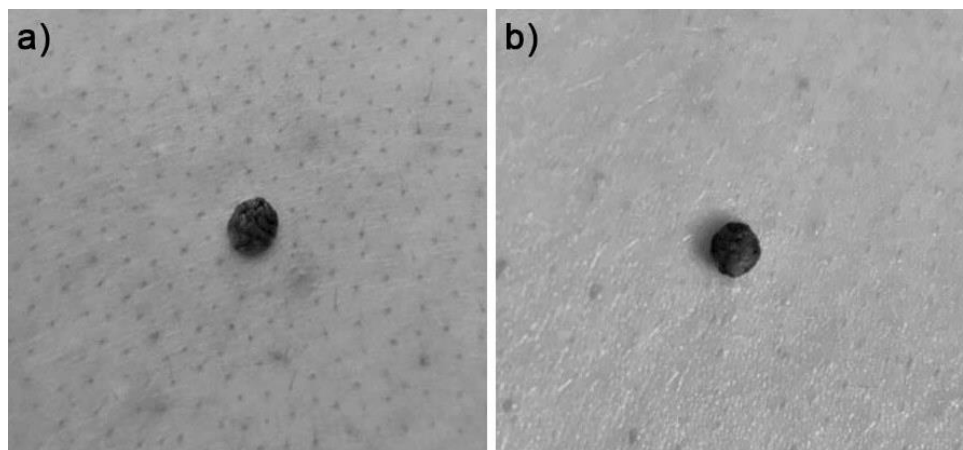


Рисунок 1. Сравнение двух видов доброкачественных новообразований: а) папилломатозный невус, б) вирусная папиллома

Вторая по величине частота встречаемости ошибок отмечается в паре «Предопухолевые состояния и злокачественные новообразования» и «Доброкачественные новообразования» и составила 2,8% в 2021 году. Анализ ошибок показал, что нейросеть ошибочно относилась самые ранние формы меланом

к доброкачественным невусам. При визуальном сравнении (Рисунок 2) макрофотографии пигментного невуса и ранней меланомы действительно могут возникнуть трудности при дифференциальной диагностике таких случаев.

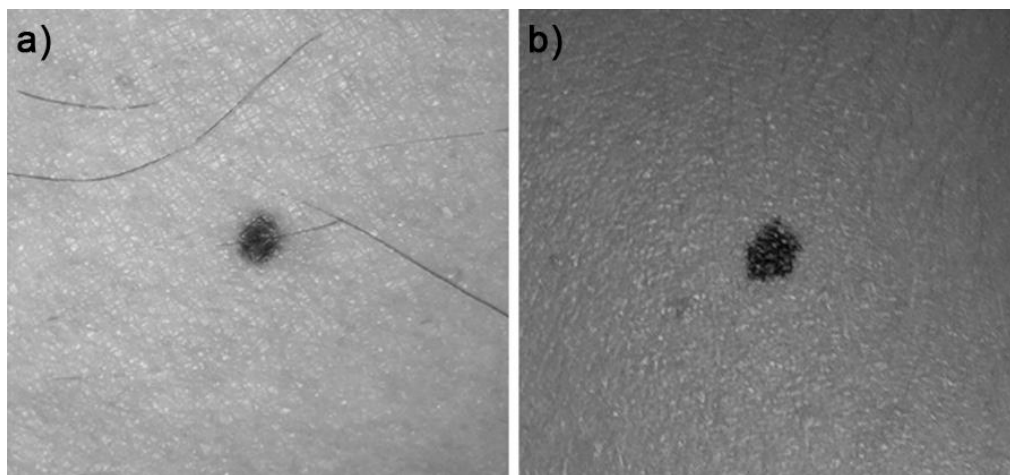


Рисунок 2. Сравнение фотографии пигментного невуса и ранней меланомы: а) пигментный невус, б) меланома (патогистологически pT1, 0,7 мм)

Эти клинические ситуации требуют проведения более точной дифференциальной диагностики с включением дерматоскопии или биопсии с гистологическим исследованием.

В то же время, при сравнении матрицы ошибок за 2020 и 2021 год (Таблица 3) становится очевидной и глобальная тенденция к уменьшению количества неправильных интерпретаций при распознавании кожной патологии.

Таблица 3.

Сравнительный анализ матрицы ошибок за 2020 и 2021 год, %

Классы кожных заболеваний	Общее количество ошибочных распознаваний (CI 95%)	
	2020	2021
Acne	13,7 (10,6-16,8)	8,8 (6,5-10,6)
Viral diseases	20,4 (16,3-24,8)	21,4 (17,6-25,2)
Precancer + cancer	10,9 (7,7-14,9)	4,6 (2,6-7,5)
Benign	6,5 (6,3-6,7)	2,9 (2,7-3,1)

Кроме того, это сравнение позволяет отметить динамику и корректность вектора проводимых работ над усовершенствованием нейросети.

Наиболее позитивные изменения произошли в группе «Предопухолевые состояния и злокачественные новообразования». Суммарное количество ошибок в 2021 году составило 4,6% (CI 95% 2,6-7,5%), что статистически достоверно меньше аналогичного значения, полученного в 2020 году (10,9% (CI 95% 7,7-14,9%)). Не менее значимо и снижение количества ошибок при распознавании доброкачественных новообразований. Так, в 2020 году уровень ошибочного распознавания в этой группе составил 13,7% (CI 95% 10,6-16,8%), в 2021 – 8,8% (CI 95% 6,5-10,6%), что также достоверно ниже предыдущего года.

При анализе общих показателей чувствительности и специфичности алгоритма Skinive сохраняется общая тенденция статистически достоверного роста показателей. Учитывая актуальность своевременного и достоверного распознавания онкологических рисков - такая тенденция указывает на реальную возможность применения нейросети Skinive в клинической практике.

5. Выводы

Результаты чувствительности и специфичности нейросети Skinive свидетельствуют о высокой точности алгоритма в определении различных новообразований и заболеваний кожи и составили более 90% практически по всем анализируемым классам кожной патологии. Получение этих результатов стало возможным благодаря постоянной работе,

направленной на всестороннее усовершенствование нейросети.

К ключевым моментам улучшения нейронных сетей, предназначенных для проведения классификации дерматологических и онкологических заболеваний кожи, следует относить постоянную и кропотливую работу над качественным составом тренировочных датасетов, количественным параметром датасетов, программным усовершенствованием алгоритмов искусственного интеллекта.

Эффект модернизации нейронных сетей Skinive подтверждается достоверной позитивной разницей между значениями чувствительности и специфичности определения кожной патологии, полученными в 2020 и 2021 году. Кроме того, необходимость модернизации алгоритмов искусственного интеллекта подтверждает и существенное снижение доли ошибочных распознаваний кожной патологии.

Полученные результаты позволяют утверждать, что наиболее значимая положительная динамика роста показателей точности отмечалась в классах онкологической патологии кожи, т.е. при распознавании доброкачественных и злокачественных новообразований, предопухоловых состояний. Это, в свою очередь, соответствует современной тенденции, направленной на актуализацию своевременного выявления кожной онкопатологии и, соответственно, сокращению расходов на ее лечение.

Высокие результаты точности в совокупности с такими достижениями, как получение сертификата безопасности ISO 13485, сертификата соответствия европейским стандартам CE-mark и выпуска мобильного приложения “Skinive MD”, доступного для скачивания в AppStore и Google Play и предназначенного для медицинских специалистов, позволило проекту Skinive начать свое успешное внедрение в медицинскую практику.

Список литературы:

1. Кубанов А.А., Богданова Е.В. Итоги деятельности медицинских организаций, оказывающих медицинскую помощь по профилю дерматовенерология, в 2020 году: работа в условиях пандемии. Вестник дерматологии и венерологии. 2021;97(4):08–32. doi: <https://doi.org/10.25208/vdv1261>
2. Chatterjee, P., et al. The 2019 novel coronavirus disease (COVID-19) pandemic: A review of the current evidence. *Indian J Med Res.* 2020 Feb & Mar;151(2 & 3):147-159.
3. Du-Harpur, X., et al. What is AI? Applications of artificial intelligence to dermatology. *Br J Dermatol.* 2020 Sep; 183(3):423-430.
4. Michael DC, Judy W. Agreement between dermatologists and primary care practitioners in the diagnosis of malignant melanoma: review of the literature. *J Cutan Med Surg.* Sep-Oct 2012;16(5):306-10.
5. Liu Y., et al. A deep learning system for differential diagnosis of skin diseases. *Nat Med* 26, 900–908 (2020). <https://doi.org/10.1038/s41591-020-0842-3>.
6. Brinker T.J., et al. Deep neural networks are superior to dermatologists in melanoma image classification. *Eur J Cancer.* 2019 Sep;119:11-17.
7. Udrea A., et al. Accuracy of a smartphone application for triage of skin lesions based on machine learning algorithms. *Journal of the European Academy of Dermatology and Venereology.* March 2020; 34-3:648-655.