DEUTSCHES KREBSFORSCHUNGSZENTRUM IN DER HELMHOLTZ-GEMEINSCHAFT

Forschen für ein Leben ohne Krebs

Studyprotocol

AI-MEL: Image analysis and machine learning for early diagnosis and risk prediction in children, adolescents and young adults

Head of Study

Dr. med. Titus J. Brinker
Nachwuchsgruppe Digitale Biomarker für die Onkologie
Deutsches Krebsforschungszentrum
Im Neuenheimer Feld 223
69120 Heidelberg
+4915175084347
Titus.brinker@nct-heidelberg.de

Other persons involved in the study

[redacted]
PD Dr.. Ines Brecht (UT, Tübingen)
[redacted]
Dr., Daniela Massi UNIFI, Florenz
[redacted]
Dr., Susana Puig (FCRB, Barcelona)

Biometrician

[redacted]

Sponsor

Horizon Europe Programm

Date and Version of Protocol

Version: 2.0
Date: 30.05.2023

Signature of study leader

1. Summary

Melanoma in childhood and adolescence is under-studied, lacking adequate preventive, diagnostic, and therapeutic strategies. The incidence of melanoma is reported to be about 1.5 per million in children under 15 years of age and 15 per million in 15-19 y.o., with increasing incidence in adolescents by 4% annually since 1997[1,2].
As benign melanocytic lesions develop in nearly every child and are present in over 1% at birth, the diagnostic accuracy of melanoma is very low in this age group. A large number of benign lesions are unnecessarily excised (593.8 benign nevi per melanoma under 20 years of age[4]). Furthermore, lower sensitivity than in adults[5] results in late diagnosis with poorer outcomes, when children, adolescents and young adults (CAYA) patients, particularly the children, could show a 90% 10-year survival rate from stage I/II disease. Thus, melanomas in CAYAs represent a medical challenge.
Increasingly, machine learning algorithms are starting to play an important role in melanoma diagnosis. However, these algorithms are based on adult data with a selection bias that could be detrimental and potentially dangerous with respect to both melanoma underdiagnosis and overdiagnosis in CAYA.
Numerous studies have shown that the analysis of medical images through algorithms based on artificial intelligence can provide important information. In oncology, a classic use case is the analysis of pigmented skin lesions by deep learning (DL) algorithms, in particular by Convolutional Neural Networks (CNNs), for early melanoma detection. Currently, available deep learning algorithms have been trained with images from open source and proprietary databases that mostly contain melanomas from patients after the age of 60 (e.g., ISIC archive 4/2022: only 111 of 5734 melanomas from patients up to 30 years of age)[27]. Furthermore, 60% of melanoma in children aged <10 years and 40% of adolescents did not meet the traditional ABCDE clinical criteria of melanoma [5]. Therefore, it is unlikely that the existing diagnostic tools will show optimal performances if challenged with images from adolescent and young adult pigmented skin lesions.
The MELCAYA project "New strategies for the treatment of melanoma in children, adolescents and young adults" thus aims to provide a necessary solution to these problems by collecting data from the CAYA cohort to compensate for its underrepresentation. Within the framework of MELCAYA, we are submitting the sub-project AI-MEL: "Imaging and machine learning for early melanoma diagnosis and risk prediction in children, adolescents and young adults" in this application for approval. The research goal of this project (AI-MEL) is to develop diagnostic assistance tools to distinguish melanomas from nevi or other benign pigmented skin lesions, especially in young patients. One of these tools will be based on macroscopic or dermatoscopic images and will be targeted to skin cancer screening in vulnerable populations. The other tool will be based on histological images and is intended to be used by pathologists on the lesions that are still suspicious of melanoma after dermatologic assessment.
During the development of the DL algorithms, a special focus is placed on explainability and robustness, in order to integrate regulatory standards at an early stage. In the long term, these DL algorithms could improve melanoma early detection in young patients by reducing the

number of unnecessary biopsies and by minimizing the number of melanomas that are overlooked. In addition, we want to create a skin age classifier as a potential determinant of melanoma risk in CAYA
reducing the number of unnecessary biopsies and minimizing the number of missed melanomas.

Study population: Mainly children (up to and including 15 years of age), adolescents (16-20) and young adults (from 21 to 30 years of age). will be included in this study.

Study-related measures.

DL algorithm development is based on macroscopic or dermatoscopic images of melanomas, nevi, or other benign pigmented skin lesions, as well as non-identifying clinical data. The collection of the specimens and the acquisition of the images of the lesions will be performed at the partner clinics (University of Florence (UNIFI-Università degli Studi di Firenze, Public University, Italy), the University of Tübingen (UT) and the Institute for biomedical research in Barcelona (FCRB-Fundació Clínic per a la Recerca Biomèdica)) and their European partner clinics in prospective collection anyway. in clinical routine and therefore does not lead to additional risks and burdens for the patients. and burdens for the patients on the one hand, and on the other hand does not lead to any additional work
clinical staff. The retrospective data are partly provided by the partners themselves or are provided by the partners themselves or collected as part of the Work Package 3 (see overall application in the
Appendix). The collection, storage, and transfer of the data is in compliance with the respective country-specific requirements and EU Regulation 2016/679, has been reviewed and approved by the respective local ethics committees, and is approved and is based on the legal basis "legitimate interests" Article 6.1 (f) GDPR in conjunction with "scientific research" (Article 9.2 (j) GDPR) (cf. Annex ethical and legal aspects clinical protocols, an assurance from all partners working on the MELCAYA project that the processing, communication and transfer of personal data personal data of all participants will be carried out in accordance with Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons).The images will be provided to the German Cancer Research Center (DKFZ) in digital form., so that no physical shipment of human tissue samples is necessary. To develop the planned classifiers, we will build on our existing adult melanoma classifiers for macroscopic/dermatoscopic images and histological images, respectively. This is necessary because the estimated number of cases that can be included in the project is comparatively small due to the rarity of melanoma in childhood, adolescence and young adult is comparatively low. For the fine-tuning and external testing of the dermatoscopic classifier we will rely on images already collected within the consortium and transferred to a research database and supplement them with images from public and intra-group databases.

2. Table of content

3. Introduction/Scientific background

Melanoma in childhood and adolescence is under-studied, lacking adequate preventive, diagnostic, and therapeutic strategies. The incidence of melanoma is reported to be about 1.5 per million in children under 15 years of age and 15 per million in 15-19 y.o., with increasing incidence in adolescents by 4% annually since 1997 [1,2].

As benign melanocytic lesions develop in nearly every child and are present in over 1% at birth, the diagnostic accuracy of melanoma is very low in this age group. A large number of benign lesions are unnecessarily excised (593.8 benign nevi per melanoma under 20 years of age[4]). Furthermore, lower sensitivity than in adults[5] results in late diagnosis with poorer outcomes, when children, adolescents and young adults (CAYA) patients, particularly the children, could show a 90% 10-year survival rate from stage I/II disease. Thus, melanomas in CAYAs represent a medical challenge.

Numerous studies have shown that the analysis of medical images through algorithms based on artificial intelligence can provide important information. In oncology, a classic use case is the analysis of pigmented skin lesions by deep learning (DL) algorithms, in particular by Convolutional Neural Networks (CNNs), for early melanoma detection.

However, the accuracy of DL algorithms is only given on data sets whose distribution corresponds to the images with which the algorithm was trained.

The diagnostic accuracy of the DL algorithms cannot be guaranteed on so-called out-of-distribution datasets, which, for example, originate from other hospitals and were thus generated with different imaging devices, or whose demographic distribution differs from the distribution of the dataset used to develop the model.

Fitting on data sets with a different distribution, in the case of this project fitting on a different age cohort, can be done by means of transfer learning. In this case, the DL-algorithms trained on the data of the original cohort are adapted to the new data set, for example by retraining only the last layers of the neural networks. In the case of very few data in the out-of-distribution domain, Few-Shot Learning approaches can be applied, where technical methods are used for adaptation via only a few new data points.

Deep neural networks as classifiers are essentially black-box systems. Especially in the medical domain, additional methods are therefore essential to explain the classifier decisions. For example, the XAI methods CAM, GradCAM, and Integrated Gradients can be used to visualize examination regions in the image as a heat map12-14. Alternatively, textual explanations can be output, for example by training an additional concept activation vector that can quantify the sensitivity of the classification model with respect to a concept (a well-known example here would be the concept "stripes" in a classifier that is supposed to recognize zebras). In addition, similar and/or counterfactual examples can be identified with content-based image retrieval (CBIR) systems.

In addition, post-hoc XAI methods can be applied to the trained models to detect and avoid model bias during the development phase.

Currently, available DL- algorithms have been trained with images from open source and proprietary databases that mostly contain melanomas from patients after the age of 60 (e.g.

ISIC archive 4/2022: only 111 of 5734 melanomas from patients up to 30 years of age)[27]. Furthermore, 60% of melanoma in children aged <10 years and 40% of adolescents did not meet the traditional ABCDE clinical criteria of melanoma[5].Therefore, it is unlikely that the existing diagnostic tools will show optimal performances if challenged with images of pigmented skin lesions in adolescents and young adults.
Other groups as well as the Brinker group at DKFZ leading this WP have already developed accurate deep learning-based image classifiers for melanomas and other skin lesions. Most of these classifiers were trained on clinical or dermatoscopic images, but first classifiers based on histological images were also generated. However, performance of these classifiers specifically on CAYA lesions or optimization of classifier performance on CAYA lesions has not been reported. Moreover, to our knowledge, specific classifiers for spitzoid lesions have not been generated yet. Classifiers for skin age are advertised by several companies that specialize in skin care products, but these are not open source and not intended as skin cancer risk predictors**.** In one of our previous studies, we have already shown that the prediction of patient age using skin images is feasible by separating patients into two age groups.

4. Objectives of this study (in general)

The research goal of this project (AI-MEL) is to develop supportive diagnostic DL algorithms to distinguish melanoma from nevi or other benign pigmented skin lesions, especially in young patients. The University Hospitals of Tübingen, Florence and Barcelona will contribute the data that will be used to train and test the classifiers. Data from open source databases will also be used. One of these algorithms will be based on macroscopic or dermatoscopic images and will be targeted to skin cancer screening in vulnerable populations. The other tool will be based on histological images and is intended to be used by pathologists on the lesions that are still suspicious of melanoma after dermatologic assessment.
We will also implement explainability methods (XAI) to enable users of such diagnostic assistance tools to comprehend the systems' decisions, avoid biases and increase trust in these applications. In the long term, these DL algorithms could be used to improve melanoma early detection in young patients by reducing the number of unnecessary biopsies on the one hand and by minimizing the number of melanomas that are overlooked. In addition, we aim to develop a skin age classifier that could determine melanoma risk in CAYA patients.

5. Target criteria

Objective	**Target criterium**	**Measures**	**Participations**
Data acquisition	Data are stored in digital form in a database provided by the DKFZ	Collection of dermatoscopic + histological images +	Participating hospitals (FCRB, UT, UNIFI will contribute data.

		associated metadata.	+ Open source data
Preprocessing of data	Data has been checked for quality and sorted out if necessary. Images are divided into patches.	Sorting the data. Only data with unique labels will be considered. Images are divided into patches.	DKFZ
Development of diagnostic and risk prediction classifiers - Diagnostic algorithms to distinguish melanoma from nevi or other benign pigmented skin lesions - Classifiers for spitzoid lesions - Skin age classifier	Classifiers have been trained, tested, validated and optimized by metadata for specific tasks	Data are classified into the training, testing and validation set. CNN and other DL- algorithms are trained and tested with the image data for different tasks (skin age classifier, diagnostic classifier- melanoma vs. nevus and spitzoid lesion classifier).	DKFZ
Implementation of explainability methods	Appropriate explanatory methods were implemented	Different explanatory methods are tested and validated for the different tasks.	DKFZ
Evaluation of methods	Feedback from Dermatologists and Pathologists are present	Dermatologists and Pathologists assess developed methods	DKFZ + Partner Dermatologists/Pathologists

6. Study related Measures AI-MEL

Study-related measures

Data sources
Retrospective data collection from:

- **SCP2 Project-Mannheim**, period 2020-2023: information pamphlets + consent form attached.
- Florence, period 2005-2021: information pamphlets + consent form attached. Passages authorizing use of data for further research purposes have been highlighted and translated in a separate document.
- **Tübingen, STEP registry** period since 2012 continuous; Dermatological Archive, period since 2016: information briefs + consent form attached.
- **Barcelona, Xarxar study**, period 1992-2017: information pamphlets + consent form attached (scripts were updated in 2015). Passages authorizing use of data for further research purposes have been highlighted and translated in a separate document.

Prospective data collection from:

- **Barcelona Biobank:** information pamphlets +consent form attached. Passages approving the use of the data for further research purposes have been highlighted and translated in a separate document.

Data processing, analysis, and collection as well as the target criteria do not differ for retrospective or prospective data.

Methodology:

1. Data Acquisition:

Data collection for the project will cover both retrospective and prospective data. Dermatoscopic and histological images with associated metadata will be acquired from participating clinics in the current project. In addition, macroscopic and histopathologic images collected during the SCP2 project in the Brinker lab and those collected during a previous project by Susana Puig (Barcelona) will be used. These data include data from the CAYA cohort and data of patients over 30, which will be used to build the adult model. We will also obtain histopathological images of the CAYA cohort from the STEP registry of the Institute of Pathology of the University of Tübingen. The Institute of Pathology of the University of Florence provides histopathological tissue images of the CAYA and adult cohorts, and specifically adult spitzoid lesions to create a classifier for spitzoid lesions. Moreover, open source dermatoscopic images from the ISIC archive (e.g. HAM10000 or BCN20000) and a few histological images from the TGCA melanoma cohort from patients below the age of 30 will be collected.
In addition, other images will be prospectively collected and provided by the participating partner hospital in Barcelona. Prospective collection will be centralized throughout the project based on specific patient consents. Macroscopic images are thereby centrally forwarded from

Barcelona. To train the macroscopic skin aging classifier, we will use open source images obtained from healthy patients with or without benign skin lesions including the patient age. These will be tested on skin sections from macroscopic CAYA melanoma images. In addition, we plan to use normal skin images collected prospectively from participating clinics.
In order to achieve a high data quality, we limit ourselves in this project to images whose labels are provided with a highly reliable label for these images. All data will be provided in digital forme1. Data augmentation by rotating and flipping images as well as other augmentations as altering the zoom or brightness will also be tested. Tumor regions will be annotated manually. For hand-crafted features, the pathology program QuPath will be trained to detect specific cells, in particular melanocytic cells and lymphocytes.

2. Development of classifiers:
In order to obtain an accurate classifier for the rare occurrences of pigmented skin lesions in young patients, the existing model for adults will be adapted to the new cases by fine-tuning, an established method to make use of a neural network trained on abundant data to a related task with limited data availability. This is necessary because the estimated number of cases that can be included in the project is rather small due to the rarity of childhood, adolescent, and young adult melanoma.
We will optimize the fine-tuning by comparatively "freezing" most of the parameters learned in the first adult model and retraining only some of the parameters in the final layers of the neural network or, alternatively, retraining all of the parameters, limiting the variances of the parameters. Also, in the first case, the ideal number of "frozen" layers can be determined using hyperparameter tuning. Similarly, we will use the higher absolute number of available spitzoid lesions in older adults to adapt our existing classifier or to train a spitzoid lesion-specific classifier on the collected adult data. This classifier can then in turn be fine-tuned using spitzoid lesions from children and AYAs from participating clinics.
We will mostly rely on lesion-containing images to train the skin age classifier. If necessary, we will remove the lesions using image segmentation or only use patches of normal skin derived from those images. However, we will at least test the classifier on images with and without benign lesions collected from the participating clinics. Importantly, we will only use images from healthy (non-melanoma) patients to avoid biases like the sun damage that is more prevalent in melanoma patients. Importantly, we must take different Fitzpatrick skin types into account, since melanomas will be more prevalent in the light-skinned population. It may be necessary to create separate classifiers for different skin types. Once the classifier is established, we will test it against skin images from young melanoma patients to find out whether age is systematically overestimated in these patients, indicating accelerated skin aging as a risk factor and possible biomarker for melanoma development.
CNNs or other deep learning architectures will be trained by supervised, semi-supervised and unsupervised training to generate binary melanoma/nevus classifiers. For that purpose, data will be split into training/validation and test sets, including at least one external test set per classifier, i. e. a test set comprised of out-of-distribution data from another cohort/clinic. The whole slide images will have to be split into tiles (=patches) because of image size. For whole

slide image classification, tile-level and slide-level approaches will be tested. The multiclass dermatoscopic skin age classifier will be trained using images from patients of all age groups. Patients with confirmed melanoma will be excluded, since it is expected that many of these patients have experienced a high amount of UV damage, which will have aged their skin more strongly than usual

In addition to an CNN-based image classification algorithm, we will also investigate whether quantifiable hand-crafted features can be identified together with dermatologists and pathologists that might be helpful to distinguish melanomas from nevi in young patients and younger from older skin, e.g. such as wrinkles or pigmented spots for the dermatoscopic images and cell densities or features connected to solar elastosis on histological slides.

3. Addition of explainable artificial intelligence (XAI) methods to the classifiers:

*D*eep neural networks as classifiers are essentially black box systems. Particularly in the medical domain, additional methods to explain the classifier decisions are thus indispensable. For instance, regions of interest can be visualized as a heatmap in the picture using CAM, GradCAM, and Integrated Gradients[12-14]. Alternatively, textual explanations can be provided, for example by training an additional concept vector identifyer[15] that can quantify the sensitivity of the classification model with respect to a concept (a well-known example here would be the concept "stripes" in a classifier that is supposed to recognize zebras) Moreover, similar and/or counter-factual examples can be identified with content-based image retrieval (CBIR) systems. In addition, post-hoc XAI methods will be applied to the trained models to detect and avoid model biases during the development phase. If the generated explanations indicate bias, for instance if the models use artifacts such as rulers on dermoscopic images for shortcut learning, the data will be balanced towards equal artefact distribution per class. Alternatively, the artifacts will be removed from the data or the classifier's objective function for training is adjusted accordingly.

The selection of the explainability method will be based on the outcome of current studies within the KTI project in the Brinker group, where the influence of different XAI methods on AI-supported classification accuracy as well as acceptance of such systems by the dermatologists and pathologists is currently investigated.The XAI methods selected for dermatologists and pathologists will be implemented on top of the final classifiers. They will be tested for fidelity and robustness before final approval. A suitable user interface will be chosen and implemented to allow the physicians to evaluate the CNN prediction, the explanation and uncertainty estimate.

4. Evaluation of the results in cooperation with dermatologists and pathologists:

By help of a questionnaire including the respective images, dermatologists and pathologists, respectively, will be asked to classify the lesions themselves. Subsequently, they will be presented with the results of the classifier(s) including the results of the XAI. The human experts will then be asked whether these results would influence their diagnostic decision and why, or why not. Thereby, the potential impact of an IA-based diagnostic assistance system for

the CAYA cohort in routine dermatological and pathological practice will be determined. Using the results of the XAI, human experts will also try to identify recognizable features that the algorithm is likely to use, and evaluate whether these features could also be used by the experts themselves without the aid of the assistance system (train the dermatologist/pathologist approach).

5. Trustworthiness and Technical robustness in AI

Trustworthiness: Based on the recommendation of the European Parliament [16] that the development of safety-relevant DL algorithms should be based on a continuous collaboration between target group and developers, we will first involve the physician perspective as well as patient advocacy groups that are part of the Melcaya consortium in the creation of the requirement for the DL system. The models developed will later be evaluated for their impact on physician diagnostic behavior.

Technical Robustness and Safety: DKFZ will train deep learning algorithms to distinguish between melanomas and nevi in CAYA. These classifiers will be generated on the level of dermatoscopic images as well as on the histopathological level. None of these classifiers is intended to be a stand-alone system, but rather a diagnostic assistance system. All results will be checked for plausibility and ultimate decision by a physician expert. XAI methods will also support such plausibility checks (see below). Different types of data augmentation will also be explored in order to ensure robustness. Bias detection after deployment will be monitored, including for instance changes in performance due to a substantial shift in the input data distribution that could occur when using different scanners or cameras or differences in staining protocols for histological slides. Testing on an external cohort will be used as a final quantification of the robustness of the DL algorithms.

Transparency: To increase transparency of the CNN-based algorithms, DKFZ will implement and compare different XAI techniques. Typical XAI approaches to enable plausibility checks for the users are ways to point out areas or structures on the images that were of particular importance for the decisions or the identification of “similar” images. As outlined above, XAI methods also allow users to monitor the quality and performance of an algorithm over time and in different settings.

Type and source of biomaterial

In this project, digitized histological and dermatoscopic images of melanoma patients are used. These are provided by the partner hospitals in Tübingen, Barcelona and Florence via an IT interface provided by the DKFZ (see next section Data transfer). An exchange of human material in the form of tissue samples is not necessary.

Data transfer

Data transfer will take place on the legal basis of bilateral Data Transfer Agreement contracts via the HIFIS (Helmholtz Federated IT Services) system Filesender. The DKFZ IT department provides this system to ensure secure, encrypted data upload. Via an authentication and role

system, designated persons of the partner institutes are granted access to the data upload, which is only possible in encrypted, password-protected form due to system restrictions. At DKFZ, the data will then be downloaded by the authenticated Data Scientist and transferred to an HD-SATA hard drive, which will be connected to the Data Scientist's local workstation for development.

Time commitment for study participants:
There is no additional time commitment for the study participants for this study.
The data used in this project will be collected in routine clinical practice anyway.

Study duration AI-MEL: 01.12.2022-30.11.2026
Project timeline: see next page

Project timeline AI-MEL

Projektlaufzeit in Jahren	Jahr 1												Jahr 2												Jahr 3												Jahr 4											
Projektlaufzeit in Monaten	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48
Organisatorische Tätigkeiten																																																
Datenerfassung und Vorverarbeitung																																																
Generierung von Klassifikatoren																																																
Implementierung von XAI																																																
Bewertung der Ergebnisse in Zusammenarbeit mit Dermatolog*innen und Patholog*innen																																																

Data protection concept

The subject of this application are pseudonymized digital data that are securely transferred to the DKFZ. Data Transfer Agreements regulate the transfer of the digital data provided by the respective partners. DL-algorithms are developed on the basis of these data.

Patient data will be collected by the respective partner clinics in Tübingen, Barcelona and Florence, which have been submitted to the local ethics committees for review and approval. The respective information pamphlets and informed consent forms submitted to the patients are attached to this application.

All partners working on the MELCAYA project assure (see Annex ethical and legal aspects of clinical protocols) that the processing, communication and transfer of personal data of all participants shall comply with Regulation EU 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of data and the Organic Law 3/2018 of December 5 on the Protection of Personal Data and guarantee of digital rights. The legal basis that justifies the processing of the data is the consent of the patient or, in the case of minors, the consent of the parents, guardian or legal representative, in accordance with the provisions of Article 9 of EU Regulation 2016/679 in accordance with the provisions of article 9 of EU Regulation 2016/679.
The names of the patients/probands and all other confidential information are subject to medical confidentiality and the provisions of the General Data Protection Regulation (GDPR) and the State or Federal Data Protection Act (LDSG or BDSG). The data collected for these studies are identified by a pseudonymization code so that no information is included that would allow identification of the participants. Only the encrypted pseudonymized data are therefore transmitted to the DKFZ or third parties, which in no case contain information that can directly identify the participant (such as first and surname, initials, address, social security number, etc.). Third parties will not be given access to original records.
Only the study physician and his collaborators with the right to access the source data (medical history) could relate the collected data with the patient's medical history. The identity of the participants will not be available to any other person except for a medical emergency or legal requirement. Health authorities, Research Ethics Committee and personnel authorized by the study sponsor may have access to the identified personal information when necessary to verify data and procedures of the study, but always maintaining confidentiality in accordance with current legislation.
As a result, a record of all the processing activities will be kept and a risk assessment of those activities will be performed to know what measures will be needed and how to implement them. In addition to the rights already provided for in the previous legislation (access, modification, opposition and cancellation of data, deletion in the new Regulation), participants can now also limit the processing of data collected for the project that is incorrect, request a copy or transfer them to a third party (portability). To exercise these rights, patients can contact the principal investigator of the study or the Data Protection Officer of the Hospital Clínic de Barcelona

through protecciodades@clinic.cat. Likewise, they have the right to contact the Data Protection Agency if they are not satisfied.The Investigator and the Sponsor are obliged to keep the data collected for the study for at least 25 years after its completion. Subsequently, personal information will only be retained by the health care facility and by the sponsor for other scientific research purposes if the patient has consented to do so, and if permitted by applicable law and ethical requirements.

7. Expected outcome

For the AI-MEL project in general: we expecta that both the dermatoscopic and histological classifier(s) generated in this project will perform well on melanomas from young patients after they have been trained or fine-tuned on sufficient images of melanomas and nevi from the pre-specified age group. These classifiers could then be tested in the clinic for their potential to reduce the number of unnecessary biopsies on the one hand and to reduce the amount of initially overlooked melanomas on the other hand in CAYA, which in turn will directly benefit patients.
The skin age classifier that is going to be trained within the current project is expected to be able to detect much narrower age classes than our preliminary one due to a much larger amount of training data and an improved methodology. This classifier will enable us to investigate the impact of skin age on melanoma risk in young patients.

The XAI methods will serve to pinpoint the morphological differences between young patients' melanomas and nevi and thereby help dermatologists and pathologists to determine critical features associated with melanoma in young patients. With respect to the skin age classifier, they may also help human experts to better understand which features are critical to determine the extent of (premature) skin aging melanoma risk in young people

Benefit for the study participants when participating in the study:
The study participants will not directly benefit from this study. However, the resulting conclusions will help improve medical knowledge about the disease so that currently available diagnoses and treatments can be improved (see expected benefits in general).

8. Possible Risks

Possible risks for the study participants when participating in the study:
Participation in this study does not involve any inconvenience or risk for the participants, since the images are collected anyway in routine clinical practice.

Potential risks to the AI-MEL project in general:
The main challenge in training an AI-based image classifier is to obtain enough data. For rare tumors, there may be too few images available and/or the available data may be too heterogeneous to train the classifiers properly.
Approach to mitigate risk: open source data and data already collected by participating clinics can be used. In addition, many clinics will contribute data themselves during the project. In

addition, our existing skin lesion classifiers trained with images of older adult skin lesions can be used as a starting point for fine-tuning the classifiers on pigmented skin lesions of young patients (see Table 1).

Table 1: Risks + Measures

Risk description	Proposed risk-mitigation measures
Too few slides / too heterogeneous data to train classifiers properly	Open source data and data collected by participating clinics exist, many clinics will contribute data
Classifiers do not perform well	Prototypes for melanoma/nevus classification and a basic age classifier have been trained successfully; these can be used as basis to fine-tune the classifiers for young patients

Low burden on study participants:
MELCAYA will use existing and newly collected data from melanoma patients, healthy volunteers, and children/minors. The age of the subjects ranges from 0 to 30 years. MELCAYA will obtain written informed consent from all subjects involved in prospective data and sample collection. This consent will explain the purpose of the study and the nature and extent of participation. In the case of minors, consent will be obtained from the parent, guardian or legal representative. Subjects will be informed that participation in MELCAYA is completely voluntary and that they may withdraw their consent at any time without giving any reason or disadvantage for further medical care. Informed consent is provided in accordance with the provisions of Regulation (EU) No. 536/2014 on the protection of individuals and informed consent. All activities related to the use of subjects are monitored, including activities such as obtaining ethics approval from the relevant national/local ethics authorities and compliance with country-specific requirements and legislation (see Appendix ethical and legal aspects clinical trials).

9. Studytyp/Studydesign

The majority of the data collected in this project will be retrospective data collected from the partner clinics or from public databases. In addition, further images will be provided and collected prospectively during the project period by the partner clinic in Barcelona. The data collection will be non-interventional, i.e. the data collection will neither influence the diagnosis nor the treatment of the patients.
The study design is non-interventional. Both the macroscopic images and the scanning of the histopathological tissue sections are already performed in routine practice. The DL assistance

systems will be tested exclusively under artificial conditions with physicians. The project has no influence on diagnosis and therapy decisions.
The classification of patient data is based on age and hospital affiliation. The basic classifiers are developed on basis of public and project internal data of patients over 30. For the fine-tuning of the models, the data of the CAYA cohort are used. For the external test sets, data from external clinics with sufficient data set size will be ready for final testing of the models.

10. Inclusion and exclusion criteria

10.1 Inclusion Criteria

The focus of this study is to develop DL algorithms for melanoma early diagnosis and risk prediction in children, adolescents and young adults (CAYA). Therefore, data from patients under and including 30 years of age will be primarily included in this study. Data from adults will be used to develop algorithms for spitzoid lesions or the skin-age classifier. In addition, only macroscopic images of lesions suspicious for melanoma that can be assigned to the diagnosis of melanoma or nevus after biopsy verification will be included. For histopathologic images, the pathologic diagnosis of melanoma or nevus is the determining factor for inclusion in the study database.

10.2 Exclusion criteria

Patients without a melanoma or nevus diagnosis will not be considered for the diagnostic classifiers of this study. In addition, images with insufficient image quality will be excluded. For post-training of diagnostic classifiers, dermotoscopic and histologic images of patients under and including 30 years of age, will be used. For these classifiers, the base model will be trained on older patients. The algorithms for spitzoid lesions and the multiclass classifier for dermatoscopic skin age are trained with images from patients of all ages. Patients with confirmed melanoma will be excluded for training of the skin age classifier.

11. Randomization procedure/plan

The need for a randomization procedure does not apply to this study

12. Termination criteria

12.1 Individual discontinuation criteria

Withdrawal of consent for the use of data is possible at any time by the patients themselves or by the legal guardians or legal representatives without giving reasons and without disadvantages for further medical care. Withdrawal of consent will exclude the patient data from the study.

12.2 12.2 Discontinuation criteria for the study as a whole.

Unexpected and extensive difficulties with the data transfer on the project partners part could lead to the termination of the study, as this would mean that the basis for training the classifiers would no longer exist.
However, this is very unlikely, as the DKFZ offers an IT solution, the HIFIS (Helmholtz Federated IT Services) system Filesender for data transfer. This ensures a secure, encrypted data upload.
Excessively poor image quality of the data could also make it difficult to train the classifiers and thus hinder the progress of the project. Since we are obtaining data from various partners and from open source databases in this study, it is highly unlikely that all partners and the open source databases will provide poor image quality.

13. Statistical design

13.1 Statistical Methods

The evaluation of the adapted diagnostic classifiers is performed on internal and external test datasets. Internal test datasets follow the same distribution as the training dataset. In contrast, an external dataset should be unknown to the model, for example, by being collected from a clinic whose data were not used to develop the model. This allows a better assessment of the generalization ability of the models. Sensitivity, specificity, accuracy weighted by class ("balanced accuracy"), and area under the receiver operator curve (AUROC) are measured. Accuracy metrics of both the base models for the adult cohort and the adjusted classifiers are calculated. For the adjusted classifiers, the evaluation is performed exclusively on data from the CAYA cohort, in case of sufficient numbers of cases also separately for sub-age groups children (up to and including 15 years of age), adolescents (16-20) and young adults (from 21 to 30 years of age). As expected, there will be more cases for the cohort of young adults than for children and adolescents due to the incidence.
For the skin risk classifier, we will systematically assess whether there is an overestimation of skin age in young melanoma patients. Therefore, in the first step, the accuracy of the classifier on healthy skin will again be assessed using internal and external test data sets. A Wilcoxon Mann-Whitney U-test will be used to test the hypothesis of systematic overestimation of skin age in the presence of a malignancy.
To determine the usefulness of the final classifiers as assistive systems, they will be evaluated with pathologists as well as dermatologists. The participating clinicians will diagnose the same cases with sufficient time interval once without and once with AI assistance. Pairwise significance tests using the two-sided paired t-test will be performed to determine the effect of the system on diagnostic sensitivity, specificity, and accuracy.
The metrics as well as the statistical tests will be calculated using the Python programming language.

13.2 Case number planning

A precise case number planning of the data is only possible to a limited extent for this study. For the part of the data that we will obtain from databases, the case numbers for each age group are as follows: **ISIC database**: age>30: 5329 images; age 18-30: 445 images; age 15-18: 3 images; age <15: 0 images; for the **SCP2 database**: age >30: ~1900 images; age 18-30: 130 images; **TCGA database**: age:>30 ~450 images; age 18-30: 25 images, age 15-18: 1 image, age <15: 0 images.
Since melanomas in CAYAs are basically very rare, we have to expect that partner hospitals (UT, UNIFI and FCRB) can only provide us with a limited amount of data. In this case, case number planning is not useful as we will collect as many images as possible to ensure a sound data basis for training the classifiers.

14. Legal and ethical aspects

Diversity, Non-discrimination and Fairness: Skin cancer screening is already implemented in many European countries. Thus, the data that is publicly available and the new cases that will be collected in this project will have a bias towards “European” or even “Northern European” skin types. However, some public datasets also contain images of nevi and melanomas on dark skin that we can at least use to investigate the performance of our algorithms on dark skin. The skin age classifier will also be tested against sex, as factors such as skin thickness and amount of hair may differ between male and female patients. The required techniques to actually use the respective classifiers are not expensive and do not require extensive training and therefore have the potential to be used in most countries

14.1 Declaration of Helsinki

The investigation will be conducted in accordance with the Declaration of Helsinki in its current version.

14.2 Review by the Ethics Committee*.

The documents relating to the research project (esp. the study protocol) will be submitted to the Ethics Committee of the Medical Faculty of Heidelberg for review prior to the start of the study.

14.3 Information on the voluntary nature of participation

The participation of the patients/test persons is voluntary.

14.4 14.4 Information on informed consent or justification for waiving study-specific consent.

The subject of this study is pseudonymized digitized data that will be securely transmitted to the DKFZ. There will be no exchange of physical tissue samples. Furthermore, the DKFZ has no direct patient contact.

The consent of the study participants will be collected by the respective partner hospitals in Tübingen (UT), Barcelona (FCRB) and Florence (UNIFI). **The respective information leaflets and consent forms for the different age groups are attached to this application.**

The collection of data from subjects during regular clinical practice for research purposes will only be carried out through informed consent (according to Article 9.2 (a) of the GDPR). In the informed consent, the study participants are informed in writing and verbally about the nature and scope of the planned study, the purpose of the study and the nature and extent of their participation, in particular about the possible benefits for their health and any risks, in compliance with the provisions of Regulation (EU) No. 536/2014. Their consent will be documented by signing the consent form. In the case of minors, the consent of the parents, guardian or legal representative will be obtained. Study participants will be informed that participation in the study is completely voluntary and that they can withdraw their consent at any time without giving reasons and without disadvantages for their further medical care. In case of withdrawal from the study, (data) material already obtained will be destroyed or the patient/subject will be asked whether he/she agrees with the analysis of the material.
All study materials, including clinical and laboratory protocols, will be submitted to the appropriate institutional review boards (IRBs) for review and approval. Approval of the study protocol will be obtained prior to participant/case selection. All changes to the study protocol, materials, etc., will undergo ethical review and approval before the changes are incorporated into the study. All participating institutions will adhere to international ethical standards regarding principles for medical research involving human subjects and data (Declaration of Helsinki, 2013).

14.5 Information on the right of withdrawal and data destruction in the event of withdrawal

Withdrawal of consent to the use of data is possible at any time by the patient him/herself or by the legal guardian or legal representative without giving reasons and without disadvantages for further medical care. By withdrawing consent, the patient data will be excluded from the study and deleted from the study database. Data sets that have already been used for own publications can no longer be tracked due to the random ID assigned. The DKFZ reports the successful deletion of the data set back to the local study secretariat of the treating institution, which informs the study participant of the successful deletion.

14.6 Data protection details

The subject of this application are pseudonymized digital data that are securely transferred to the DKFZ. Data Transfer Agreements regulate the transfer of the digital data provided by the respective partners. DL-algorithms are developed on the basis of these data.

Patient data will be collected by the respective partner clinics in Tübingen, Barcelona and Florence, which have been submitted to the local ethics committees for review and approval. The respective information pamphlets and informed consent forms submitted to the patients are attached to this application.

All partners working on the MELCAYA project assure (see Annex ethical and legal aspects of clinical protocols) that the processing, communication and transfer of personal data of all participants shall comply with Regulation EU 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of data and the Organic Law 3/2018 of December 5 on the Protection of Personal Data and guarantee of digital rights. The legal basis that justifies the processing of the data is the consent of the patient or, in the case of minors, the consent of the parents, guardian or legal representative, in accordance with the provisions of Article 9 of EU Regulation 2016/679 in accordance with the provisions of article 9 of EU Regulation 2016/679.
The names of the patients/probands and all other confidential information are subject to medical confidentiality and the provisions of the General Data Protection Regulation (GDPR) and the State or Federal Data Protection Act (LDSG or BDSG). The data collected for these studies are identified by a pseudonymization code so that no information is included that would allow identification of the participants. Only the encrypted pseudonymized data are therefore transmitted to the DKFZ or third parties, which in no case contain information that can directly identify the participant (such as first and surname, initials, address, social security number, etc.). Third parties will not be given access to original records.
Only the study physician and his collaborators with the right to access the source data (medical history) could relate the collected data with the patient's medical history. The identity of the participants will not be available to any other person except for a medical emergency or legal requirement. Health authorities, Research Ethics Committee and personnel authorized by the study sponsor may have access to the identified personal information when necessary to verify data and procedures of the study, but always maintaining confidentiality in accordance with current legislation.
As a result, a record of all the processing activities will be kept and a risk assessment of those activities will be performed to know what measures will be needed and how to implement them. In addition to the rights already provided for in the previous legislation (access, modification, opposition and cancellation of data, deletion in the new Regulation), participants can now also limit the processing of data collected for the project that is incorrect, request a copy or transfer them to a third party (portability). To exercise these rights, patients can contact the principal investigator of the study or the Data Protection Officer of the Hospital Clínic de Barcelona

through protecciodades@clinic.cat. Likewise, they have the right to contact the Data Protection Agency if they are not satisfied.The Investigator and the Sponsor are obliged to keep the data collected for the study for at least 25 years after its completion. Subsequently, personal information will only be retained by the health care facility and by the sponsor for other scientific research purposes if the patient has consented to do so, and if permitted by applicable law and ethical requirements.

14.7 Financing/institutional connections/conflicts of interest

Dr. Brinker owns a company that develops mobile apps in the healthcare sector (Smart Health Heidelberg GmbH, Handschuhsheimer Landstr. 9/1, 69120 Heidelberg, Germany). There are no conflicts of interest.

15. List of References

1. Pappo, A. S. Melanoma in children and adolescents. *Eur. J. Cancer* **39**, 2651–2661 (2003).
2. Ferrari, A. *et al.* Defining and listing very rare cancers of paediatric age: consensus of the Joint Action on Rare Cancers in cooperation with the European Cooperative Study Group for Pediatric Rare Tumors. *Eur. J. Cancer* **110**, 120–126 (2019).
3. Vries, M. de, de Vries, M., Vonkeman, W. G., van Ginkel, R. J. & Hoekstra, H. J. Morbidity after inguinal sentinel lymph node biopsy and completion lymph node dissection in patients with cutaneous melanoma. *European Journal of Surgical Oncology (EJSO)* vol. 32 785–789 Preprint at https://doi.org/10.1016/j.ejso.2006.05.003 (2006).
4. Moscarella, E. *et al.* Excised melanocytic lesions in children and adolescents - a 10-year survey. *Br. J. Dermatol.* **167**, 368–373 (2012).
5. Cordoro, K. M., Gupta, D., Frieden, I. J., McCalmont, T. & Kashani-Sabet, M. Pediatric melanoma: results of a large cohort study and proposal for modified ABCD detection criteria for children. *J. Am. Acad. Dermatol.* **68**, 913–925 (2013).
6. Howlader, N., Noone, A. & Krapcho, M. Surveillance, Epidemiology, and End Results program cancer statistics review, 1975-2014. 2017. *Bethesda, MD: National Cancer Institute*.
7. Höhn, J. *et al.* Combining CNN-based histologic whole slide image analysis and patient data to improve skin cancer classification. *Eur. J. Cancer* **149**, 94–101 (2021).
8. Hekler, A. *et al.* Deep learning outperformed 11 pathologists in the classification of histopathological melanoma images. *European Journal of Cancer* vol. 118 91–96 Preprint at https://doi.org/10.1016/j.ejca.2019.06.012 (2019).
9. Hekler, A. *et al.* Superior skin cancer classification by the combination of human and artificial intelligence. *Eur. J. Cancer* **120**, 114–121 (2019).
10. Brinker, T. J. *et al.* Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. *Eur. J. Cancer* **113**, 47–54 (2019).
11. Schmitt, M. *et al.* Hidden Variables in Deep Learning Digital Pathology and Their Potential to Cause Batch Effects: Prediction Model Study. *J. Med. Internet Res.* **23**, e23436 (2021).
12. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A. & Torralba, A. Learning deep features for discriminative localization. in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2921–2929 (IEEE, 2016).
13. Sundararajan, M., Taly, A. & Yan, Q. Axiomatic Attribution for Deep Networks. in *Proceedings of the 34th International Conference on Machine Learning* (eds. Precup, D. & Teh, Y. W.) vol. 70 3319–3328 (PMLR, 06--11 Aug 2017).
14. Selvaraju, R. R. *et al.* Grad-CAM: Visual explanations from deep networks via Gradient-based localization. *arXiv [cs.CV]* 618–626 (2016).
15. Kim, B. *et al.* Interpretability Beyond Feature Attribution: Quantitative Testing with Concept Activation Vectors (TCAV). in *Proceedings of the 35th International Conference on Machine Learning* (eds. Dy, J. & Krause, A.) vol. 80 2668–2677 (PMLR, 10--15 Jul 2018).
16. Lekadir, K., Quaglio, G., Garmendia, A. T. & Gallin, C. Artificial intelligence in healthcare applications, risks, and ethical and societal impacts. *EPRS (European Parliamentary Research Service)* (2022).