

**Influence of Algorithms: Empirical study on the influence of algorithms' reliability
and transparency on the users' decision-making process**

Sören M. Schröder (399838)

Computer Science

Individual and Technology, RWTH Aachen University

Empirical Investigation of Communication in Human-Robot-Interaction

Prof. Dr. Astrid Rosenthal-von der Pütten

28. August 2020

Eidesstattliche Versicherung

Statutory Declaration in Lieu of an Oath

Schröder, Sören

Name, Vorname/Last Name, First Name

399838

Matrikelnummer (freiwillige Angabe)

Matriculation No. (optional)

Ich versichere hiermit an Eides Statt, dass ich die vorliegende Arbeit ~~Bachelorarbeit~~/
~~Masterarbeit~~* mit dem Titel

I hereby declare in lieu of an oath that I have completed the present paper/Bachelor thesis/Master thesis* entitled

Influence of Algorithms: Empirical study on the influence of algorithms' reliability
and transparency on the users' decision-making process

selbstständig und ohne unzulässige fremde Hilfe (insbes. akademisches Ghostwriting) erbracht habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt. Für den Fall, dass die Arbeit zusätzlich auf einem Datenträger eingereicht wird, erkläre ich, dass die schriftliche und die elektronische Form vollständig übereinstimmen. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

independently and without illegitimate assistance from third parties (such as academic ghostwriters). I have used no other than the specified sources and aids. In case that the thesis is additionally submitted in an electronic format, I declare that the written and electronic versions are fully identical. The thesis has not been submitted to any examination body in this, or similar, form.

Aachen, 28. August 2020

Ort, Datum/City, Date

Unterschrift/Signature

*Nichtzutreffendes bitte streichen

*Please delete as appropriate

Belehrung:

Official Notification:

§ 156 StGB: Falsche Versicherung an Eides Statt

Wer vor einer zur Abnahme einer Versicherung an Eides Statt zuständigen Behörde eine solche Versicherung falsch abgibt oder unter Berufung auf eine solche Versicherung falsch aussagt, wird mit Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft.

Para. 156 StGB (German Criminal Code): False Statutory Declarations

Whoever before a public authority competent to administer statutory declarations falsely makes such a declaration or falsely testifies while referring to such a declaration shall be liable to imprisonment not exceeding three years or a fine.

§ 161 StGB: Fahrlässiger Falscheid; fahrlässige falsche Versicherung an Eides Statt

(1) Wenn eine der in den §§ 154 bis 156 bezeichneten Handlungen aus Fahrlässigkeit begangen worden ist, so tritt Freiheitsstrafe bis zu einem Jahr oder Geldstrafe ein.

(2) Straflosigkeit tritt ein, wenn der Täter die falsche Angabe rechtzeitig berichtigt. Die Vorschriften des § 158 Abs. 2 und 3 gelten entsprechend.

Para. 161 StGB (German Criminal Code): False Statutory Declarations Due to Negligence

(1) If a person commits one of the offences listed in sections 154 through 156 negligently the penalty shall be imprisonment not exceeding one year or a fine.

(2) The offender shall be exempt from liability if he or she corrects their false testimony in time. The provisions of section 158 (2) and (3) shall apply accordingly.

Die vorstehende Belehrung habe ich zur Kenntnis genommen:

I have read and understood the above official notification:

Aachen, 28. August 2020

Ort, Datum/City, Date

Unterschrift/Signature

Abstract

Algorithms are a part of our everyday life. So-called algorithm decision systems (ADS) are systems that are designed to make a decision, based on input data. These systems are used to support professionals in the difficult decisions they have to make. Often these systems are developed through machine learning, which employs huge datasets to train an algorithm to detect certain patterns. One of the most remarkable achievements of machine learning was the development of an algorithm that can detect melanoma during an early stage with similar success rates to medical professionals. When using such a system a medical professional might be influenced in their decisions by the ADS. To investigate how transparency and reliability of the ADS influence the professionals' decision-making process we performed an empirical online study. In the study, medical students assessed pictures of nevi with the help of a mocked-up ADS. During the experiment factors like fairness, confidence, and conformity were measured. We found that people do not conform with unreliable ADS and that the gain in understanding of the algorithm by using it is influenced by its transparency. Furthermore, our findings indicate that people's self-reported confidence in the algorithm diverges from their actual behavior.

Keywords: *algorithmic decision making, algorithms, transparency, reliability, fairness, conformity, confidence, understanding, skin cancer, melanoma*

Table of Contents

1. Introduction	3
1. Algorithmic decision making	3
2. Transparency	5
3. Fairness	5
4. Reliability	5
5. Confidence	6
6. Conformity	6
7. Understanding	6
2. Method	7
1. Experimental design	7
2. Procedure	9
3. Operationalization dependent variables	10
4. Participants.....	11
3. Results	12
1. Fairness of the algorithm	12
2. Confidence in the algorithm	12
3. Conformity	13
4. Understanding of the algorithm.....	14
4. Discussion	15
1. Transparency	15
2. Reliability	17
3. Limitations & further research.....	18
5. Conclusion	18
6. Acknowledgments	19
7. References.....	20
8. Appendix A.....	21

1. Introduction

In many areas important decisions have to be made regularly. Whether an applicant is going to be hired, a defendant is released on bail, or a spot on the skin is a nevus or a melanoma. All these cases must be decided daily by professionals. These decisions can have serious consequences for those affected. This makes it all the more important that these decisions are made as carefully as possible. Together with the ongoing digitization of society, algorithms are also gaining ground in the aforementioned areas. Furthermore, the algorithms are used to support the professionals in their decision-making process (see Kirkpatrick, 2017, and Esteve et al., 2017). In addition to the expected simplification of work, there is also reason to believe that certain properties of the algorithms can influence the decision making of the users. Prior research showed, that transparency (Wang, Harper, & Zhu, 2020) and reliability (Dietvorst, Simmons, & Massey, 2015) can influence the perception of the users. This might result in influencing them when making their decision. The goal of this study is to investigate how transparency and reliability of algorithms influence the decision-making process of the people who use them.

1.1 Algorithmic decision making

In general, an algorithm is a series of calculations that receive a series of inputs and transform them into a series of outputs (Cormen, Leiserson, Rivest, & Stein, 2013). For us, the most interesting algorithms are those which are defined by Cormen et al. (2013) as correct algorithms. They define them as algorithms that stop for a certain input and deliver a correct result and solving the calculation. Another characteristic of an algorithm is that for the same input the same corresponding output is returned (Rogers, 1967), in other words, the algorithm is deterministic.

Algorithms can not only be manually written and executed by humans. They can also be generated by analyzing data and recognizing patterns in it. This approach (often referred to as *machine learning*) can be used to develop algorithm decision systems (ADS), which are involved in the process of decision making (Castelluccia & Le Métayer, 2019). They describe that the involvement of humans can be stated in a spectrum that ranges from systems that provide advice for a human who is responsible for the final decision to systems that make decisions fully automatically.

How algorithms on the autonomous end of the described spectrum can be used for job applications was shown by Wang, Harper, & Zhu (2020) in their work about how the fairness of algorithmic decision systems is perceived by concerned individuals. In their work, they created a mock system that behaved as if it could promote workers on a crowdsourcing workplace (an online platform where workers get paid for fulfilling micro tasks (Wang et al., 2020)), based on worker's data provided to the algorithm. Although this study did not use a real algorithm, because their research focused on the perceived fairness, it is indicative that ADS for job applications will eventually become reality. In other fields, ADS are already in use. One of these systems is the COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) software developed by Northpoint, Inc. (Kirkpatrick, 2017). The tool is used to decide whether a defendant is allowed to be released on bail or should be kept in (Hao & Stray, 2019). Similarly, the field of medical algorithmic decision systems has gathered significant attention. Esteva et al. (2017) developed an ADS that is able to detect skin cancer by analyzing images of the corresponding skin areas. Achieving success rates approximating those of experts they have tested. This topic is of special interest since skin cancer is the most common malignant disease among humans (Esteva et al., 2017). A dermatologist diagnoses melanomas primarily visually by applying the ABCDE method (Esteva et al., 2017). ABCDE is an abbreviation for different characteristics of nevi which can be used as an aid to detect melanoma in an early stage (Rigel, Friedman, Kopf, & Polsky, 2005). The characteristics of the nevus which are used by the method to assess them are **a**symmetry, **b**order irregularity, **c**olor, **d**iameter, and **e**volution (of the nevus over time).

The development of algorithms as presented by Esteva et al. (2017) is most likely to be continued. Either to provide easy access to early skin cancer detection for many humans by using smartphones (Esteva et al., 2017) or to support dermatologists in their work. For the latter, it would be interesting to investigate several effects that emerge from their use. Therefore, this study investigates the use of ADS in the context of medicine on the example of the task of assessing spots on the skin for melanoma. We set a focus on the transparency and reliability of the ADS and how they influence the decision-making process of medical professionals. In the following, we use the algorithm as a synonym for ADS, since we did not introduce the term ADS during the study because it was not necessary to know for the participants.

1.2 Transparency

In the application area of jurisdiction, the final decision is made by a judge (Kirkpatrick, 2017) and the same would apply when using ADS in a medical context. For COMPAS we know, that it is biased against particular subgroups (Angwin, Larson, Mattu, & Kirchner, 2016) and similar problems will likely occur in other systems for different application areas. Since the user of an ADS (e.g. judges or doctors) usually do not have the necessary technical knowledge to understand how the algorithm comes to its decision, the level of transparency the algorithm provides could be of significant interest regarding the question of how they are going to be used. Kizilcec (2016) showed that the provided transparency level of the ADS influences the trust in it. So we state our *Research Question 1: How is the user's decision-making influenced by the transparency of the involved algorithm?*

1.3 Fairness

Prior research could not find a relation between transparency and perceived fairness in the context of ADS (Wang et al., 2020). Two things might have led to this result. First, the operationalization of transparency by stating whether the algorithm was developed transparently or not might have been too weak a stimulus to measure an effect. Second, their participants were directly affected by the result which might interfere with the perceived fairness. By overcoming these two problems by giving a stronger stimulus and by using a setting where the user is not directly affected we state *Hypothesis 1: Transparency of the algorithm is positively associated with the perception of fairness.*

1.4 Reliability

Another factor that might influence the perception of the algorithm is how well it performs (i.e. how reliable it is). Studies have shown that the performance of an algorithm can influence how it is perceived by the user. Users were made aware of performance issues either directly by seeing that the algorithm fails (Dietvorst et al., 2015), or by having the information that the algorithm is known to make errors by providing information about error rates and biases (Wang et al., 2020). Because it is difficult to prevent errors during the development of software in general and for ADS in particular, since an already biased dataset can lead to a biased algorithm (Hao & Stray, 2019), it is important to know how the users of an ADS are going to

cope with the shortcomings of the algorithm and how it influences their decision-making process. So this study also investigates *Research Question 2: How is the user's decision-making influenced by the reliability of the algorithm involved?*

1.5 Confidence

When people recognize that an algorithm makes an error, they will lose confidence in the algorithm (Dietvorst et al., 2015). In their study, they found out that the loss of confidence is even higher than if the same mistake was made by a human. We suggest that this will also be the case if the participant witnesses a bad performance, which leads us to *Hypothesis 2: The users will show less confidence in decisions made by an unreliable algorithm*, or if they know that the algorithm will make errors due to a known error rate. Wang et al. (2020) reported that a known bias influenced the user perception of the algorithm. So we state *Hypothesis 3: The confidence in the algorithm will be less when the algorithm's error rates are known as compared to when no information about error rate is provided.*

1.6 Conformity

Since the confidence in the algorithm is expected to decrease for unreliable algorithms we also expect the users to deviate with their predictions from the ones stated by the algorithm. This lack of conformity gave us *Hypothesis 4: The user's predictions (of the probability of the nevus being melanoma) will deviate more from the predictions of a less reliable algorithm*

1.7 Understanding

When using ADS as an aid it would be important that the users understand how the algorithm is used and how the results need to be interpreted. This is necessary to use the algorithm's output for their own decisions. Transparency could be a key factor for understanding an ADS. Wortham, Theodorou, & Bryson (2017) showed that providing an insight into the decision-making process of an algorithm can increase the user's understanding. Since insight into an algorithm is a way of making the algorithm more transparent we expect that this will also be the case when providing transparency not during the use but in advance in the form of information about how the algorithm was developed. Therefore, we state our *Hypothesis 5: The understanding of the algorithm, before it is used, will be higher for an algorithm with high transparency than for one with low transparency*. Additionally, we also want to have a look into how the

understanding of the algorithm is changed by working with it (i.e. using the algorithm). This led us to *Research Question 3: How does the use of algorithms in the decision-making process change its understanding?*

2. Method

To answer the stated hypotheses and research questions an empirical study was performed as a randomized between-subjects online experiment. Performing the study online had the advantage of an easy distribution to more potential participants than by conducting the experiment in the lab. The provided background story of the experiment was to evaluate the performance of a new algorithm, which assesses the likability of skin cancer. The experiment was developed with SoSciSurvey¹ and can be found in Appendix A (see page 21).

2.1 Experimental design

In our between-subjects experiment, the participants were randomly assigned to a 2 (low transparency, high transparency) x 2 (unreliable, reliable) design. The distribution of the participants is shown in table 1. This design gave us the ability to investigate on the one hand the effect of the transparency of the algorithm on the perceived fairness (Hypothesis 1), confidence (Hypothesis 3), and the understanding of the algorithm before using it (Hypothesis 5), and on the other hand the effects of the algorithm's reliability on the confidence (Hypothesis 2) and the deviation of the user's prediction from the algorithm (Hypothesis 4).

Table 1

Number of participants with a specific combination of conditions.

	Low transparency	High transparency	Total
Unreliable	15	16	31
Reliable	15	15	30
Total	30	31	61

It was the participants' task to assess cases, which consists of a picture showing a section of human skin with a nevus, information about the symptoms, and an assessment by the

1. <https://soscisurvey.de>

algorithm (see Figure 1). The algorithm states to which degree (0% – 100%) it rates the nevus as a melanoma. The 15 cases were divided into 5 negative cases (clearly no melanoma), 5 positive cases (clearly melanoma), and 5 ambiguous (not clearly) cases. The classification and images were provided by a dermatologist from the RWTH Aachen University.

Figure 1

Example of a case showed during the experiment

Bitte sehen Sie sich das Bild an und lesen Sie die untenstehenden Informationen sorgfältig durch. Und beantworten Sie die folgenden Fragen:



- Der Algorithmus sagt voraus, dass es sich bei diesem Muttermal mit einer Wahrscheinlichkeit von **12%** um ein Melanom handelt.
- Der Fleck auf der Haut ist im Laufe der Zeit verblasst.

Note. The image shows a *correct negative case*. The shown nevus is no melanoma and the assessment of the algorithm reflects that.

Transparency: The participants performed the experiment either in the low or high transparency condition. Therefore, they were provided with different information about how the algorithm was developed. In the low transparency condition, the participants only were informed that the algorithm was developed by "us". In the high transparency condition, the participants additional were informed about which factors were used to develop the algorithm, how the algorithm was trained, and how it performed. We indicated that the performance was measured with a subset of the data not used during the training.

Reliability: Besides the characteristic of low or high transparency the algorithm had also one of the two possible reliability levels. In the good reliability condition, the algorithm did not make any mistake. In the unreliable condition, the algorithm assessed three cases obviously wrong. One of the positive cases was changed to a false positive (clearly no melanoma, but the algorithm state the opposite) and two of the negative cases were changed to false negatives (vise versa). We decided not to add any more mistakes to avoid destroying the confidence in the algorithm, which might have led to ignoring the algorithm altogether. Also then the error rate would have diverged too much from the rate stated in the algorithm description for the high transparency condition what could make the participants suspicious. Having one more false negative than false positive is because a false negative assessment has severe implications for the patient

2.2 Procedure

The study consists of three parts. Starting with an introduction containing information about the experiment and basic knowledge followed by the task and its explanation, and completed by a questionnaire after the task, regarding the perception of the algorithm and demographical questions. As seen in Appendix A (see page 21) several more questions were asked during the experiment. As they were not used in this particular study, they are not mentioned here and we do not expect that they had any impact on our results.

On the introduction page, visitors were informed about the conditions under which they can participate in the study (participation was anonymous, needed time was about 25 minutes, the rough structure of the experiment, a brief description of the task, etc.). They were also informed that the purpose of the study is to evaluate an algorithm for detecting melanoma, which was developed at the computer science department at RWTH. This deception was resolved at

the end of the experiment, but necessary to keep the participants uninformed during the experiment to prevent bias by knowing the actual research goals. Further, they were informed about the optional lottery. The participants were asked to read basic information about the ABCDE-Method to provide them a short recap and ensure some common base knowledge level. To ensure this, a knowledge check was performed before continuing.

The second part started with the explanation of the task and an example of the later provided cases. Afterward, according to the assigned transparency condition, the participants got the corresponding explanation of the algorithm's development. To ensure the information was read attentively the participants had to answer one (low transparency) or three (high transparency) corresponding questions correctly. We accepted the different amount of questions since we wanted to ensure that in the high transparency condition all given information was read and at the same time we wanted to keep the information for the low transparency condition as sparse as possible. We do not expect this to be a significant confounding factor in our study. Afterward, the participants stated how confident they felt in using the algorithm, before assessing of the 15 cases.

After performing the task the participants stated again how confident they felt in using the algorithm and asked them to provide several demographic data (gender, age, fields of studies, setting they learned ABCDE method, etc.). In the debriefing, the occlusion of the *algorithms* nature and the experiment's conditions and which were used in their case was revealed. Finally, the participants were asked to visit a dermatologist if they observe conspicuous nevus on their skin.

2.3 Operationalization dependent variables

Perceived fairness: The participants rated the fairness of the algorithm after they perceived the algorithm by assessing the cases. They were asked to consider how fair they would rate the algorithm on a 5-point Likert scale (*Very unfair* to *Strongly fair*).

Conformity: For each case, the participants had to state a prediction, similar to the one provided by the algorithm, how likely the nevus is a melanoma (0% – 100%). The answer was used to calculate the deviation from the algorithm's prediction.

Confidence in the algorithm: Besides the assessment of the nevus, the participants also stated how reliable they consider the algorithm. They provided the answer for each case on a 5-point Likert scale (*Not at all reliable* to *Very reliable*).

Understanding the algorithm: The participants self-estimated how good they have understood how to use the algorithm after they were informed about how the algorithm was developed and after they performed the cases. Three statements (*I think I know how the algorithm works*, *I think I have a good grasp of the algorithm*, and *I think I know how to use the algorithm*) were rated on a 7-point Likert scale, from *completely disagree* to *completely agree*. At both points the measurement had a high internal consistency ($\alpha = .86$ and $\alpha = .87$).

2.4 Participants

We recruited 61 participants from which 25 stated their gender as male and 36 as female. From the participants, 29 indicated their age at 18 - 24 years, 30 at 25 – 34 years, and 2 at 35 – 44 years. The absence of older participants was expected and was due to the recruitment strategy. The taken convenience sample was recruited in several ways. The experiment was posted in Facebook groups of medical students, sent to faculties of medical students with the request to distribute the experiment, and it was posted on online bulletin boards of medical faculties. Furthermore, personal contacts were asked to participate and distribute the experiment, and a class coordinator from the medical faculty of the RWTH Aachen University distributed the experiment to their students. The study was conducted by 21 participants in English and by 40 in German.

The participants were asked only to participate if they were familiar with the ABCDE-Method, which is the reason the recruitment was limited to the described ways. This requirement was stated in the introduction and verified during the experiment by asking where the participants have learned about the method and what their educational background is. From this, we could see that 52 of our participants already knew the ABCDE method. Nevertheless, we have chosen not to remove the other participants from the sample due to the already low number of participants and the resulting imbalance of the conditions. As compensation for their time, the participants had the option to register their mail address, which was stored separately

from the experimental data, to a lottery with which they had the chance to win one of five Amazon gift cards with a total value of 150 Euro (1 x 50 Euro and 4 x 25 Euro).

3. Results

The results of our measurements show that overall there is less impact of transparency and on reliability than we expected beforehand. Nevertheless, the collected data gave some insights into how the user's perception of the algorithm is influenced and the resulting decisions and actions.

3.1 Fairness of the algorithm

After performing all 15 cases the participants were asked to rate the fairness of the algorithm. The measured difference between the high transparency condition ($M = 3.52$, $SD = 0.72$), and the low transparency condition ($M = 3.70$, $SD = 0.65$) was quite small and therefore even higher for the low transparency condition. Likewise, a two-way analysis of variance (ANOVA) did not reveal any support for a significant effect of transparency on the perceived fairness of the algorithm, $F(1, 57) = 1.04$, $p = .313$, $\eta^2 = .018$, which indicates that the perceived fairness is not influenced by the algorithm's transparency. So, there is no support for Hypothesis 1.

3.2 Confidence in the algorithm

The confidence in the algorithm was stated by the participants for each case. Among all cases, the result was between *moderate* and *reliable* ($M = 3.48$, $SD = 0.42$). A two-way ANOVA did not show any significant effect on the confidence in the algorithm. Neither by transparency, nor by reliability, nor by an interaction of both (see table 2). Therefore, no support for Hypothesis 3 and Hypothesis 2 could be found. Nevertheless, the result ($p = .140$) suggests at least some effect on reliability exists. So, we looked only at the cases where the unreliable algorithm did obvious mistakes. A two-way ANOVA showed that reliability had a significant effect on the confidence in the algorithm, even with a medium effect size (see table 2). But the confidence for the unreliable algorithm was close to *moderate* ($M = 2.92$, $SD = 0.72$) and only a bit below the confidence into the reliable algorithm ($M = 3.35$, $SD = 0.55$).

Table 2*Statistical analysis of confidence in the algorithm*

Variable	All cases				Unreliable cases			
	df	F(1, 57)	Significance	Partial Eta Square	df	F(1, 57)	Significance	Partial Eta Square
Transparency	1	0.839	.363	.015	1	2.490	.120	.042
Reliability	1	2.234	.140	.038	1	6.379	.014	.101
Transparency*Reliability	1	0.063	.802	.001	1	0.129	.720	.002

3.3 Conformity

We used the participants' assessments of the cases to calculate the deviation from the algorithm by taking the absolute difference in percentage points between the assessments of the participants and the algorithm. The deviation can be described as moderate ($M = 17.2\%$, $SD = 7.3$). For the reliable algorithm, we can report even lower deviations in the assessments ($M = 14.25\%$, $SD = 7.36$). In contrast the deviation for the unreliable algorithm shows a 5.96 percentage points higher deviation ($M = 20.1\%$, $SD = 6.2$). A two-way ANOVA revealed a significant effect of reliability on the deviation of assessments, $F(1, 57) = 11.17$, $p = .001$, $\eta_p^2 = .164$, with a large effect size. But no significant effect on the deviation was caused by transparency or interaction between transparency and reliability (see table 3). These results support Hypothesis 4.

Table 3*Statistical analysis of prediction deviation*

Variable	df	F(1, 57)	Significance	Partial Eta Square
Transparency	1	0.41	.522	.007
Reliability	1	11.17	.001	.164
Transparency*Reliability	1	1.31	.267	.022

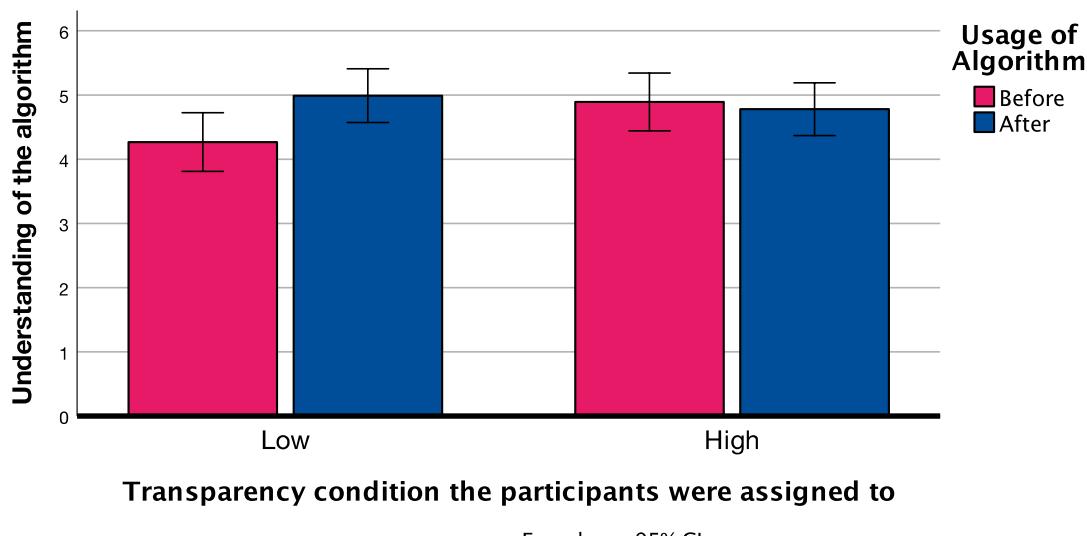
3.4 Understanding of the algorithm

For the measurement before assessing the cases, the understanding of the algorithm in the low transparency condition ($M = 4.27$, $SD = 1.41$) was below the understanding in high transparency condition ($M = 4.89$, $SD = 1.07$). But a two-way ANOVA revealed no significant significant effect of transparency on the understanding, $F(1, 57) = 3.81$, $p = .056$, $\eta_p^2 = .063$, which gave no support for Hypothesis 5.

Although, the result indicated that there might be some effect of the transparency on the understanding of the algorithm. Comparing the change of reported understanding between the two measurements, before and after the cases, show that for the low transparency group the understanding was higher after using the algorithm's assessment compared to before, with only the explanation. In contrast to the high transparency group, the value dropped a little bit from before to after (see Figure 2). A two-way repeated-measures ANOVA revealed there is a significant effect of the interaction between the transparency and using the algorithm on the understanding of it, $F(1, 57) = 8.32$, $p = .006$, $\eta_p^2 = .127$.

Figure 2

Understanding of the algorithm before and after using it



Note. The understanding of the algorithm has increased for the low transparency group from before to after using the algorithms results. For the high transparency group, the reported understanding decreased a bit from before to after, but not significantly.

4. Discussion

The goal of this study was to get a better insight into how ADS are used by people in general and by medical professionals in particular. We looked at transparency and reliability as factors that influence the perception of the algorithm and thereby how it is used. In doing so we observed several dependent factors: fairness, confidence, conformity, and understanding.

4.1 Transparency

The study did not find support for the claim that transparency affects perceived fairness (Hypothesis 1). This result is unexpected since we know that other research has observed such an effect (Wang et al., 2020). One noted difference between the studies is that in our scenario the user of the algorithm was not directly affected by the outcome of the result, which might have led to another perception of fairness in general (Greenberg, 1983). Another reason that could explain the differing results could be the kind of decision that was made. The decision made in the study by Wang et al. (2020) was not either right or wrong. To promote someone or not is always a balancing of arguments. In contrast, the assessment of a nevus being a melanoma could be either correct or incorrect. So the perception of fairness in the context of medical assessments might not be connected to the error rate itself. To investigate this topic further we suggest investigating how different error rates for certain subgroups (e.g. skin color) would influence the perceived fairness. Furthermore, the perception of fairness could be compared between after stating the error rates but before using the algorithm and after using the algorithm to see how the usage of the algorithm changes the perception of fairness.

The confidence of the participants in the algorithm was not significantly affected by the level of transparency, which does not support Hypothesis 3. This could be a hint that users, even if they know that the algorithm will make mistakes, are not including this information in their assessment of their confidence in the algorithm. Since we only observed no significant differences in the confidence we have to assume that this kind of transparency is not an influencing factor for users of ADS. On the other hand, there could have been other reasons why we did not observe any influence here. One might be that manipulation by stating error rates was too abstract and therefore too weak to influence the judgment. Since during this time some cognitive load was put on the participants, they had to assess the cases, another reason for observing no effect could be that they forgot about the information that was presented at the

beginning of the task section. Further research should investigate if a continuous reminder (e.g. displaying error rates during usage) would affect the confidence in the algorithm.

Regarding the initial assessment of how well the participants understood the algorithm before they used it we could not find support for Hypothesis 5. It seems that a detailed explanation of how the ADS was trained did not lead to significantly better understandings. Since the result was quite close to being significant ($p = .056$) we assume that our manipulation was not ideal to improve the understanding of the algorithm, but that in general, more information about the algorithm will help users to understand how to use it. To gain a better effect an approach similar to that chosen by Wortham et al. (2017) might be more effective. They provided for each decision of the algorithm an insight into how this decision was achieved. Regarding the field of medicine, it might be helpful for medical professionals to have more information on the decision that the ADS states. It could show how this decision was made and which factors had which amount of influence on it. For further research, we suggest performing experiments where the participants are provided with more information during the cases to see if this would help them gain a better understanding of the algorithm. To answer Research Question 1, we can determine that transparency seems to have less effect on the user's decision-making process than we expected beforehand. On the other hand, we saw that transparency interferes with some kind of learning effect, which suggests that transparency is still important when developing ADS for people who are non-experts regarding algorithms.

Regarding Research Question 3 we found that there is an interference effect between using the algorithm and transparency. We saw that the understanding increased significantly more for the low transparency condition by using the algorithm compared to the high transparency condition. After using it the knowledge for both groups was on a quite similar level. Even if the understanding of the algorithm before using it was not significantly influenced by the transparency we can see that more information, in the beginning, reduces the increase of understanding during the usage. If the understanding of the algorithm increases during the usage this could mean that especially the first cases a medical professional assess by using ADS might suffer due to the lack of understanding. So we suggest providing information about how the algorithm was trained to help the professionals to early gain an understanding of how to use the algorithm. So we can answer Research Question 3 by stating that the use of the algo-

rithm influences the understanding of it but that also other factors (here transparency but there might be others) have an influence on how this change in understanding is expressed.

4.2 Reliability

Besides the influence of transparency, the study investigated reliability as another factor that might influence the usage of ADS. Here we could not find support for Hypothesis 2, which was quite unexpected, since prior research suggested that unreliable ADS will result in less confidence (Dietvorst et al., 2015). A possible explanation could be that the participants did not perceive the algorithm as performing badly, since they did not get feedback after assessing the cases. For further research, we suggest changing the experiment, such that the assessment of the algorithm's quality does not rely on the knowledge of the participants. With a more carefully selected sample to ensure a higher common basic knowledge, this problem could be also solved. Besides this, it is interesting that the participants perceiving the unreliable algorithm have deviated significantly in assessing the cases from those with a reliable algorithm and still perceived the reliability of the algorithm as close to *moderate*. This shows a deviation from what they reported to how they behaved and suggests that self-reporting might not be ideal for measuring confidence in the algorithm (and perhaps also not ideal for measuring self-confidence). Since the behavior is closer to what prior research suggested (Dietvorst et al., 2015) we suggest measuring confidence in a behavioral manner in future research.

As already mentioned, the reliability of the algorithm influences the conformity of the assessments. These results give us support for Hypothesis 5. We can tell that the participants did not just follow the assessments of the ADS but used the ABCDE method to give their assessment. Another reason for the divergence of conformity and confidence could be that the participants just ignored the algorithm. In the presented study, the stimulus material was on a basic visually level. The image and the results of the algorithm were obviously part of the user interface from SoSciSurvey, which did not make them stand out enough which might explain why they may have been overlooked by the participants. To overcome this weakness and to even provide a more realistic scenario we suggest mocking a user interface of some application and including it into the cover story. This will increase the credibility of the cases and the algorithm.

4.3 Limitations & further research

The most obvious and severe limitation of this work is the number of participants. This limitation might be a reason why several measurements showed some differences but often not significant. We think that the lack of significant findings can be solved by performing the study with more participants. Using a special scenario like detecting skin cancer has, on the one hand, the advantage, that it is easier to give the participants a valid reason why they are needed in the study. On the other hand, it reduces the number of people who can participate drastically. We, therefore, propose to study the field of ADS with more general topics. To convince the people that their participation is of importance might be more challenging but it is outweighed by the advantage of a bigger sample.

By following this suggestion the second limitation of our work, which was the participants without prior knowledge of the ABCDE method, would not have occurred. For studies where particular information known to the participants is important and small sample sizes are expected, we suggest the following to prevent imbalanced conditions. We saw that asking where the needed knowledge was acquired can be used to filter out unsuitable participants. Here it is crucial to check the gathered data on a very regular basis (several times a day) to be able to react to irregular participation and adjust the randomization correspondingly to avoid imbalanced conditions.

In this work, only two levels of reliability were used, which was due to the expected low number of participants. Since the reliability only had a significant influence on the conformation of the assessments, more levels of *bad* reliability would help to gain more insight. Several levels with an even higher number of wrong assessed cases would show how these influence e.g. the confidence in the algorithm. Also, the strength of the error could have been varied. Future research could investigate these more differentiated levels of error to gain knowledge of how different types of errors influence the use of ADS.

5. Conclusion

We showed in the presented study how the medical professionals' perception of ADS is influenced by its transparency and reliability. By manipulating these and measuring other factors (fairness, confidence, conformity, and understanding) we could gain some insight into the user's perception. Besides several insignificant influences of the two factors, the main findings

in the study were that people do not agree with unreliable algorithms when they made mistakes and that the understanding of these algorithms while using them is influenced by the amount of information about the algorithm provided. We also found an interesting divergence between how confident people are in an algorithm and how they behave while using it.

6. Acknowledgments

I would like to thank Sourabh Zanwar for the great cooperation during the semester in this project. Furthermore, I want to thank Prof. Dr. Astrid Rosenthal-von der Pütten and Nikolai Bock for their supervision of this project and for providing their help to us throughout the seminar. Finally, I would like to thank Kailex Johnston for proofreading this term paper.

7. References

- Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias. *ProPublica*, May, 23, 2016.
- Castelluccia, C., & Le Métayer, D. (2019). *Understanding algorithmic decision-making: Opportunities and challenges*. European Parliament. <https://www.doi.org/10.2861/536131>
- Cormen, T. H., Leiserson, C. E., Rivest, R., & Stein, C. (2013). *Algorithmen - Eine Einführung*. Walter de Gruyter GmbH & Co KG. <https://doi.org/10.1515/9783110522013>
- Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1), 13. https://repository.upenn.edu/cgi/viewcontent.cgi?article=1392&context=fnce_papers
- Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115-118. <https://doi.org/10.1038/nature21056>
- Greenberg, J. (1983). Overcoming Egocentric Bias in Perceived Fairness Through Self-Awareness. *Social Psychology Quarterly*, 46(2), 152-156. <https://doi.org/10.2307/3033852>
- Hao, K., & Stray, J. (2019). *Can you make AI fairer than a judge? Play our courtroom algorithm game*. Retrieved 2020-08-16 from <https://www.technologyreview.com/2019/10/17/75285/ai-fairer-than-judge-criminal-risk-assessment-algorithm/>
- Kirkpatrick, K. (2017). It's not the algorithm, it's the data. *Commun. ACM*, 60(2), 21-23. <https://doi.org/10.1145/3022181>
- Kizilcec, R. F. (2016). How Much Information? Effects of Transparency on Trust in an Algorithmic Interface. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2390-2395. <https://doi.org/10.1145/2858036.2858402>
- Rigel, D. S., Friedman, R. J., Kopf, A. W., & Polsky, D. (2005). ABCDE—An Evolving Concept in the Early Detection of Melanoma. *Archives of Dermatology* *Arch Dermatol*, 141(8), 1032-1034. <https://doi.org/10.1001/archderm.141.8.1032>
- Rogers, H. (1967). *Theory of Recursive Functions and Effective Computability*. <https://doi.org/10.1137/1011079>

- Wang, R., Harper, F. M., & Zhu, H. (2020). Factors Influencing Perceived Fairness in Algorithmic Decision-Making: Algorithm Outcomes, Development Procedures, and Individual Differences. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1-14. <https://doi.org/10.1145/3313831.3376813>
- Wortham, R. H., Theodorou, A., & Bryson, J. J. (2017). *Robot Transparency: Improving Understanding of Intelligent Behaviour for Designers and Users Towards Autonomous Robotic Systems*. In Y. Gao, S. Fallah, Y. Jin, & C. Lekakou (pp. 274-289). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-64107-2_22

8. Appendix A

Below is the start page common for all conditions, followed by the high transparency and reliable condition experiment and the low transparency and unreliable condition experiment. The other two possible combinations of conditions result by combining them accordingly.

Welcome!

Thank you for taking the time to support us in our research!

The early detection of skin cancer is important, as failure to detect it can have serious consequences. To support physicians in this task, we have developed an **algorithm-based recommendation system for the identification of melanoma**. With your help, we would like to **test the accuracy and reliability of this system**.

The algorithm has been developed in cooperation with the University Hospital Aachen and the Department of Computer Science at the RWTH Aachen University.

We now need the expertise of students from the medical faculty to better evaluate the quality of the results. For this purpose, the following explains the decision situation (does a biopsy have to be performed?) and the relevant factors for the decision, so that you can then independently assess some case descriptions (picture + information). Both before and after the assessment you will also answer a few additional questions.

- To participate, you must be **familiar with the basics of the ABCDE method** for the detection of melanomas.
- Participation in the study **will take about 25 minutes**.
- The information you provide **will be stored completely anonymously and will only be used for scientific purposes**. At no time will it be possible to draw conclusions about your person.
- You can **cancel your participation at any time**.
- You must be **at least 18 years old**.
- You should use a **tablet or computer** for optimal viewing. If you use a **mobile**, please make sure to use it in **landscape**.

As a thank you for your participation, we are giving away Amazon vouchers worth a total of 150 euros (1 x 50 euros and 4 x 25 euros). To take part in the raffle, you must enter a valid e-mail address at the end of the questionnaire. This e-mail address will be stored separately from your other entries and will not be passed on to third parties.

If you want to start the survey, click on "Continue".

We would like to know how you generally think about algorithms.

Please indicate how much you agree with the following statements.

I think..	completely disagree	strongly disagree	somewhat disagree	undecided	somewhat agree	strongly agree	completely agree
algorithm-based decisions are not transparent.	<input type="radio"/>						
People could let themselves be determined by algorithms.	<input type="radio"/>						
Algorithms can make more precise decisions than a human.	<input type="radio"/>						
algorithm-based decisions are too uncertain for me.	<input type="radio"/>						
Algorithms are not aware of the responsibility of a decision.	<input type="radio"/>						
Algorithms prefer no one.	<input type="radio"/>						
Algorithms can relieve people of difficult decisions.	<input type="radio"/>						
Algorithms treat all people equally.	<input type="radio"/>						
Algorithms are completely rational and therefore comprehensible.	<input type="radio"/>						
Algorithms can make decisions that no human should have to make.	<input type="radio"/>						
that certain decisions should only be made by people.	<input type="radio"/>						
Algorithms are less flexible than humans in evaluating decision factors.	<input type="radio"/>						
Algorithms can save decision makers a lot of work.	<input type="radio"/>						
Algorithms cannot be bribed.	<input type="radio"/>						

that many people would simply follow algorithm-based recommendations.	<input type="radio"/>						
it is problematic that algorithms cannot be held responsible	<input type="radio"/>						
Algorithms can process more data than a human.	<input type="radio"/>						
Algorithms cannot consider the consequences of a decision.	<input type="radio"/>						
Algorithms can not be held responsible.	<input type="radio"/>						
Algorithms are not suitable for making personal decisions.	<input type="radio"/>						
Algorithms can analyze data faster than a human.	<input type="radio"/>						
Algorithms apply the same scale to everyone.	<input type="radio"/>						
Algorithms should not make morally difficult decisions.	<input type="radio"/>						
Algorithms make decisions more responsibly than humans.	<input type="radio"/>						
algorithm-based decisions are too impersonal for me.	<input type="radio"/>						
Recommendations by an algorithm lead to people thinking less about decisions.	<input type="radio"/>						
Algorithms have no good and no bad days.	<input type="radio"/>						
Algorithms have no prejudices.	<input type="radio"/>						

This algorithm was developed with the "ABCDE method" in mind. This is a common method for detecting melanoma by visual inspection. It can help both medical laypersons and clinicians to identify features in a skin lesion that could indicate melanoma.

Please read the following text carefully.

ABCDE Method

Asymmetry:

A stands for asymmetry. A melanocytic nevus (harmless mole) is usually symmetrical, while a melanoma often has an irregular or asymmetrical shape.

Border Irregularity:

B stands for Border Irregularity, i.e. irregularity at the border. A melanocytic nevus (harmless liver spot) has smooth and even edges, while a melanoma often has irregular and difficult to define edges.

Color Variation:

C stands for Color Variation. A melanocytic nevus (harmless mole) usually has a single shade or two shades, one entering the other or repeating regularly (generally pink, brown, or tan). Melanoma can be brown but can have up to five or six colors (blue, black, brown, tan, grey, pink, and red). These colors are unevenly or irregularly distributed.

Diameter greater than 6mm:

D stands for diameter. Most melanomas have a diameter of more than 6 mm when they are diagnosed.

Evolution:

E stands for Evolution or change. A melanocytic nevus (harmless mole) is usually stable and does not change in size, shape, or color, whereas a melanoma changes over time. Changes in size, color, shape, or structure can become noticeable over months to years.

1. The following are the diameters measured for spots on 5 different people for the diagnosis of Melanoma. Which is most likely to be a Melanoma based on the diameter?

- 4.0 mm
- 5.1 mm
- 6.9 mm
- 5.5 mm
- 3.6 mm

Please read the following task description carefully.

We would like to ask you to put yourself in the role of the treating dermatologist.

Based on the information available to you, you will decide how likely the birthmark could be melanoma and whether you would have a biopsy performed.

Below are some cases where you should make this decision. In the example below you can see that the following information is available to you:

- You can analyze the image of the melanoma using the well-known ABCDE formula.
- You can ask the patient whether the melanoma has changed over time in terms of color, shape or size.
- You can ask the patient whether there has been any itching or bleeding at the birthmark.

Example: Image of a birthmark for analysis with accompanying information



- The spot on the skin has slightly increased in size over the last four months (6.2mm to 6.6mm). It does not itch or bleed.

MANIPUL

In addition to the case descriptions you will receive a risk assessment of our algorithm in the following form:

The risk assessment algorithm predicts that there is a 20% probability that this mole is a melanoma.

This algorithm was developed by us to help dermatologists in the assessment of moles.

To this end, factors were first determined which are taken into account in the analysis by the algorithm. The following factors turned out to be particularly relevant:

The ABCDE formula:

A = Asymmetry

B = limitation

C = Colour (English colour)

D = diameter or dynamic

E = sublimity

changes of the birthmark regarding size, shape, colour

itching or bleeding from the birthmark

The next step was to choose a suitable model for the algorithm.

This model was then trained on a database consisting of 22,783 real cases from the dermatology clinics of the UKA and the other clinics of the Aachen city region.

One case consisted of a photo of the affected skin area, additional information about changes and symptoms and the decision made by a dermatologist.

The cases were divided into training data sets and verification data sets in a ratio of 80/20 before the training. So the algorithm was trained with 80% of the data sets. The remaining 20% were used to verify the performance of the algorithm.

Through this process a prediction accuracy of 86% was achieved. This means that the probability of the algorithm making a mistake is 14%. Thus, out of 100 moles, the algorithm would make an incorrect recommendation in 14 cases.

The error can either be due to the fact that the mole is biopsied even though there is no risk of melanoma and thus an unnecessary intervention has been performed. Or the error is that an existing melanoma is not detected and not treated.

2. If the algorithm were to evaluate 200 cases, how many errors would it likely make based on the information above?

8

14

20

28

42

80

3. How many real cases were used to train the algorithm?

10000

15153

22783

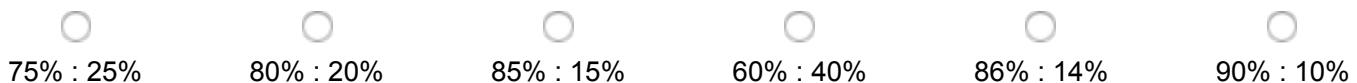
32782

20145

50000

4. How was the training and verification data from the cases split?

The following options are in format (training data : Verification data)



Please evaluate the information you have received about the algorithm:

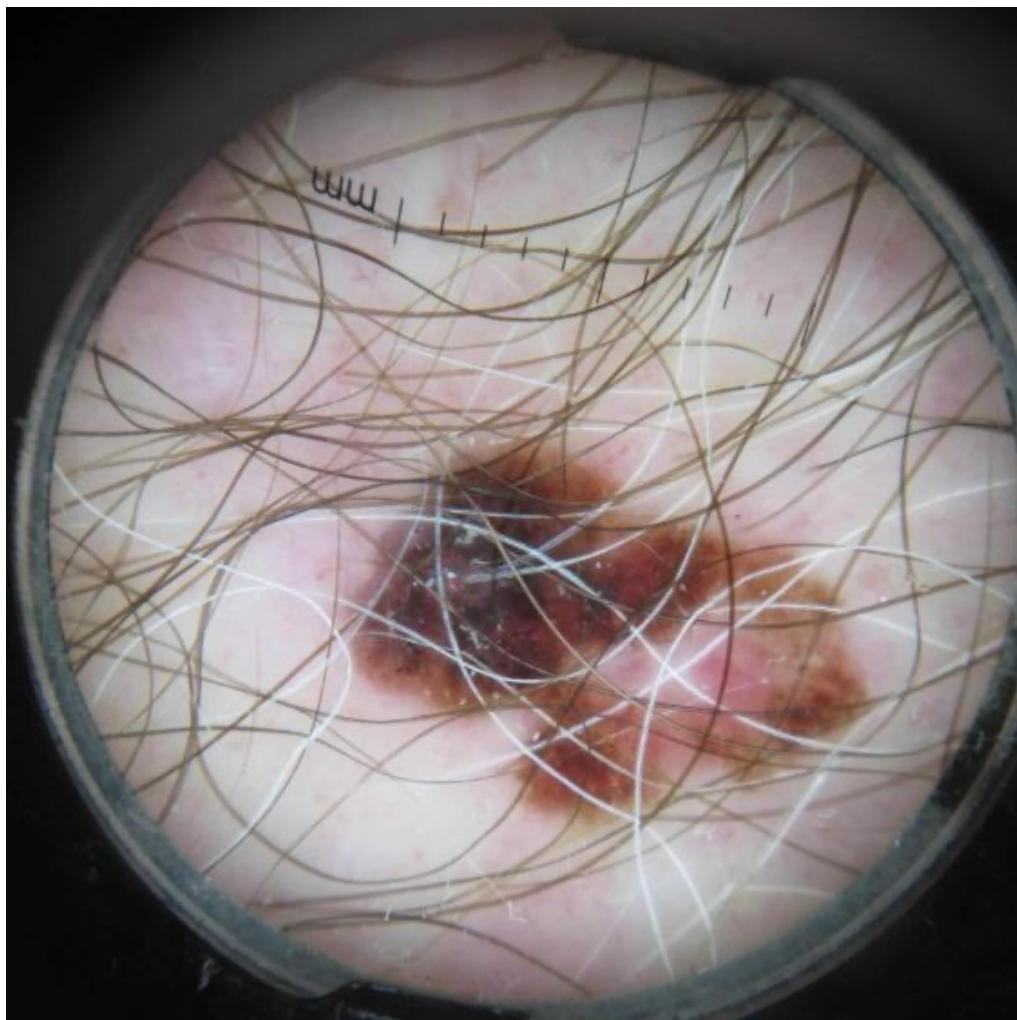
The explanation was...

overwhelming	<input type="radio"/>	Underchallenging						
uninformative	<input type="radio"/>	informative						
not useful	<input type="radio"/>	helpful						
hard to understand	<input type="radio"/>	easily understandable						
misleading	<input type="radio"/>	unambiguous						
not understandable	<input type="radio"/>	understandable						
very short	<input type="radio"/>	very long						

Please indicate to what extent you agree with the following statements.

	completely disagree	strongly disagree	somewhat disagree	undecided	somewhat agree	strongly agree	completely agree
I think I know how to use the algorithm.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I think I know how the algorithm works.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I think I have a good grasp of the algorithm.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Please look at the image and read the information given below carefully. And answer the following questions:



- Prediction from the Algorithmic Decision Making by AI for Melanoma: 84%.
- The spot on the skin has been growing in size and the borders of the spot are irregular.

Make a prediction

What, according to you, is the probability that the spot on skin in the image shown above is Melanoma? (in %)

0%

100%

Make a decision

A biopsy should be performed on the patient

No biopsy should be performed on the patient.

How sure are you of your decision?

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Not at all sure/sehr unsicher	Not sure/unsicher	Moderate/moderat	Sure/sicher	Very Sure/sehr sicher

How reliable do you consider the recommendation of the algorithm to be?

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Not at all reliable/sehr unzuverlässig	Not Reliable/ unzuverlässig	Moderate/moderat	Reliable/zuverlässig	Very Reliable/sehr zuverlässig

Seite 07**CN1****Please look at the image and read the information given below carefully. And answer the following questions:**

- Prediction from the Algorithmic Decision Making by AI for Melanoma: 12%.
- The spot on the skin has faded over time.

Make a prediction

What, according to you, is the probability that the spot on skin in the image shown above is Melanoma?(in %)

0%

100%

Make a decision

A biopsy should be performed on the patient

No biopsy should be performed on the patient.

How sure are you of your decision?

Not at all
sure/sehr unsicher

Not sure/unsicher

Moderate/moderat

Sure/sicher

Very Sure/sehr
sicher

How reliable do you consider the recommendation of the algorithm to be?

Not at all
reliable/sehr
unzuverlässig

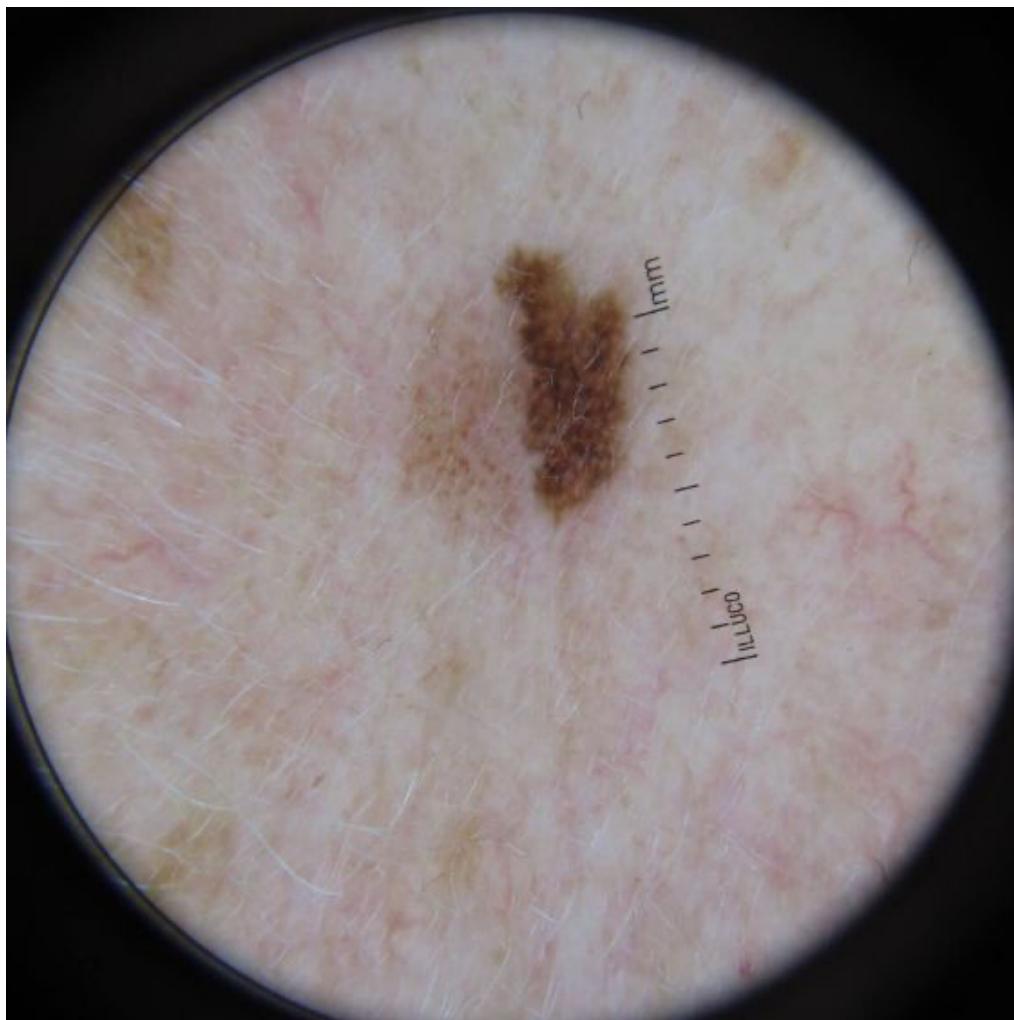
Not Reliable/
unzuverlässig

Moderate/moderat

Reliable/zuverlässig

Very Reliable/sehr
zuverlässig

Please look at the image and read the information given below carefully. And answer the following questions:



- Prediction from the Algorithmic Decision Making by AI for Melanoma: 54%
- The spot on the skin has slightly increased in 4 months (6.2mm to 6.6mm).

Make a prediction

What, according to you, is the probability that the spot on skin in the image shown above is Melanoma?(in %)

0%

100%

Make a decision

A biopsy should be performed on the patient

No biopsy should be performed on the patient.

How sure are you of your decision?

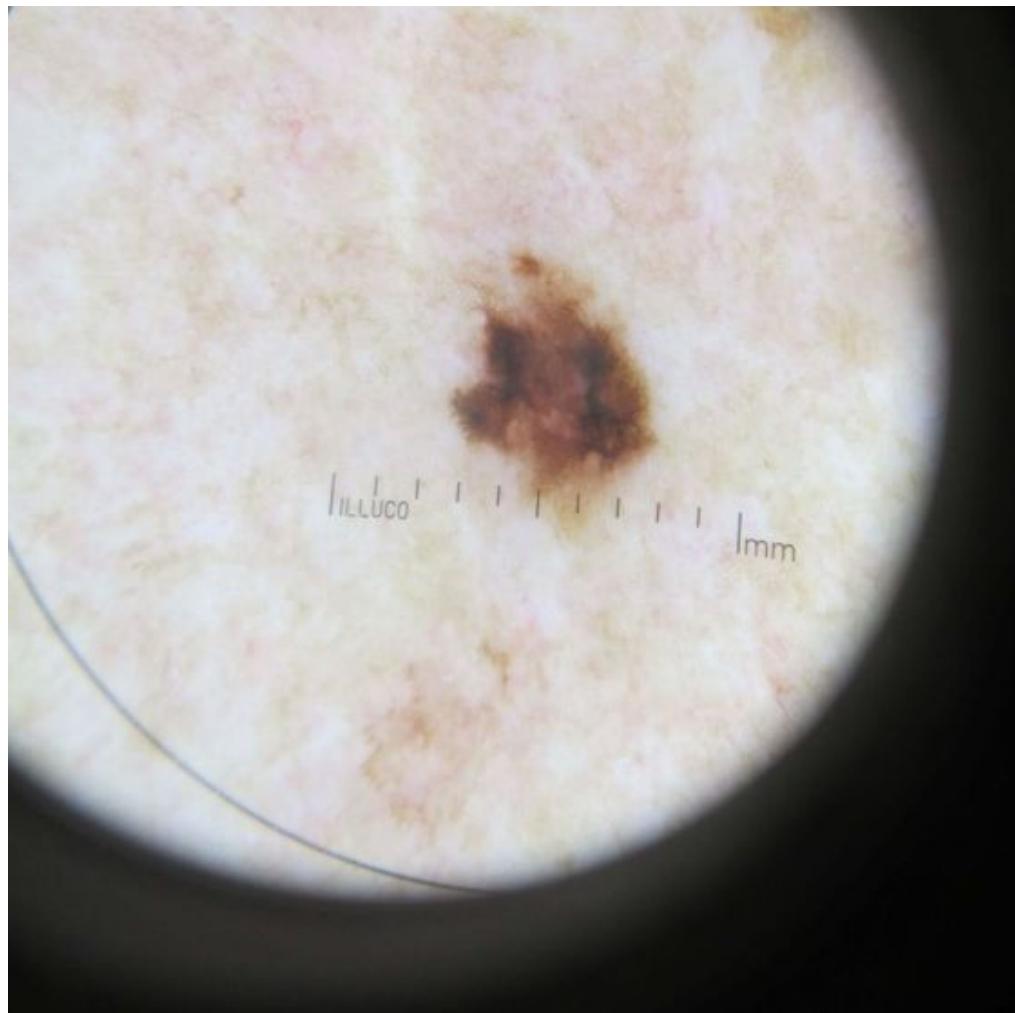
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Not at all sure/sehr unsicher	Not sure/unsicher	Moderate/moderat	Sure/sicher	Very Sure/sehr sicher

How reliable do you consider the recommendation of the algorithm to be?

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Not at all reliable/sehr unzuverlässig	Not Reliable/ unzuverlässig	Moderate/moderat	Reliable/zuverlässig	Very Reliable/sehr zuverlässig

Seite 09

CP2

Please look at the image and read the information given below carefully. And answer the following questions:

- Prediction from the Algorithmic Decision Making by AI for Melanoma: 77%.
- The size of the spot is 7mm, and the color is brown. The spot is also itchy.

Make a prediction

What, according to you, is the probability that the spot on skin in the image shown above is Melanoma?(in %)

0%

100%

Make a decision



A biopsy should be performed on the patient



No biopsy should be performed on the patient.

How sure are you of your decision?



Not at all
sure/sehr unsicher



Not sure/unsicher



Moderate/moderat



Sure/sicher



Very Sure/sehr
sicher

How reliable do you consider the recommendation of the algorithm to be?



Not at all
reliable/sehr
unzuverlässig



Not Reliable/
unzuverlässig



Moderate/moderat



Reliable/zuverlässig



Very Reliable/sehr
zuverlässig

Please look at the image and read the information given below carefully. And answer the following questions:



- Prediction from the Algorithmic Decision Making by AI for Melanoma: 6%
- No additional information was available for this case.

Make a prediction

What, according to you, is the probability that the spot on skin in the image shown above is Melanoma?(in %)

0%

100%

Make a decision

A biopsy should be performed on the patient

No biopsy should be performed on the patient.

How sure are you of your decision?

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Not at all sure/sehr unsicher	Not sure/unsicher	Moderate/moderat	Sure/sicher	Very Sure/sehr sicher

How reliable do you consider the recommendation of the algorithm to be?

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Not at all reliable/sehr unzuverlässig	Not Reliable/ unzuverlässig	Moderate/moderat	Reliable/zuverlässig	Very Reliable/sehr zuverlässig

Seite 11
CN3

Please look at the image and read the information given below carefully. And answer the following questions:



- Prediction from the Algorithmic Decision Making by AI for Melanoma: 14%
- There is no irritation and no color variation, the size is less than 5mm.

Make a prediction

What, according to you, is the probability that the spot on skin in the image shown above is Melanoma?(in %)

0%

100%

Make a decision



A biopsy should be performed on the patient



No biopsy should be performed on the patient.

How sure are you of your decision?



Not at all
sure/sehr unsicher



Not sure/unsicher



Moderate/moderat



Sure/sicher



Very Sure/sehr
sicher

How reliable do you consider the recommendation of the algorithm to be?



Not at all
reliable/sehr
unzuverlässig



Not Reliable/
unzuverlässig



Moderate/moderat



Reliable/zuverlässig



Very Reliable/sehr
zuverlässig

Please look at the image and read the information given below carefully. And answer the following questions:



- Prediction from the Algorithmic Decision Making by AI for Melanoma: 48%
- No additional information was available for this case.

Make a prediction

What, according to you, is the probability that the spot on skin in the image shown above is Melanoma?(in %)

0%

100%

Make a decision

A biopsy should be performed on the patient

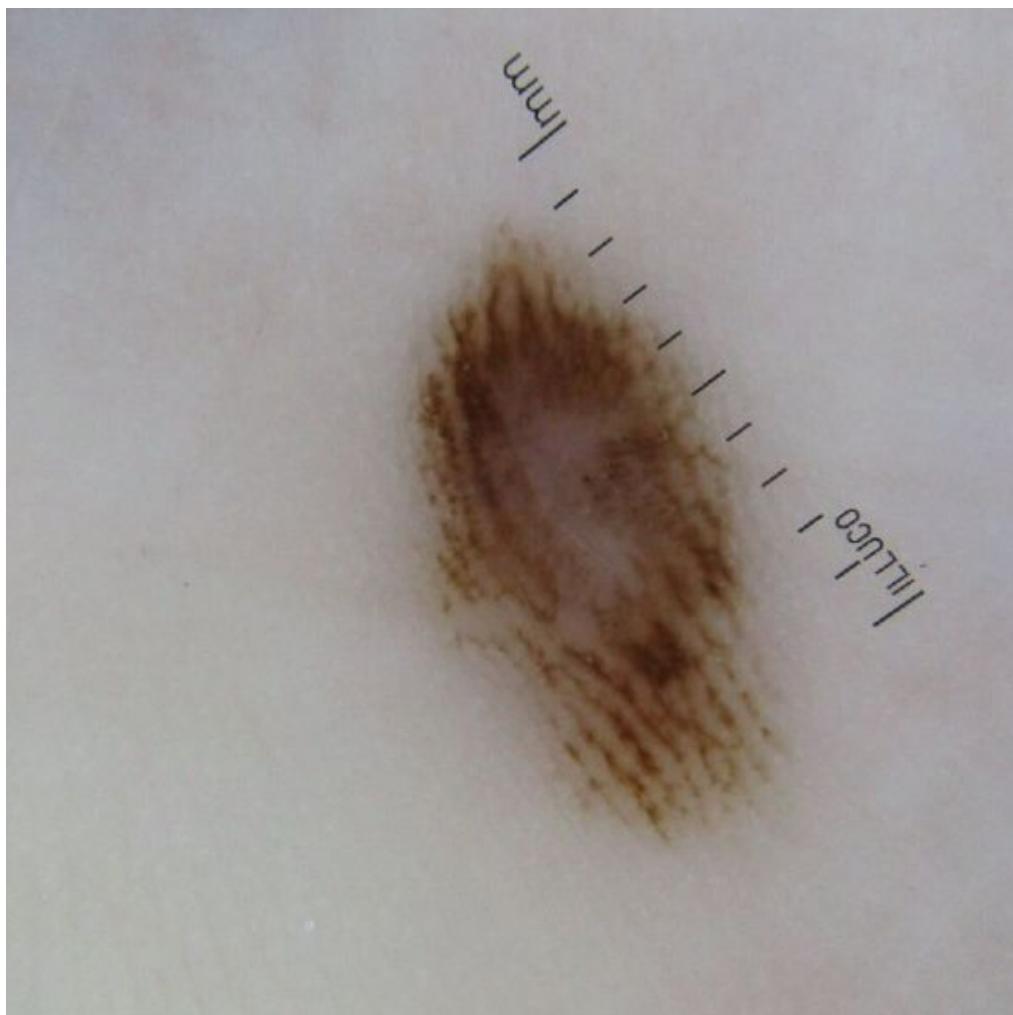
No biopsy should be performed on the patient.

How sure are you of your decision?

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Not at all sure/sehr unsicher	Not sure/unsicher	Moderate/moderat	Sure/sicher	Very Sure/sehr sicher

How reliable do you consider the recommendation of the algorithm to be?

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Not at all reliable/sehr unzuverlässig	Not Reliable/ unzuverlässig	Moderate/moderat	Reliable/zuverlässig	Very Reliable/sehr zuverlässig

Seite 13**A3****Please look at the image and read the information given below carefully. And answer the following questions:**

- Prediction from the Algorithmic Decision Making by AI for Melanoma: 51%.
- The color of lesion has become darker over a period of 2 years.

Make a prediction

What, according to you, is the probability that the spot on skin in the image shown above is Melanoma?(in %)

0%

100%

Make a decision



A biopsy should be performed on the patient



No biopsy should be performed on the patient.

How sure are you of your decision?



Not at all
sure/sehr unsicher



Not sure/unsicher



Moderate/moderat



Sure/sicher



Very Sure/sehr
sicher

How reliable do you consider the recommendation of the algorithm to be?



Not at all
reliable/sehr
unzuverlässig



Not Reliable/
unzuverlässig



Moderate/moderat

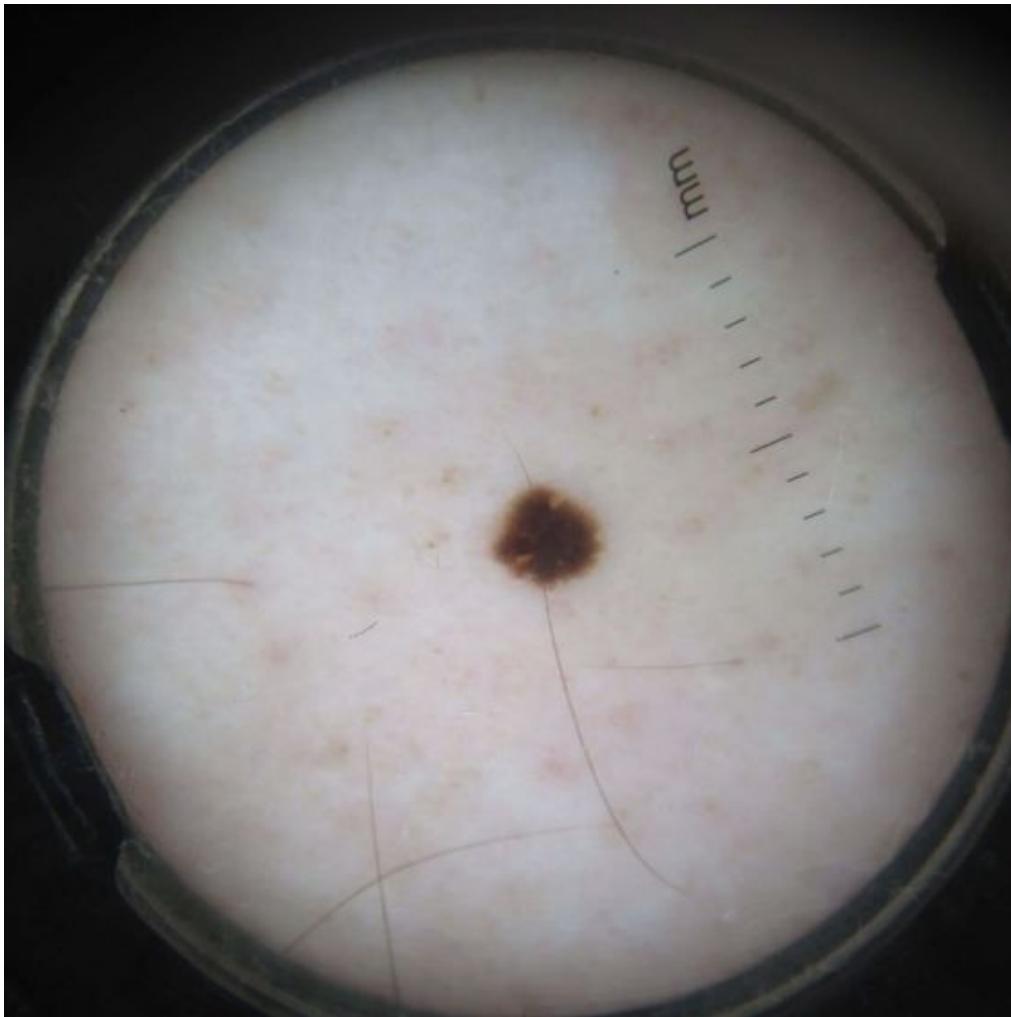


Reliable/zuverlässig



Very Reliable/sehr
zuverlässig

Please look at the image and read the information given below carefully. And answer the following questions:



- Prediction from the Algorithmic Decision Making by AI for Melanoma: 75%.
- The color of the lesion has become darker and the size has slightly grown.

Make a prediction

What, according to you, is the probability that the spot on skin in the image shown above is Melanoma?(in %)

0%

100%

Make a decision

A biopsy should be performed on the patient

No biopsy should be performed on the patient.

How sure are you of your decision?

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Not at all sure/sehr unsicher	Not sure/unsicher	Moderate/moderat	Sure/sicher	Very Sure/sehr sicher

How reliable do you consider the recommendation of the algorithm to be?

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Not at all reliable/sehr unzuverlässig	Not Reliable/ unzuverlässig	Moderate/moderat	Reliable/zuverlässig	Very Reliable/sehr zuverlässig

Seite 15

CN4

Please look at the image and read the information given below carefully. And answer the following questions:

- Prediction from the Algorithmic Decision Making by AI for Melanoma: 23%.
- The spot has smooth even borders and the pigmented component fades towards outside.

Make a prediction

What, according to you, is the probability that the spot on skin in the image shown above is Melanoma?(in %)

0%

100%

Make a decision



A biopsy should be performed on the patient



No biopsy should be performed on the patient.

How sure are you of your decision?



Not at all
sure/sehr unsicher



Not sure/unsicher



Moderate/moderat



Sure/sicher



Very Sure/sehr
sicher

How reliable do you consider the recommendation of the algorithm to be?



Not at all
reliable/sehr
unzuverlässig



Not Reliable/
unzuverlässig



Moderate/moderat

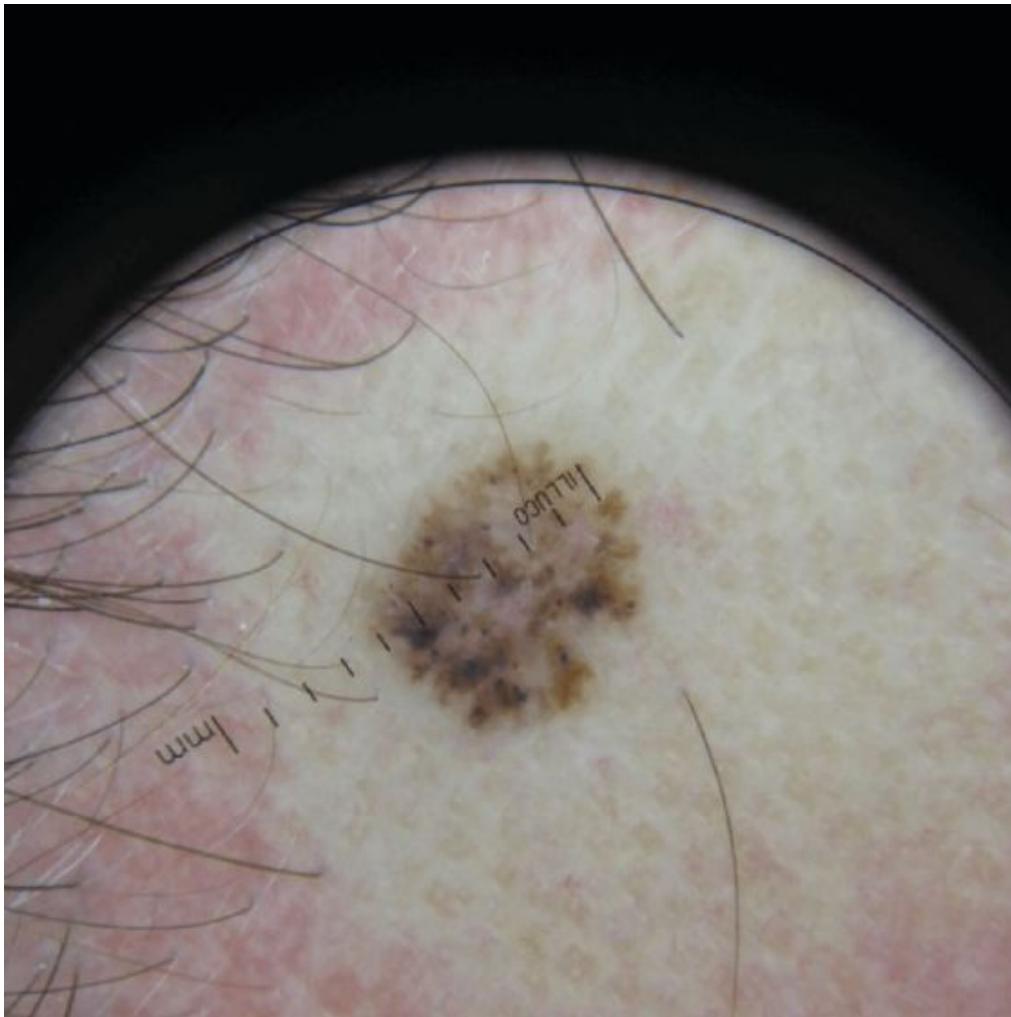


Reliable/zuverlässig



Very Reliable/sehr
zuverlässig

Please look at the image and read the information given below carefully. And answer the following questions:



- Prediction from the Algorithmic Decision Making by AI for Melanoma: 47%.
- The spot has an approximate diameter of 5.9mm

Make a prediction

What, according to you, is the probability that the spot on skin in the image shown above is Melanoma?(in %)

0%

100%

Make a decision

A biopsy should be performed on the patient

No biopsy should be performed on the patient.

How sure are you of your decision?

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Not at all sure/sehr unsicher	Not sure/unsicher	Moderate/moderat	Sure/sicher	Very Sure/sehr sicher

How reliable do you consider the recommendation of the algorithm to be?

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Not at all reliable/sehr unzuverlässig	Not Reliable/ unzuverlässig	Moderate/moderat	Reliable/zuverlässig	Very Reliable/sehr zuverlässig

Please look at the image and read the information given below carefully. And answer the following questions:



- Prediction from the Algorithmic Decision Making by AI for Melanoma: 49%.
- The size of the spot on the skin hasn't grown for over 2 years, but it has irregular boundaries and dark color.

Make a prediction

What, according to you, is the probability that the spot on skin in the image shown above is Melanoma?(in %)

0%

100%

Make a decision

A biopsy should be performed on the patient

No biopsy should be performed on the patient.

How sure are you of your decision?

Not at all
sure/sehr unsicher

Not sure/unsicher

Moderate/moderat

Sure/sicher

Very Sure/sehr
sicher

How reliable do you consider the recommendation of the algorithm to be?

Not at all
reliable/sehr
unzuverlässig

Not Reliable/
unzuverlässig

Moderate/moderat

Reliable/zuverlässig

Very Reliable/sehr
zuverlässig

Please look at the image and read the information given below carefully. And answer the following questions:



- Prediction from the Algorithmic Decision Making by AI for Melanoma: 96%.
- The spot has become firm and is continuously growing in size.

Make a prediction

What, according to you, is the probability that the spot on skin in the image shown above is Melanoma?(in %)

0%

100%

Make a decision

A biopsy should be performed on the patient

No biopsy should be performed on the patient.

How sure are you of your decision?

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Not at all sure/sehr unsicher	Not sure/unsicher	Moderate/moderat	Sure/sicher	Very Sure/sehr sicher

How reliable do you consider the recommendation of the algorithm to be?

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Not at all reliable/sehr unzuverlässig	Not Reliable/ unzuverlässig	Moderate/moderat	Reliable/zuverlässig	Very Reliable/sehr zuverlässig

Seite 19

CN5

Please look at the image and read the information given below carefully. And answer the following questions:

- Prediction from the Algorithmic Decision Making by AI for Melanoma: 4%.
- No change in size, color, shape, or structure noted.

Make a prediction

What, according to you, is the probability that the spot on skin in the image shown above is Melanoma?(in %)

0%

100%

Make a decision



A biopsy should be performed on the patient



No biopsy should be performed on the patient.

How sure are you of your decision?



Not at all
sure/sehr unsicher



Not sure/unsicher



Moderate/moderat



Sure/sicher



Very Sure/sehr
sicher

How reliable do you consider the recommendation of the algorithm to be?



Not at all
reliable/sehr
unzuverlässig



Not Reliable/
unzuverlässig



Moderate/moderat



Reliable/zuverlässig



Very Reliable/sehr
zuverlässig

Please look at the image and read the information given below carefully. And answer the following questions:



- Prediction from the Algorithmic Decision Making by AI for Melanoma: 97%.
- The spot has been itchy and shows no signs of healing. The skin has become darker around the spot, with blood clots forming around the spot.

Make a prediction

What, according to you, is the probability that the spot on skin in the image shown above is Melanoma?(in %)

0%

100%

Make a decision

A biopsy should be performed on the patient

No biopsy should be performed on the patient.

How sure are you of your decision?

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Not at all sure/sehr unsicher	Not sure/unsicher	Moderate/moderat	Sure/sicher	Very Sure/sehr sicher

How reliable do you consider the recommendation of the algorithm to be?

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Not at all reliable/sehr unzuverlässig	Not Reliable/ unzuverlässig	Moderate/moderat	Reliable/zuverlässig	Very Reliable/sehr zuverlässig

Seite 21**POSTSURVEY**

You have now processed all 15 cases. Please answer the questions below.

Please indicate to what extent you agree with the following statements.

	completely disagree	strongly disagree	somewhat disagree	undecided	somewhat agree	strongly agree	completely agree
I think I know how the algorithm works.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I think I have a good grasp of the algorithm.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I think I know how to use the algorithm.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

	completely disagree	strongly disagree	disagree	undecided	agree	strongly agree	completely disagree
I have largely ignored the algorithm in my decisions	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The algorithm was very helpful in the decision making process.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have incorporated the recommendations of the algorithm into my decision-making process.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I found the recommendations of the algorithm reasonable.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The recommendations of the algorithm were in line with my assessment.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The recommendations of the algorithm were easy to understand.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
In my opinion the algorithm did not give good recommendations.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The algorithm made errors.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The algorithm was unreliable.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Please indicate how much you agree with the following :

“The decisions of algorithm influenced my decisions”

<input type="radio"/>				
Strongly Disagree	Somewhat Disagree	Moderate	Somewhat Agree	Strongly agree

Please indicate how fair you consider the algorithm to be:

<input type="radio"/>				
Very Unfair	Somewhat Unfair	Moderate	Somewhat Fair	Strongly Fair

I think it is good when decision-makers in the medical system receive assistance from algorithm-based recommendation systems.

Please indicate how much you agree with the following statement.

<input type="radio"/>				
Strongly Disagree	Somewhat Disagree	Moderate	Somewhat Agree	Strongly Agree

Algorithm-based recommendation systems are also used in many other areas. We have listed four of these use cases below.

Please sort them according to their severity. The most serious use case should be placed on rank 1.

There are two ways to sort the terms. Either (a) you drag the cards to a free rank or (b) you click on them one after another with a double click.

*Triage = The prioritization of medical aid in case of resource shortage. Example: Who will receive life-sustaining treatments during the Covid-19 pandemic and who will not.

Diagnosis of skin cancer	Selection of applicants in a company	1
Decisions in the criminal justice system	Recommendations in a dating app	2
Triage*		3
		4
		5

Please sort the use cases now according to which case you would most likely agree to the use of a recommendation system.

Diagnosis of skin cancer	Selection of applicants in a company	1
Decisions in the criminal justice system	Recommendations in a dating app	2
Triage*		3
		4
		5

In the following, we would like to collect some demographic information

5. Please indicate your gender**6. Please indicate the age group you belong to.**

[Please choose] 

7. In what setting did you learn the basics of the ABCDE method?

For example, course module name, event or training.

8. Please name the institution where you learned the basics of the ABCDE method.

For example, name of the university, clinic or practice.

9. Please indicate your field of study.**10. If you are currently studying, please indicate your study semester.****11. How would you rate your knowledge in the following areas?****Knowledge in the area of**

non-existent minimal moderate advanced very advanced

Computer Science (general)

Machine Learning

Dermatology

12. Please state your general opinion on the use of algorithm-based recommendation systems.

I am more **in favor** of the use of algorithm-based recommendation systems.



I am rather **against** the use of algorithm-based recommendation systems.

Please give a short justification for your decision.

Why:

Seite 24
DEB

Thank you very much!

Thank you for taking the time to participate in this survey.

At the end of the survey we would like to inform you about the background of our scientific question.

This study investigates to what extent people include the decision template of an algorithm in their own decision making. In this study we investigate the influence of:

1) little as opposed to detailed information about the algorithm on the probability that people will include the algorithm's recommendation in their decision making.

2) the reliability of the algorithm (algorithm makes no mistakes or makes mistakes) on the probability that people will include the algorithm's recommendation in their decision.

They were given detailed information and the algorithm made no mistakes.

To ensure that all study participants receive the same cases and recommendations, no real algorithm was used in this study. The cooperation with the Department of Computer Science of the RWTH Aachen University and the Department of Dermatology of the RWTH Aachen University does not exist and served to make the cases appear credible.

Even though no algorithm was developed in the context of this survey, algorithms of this kind exist with a very good prediction probability. Our investigation helps to find out how such systems can be used responsibly by experts in the future.

Should you discover conspicuous skin marks on yourself after this study, please contact your family doctor or a dermatologist.

If you have further questions about the content of this study, please send us an email with the subject "Question about the Derma Algorithm Study" to itec@humtec.rwth-aachen.de.

On the following page, you have the opportunity to participate in the lottery mentioned at the beginning of the survey. Thank you again for taking the time to support us in our research.

- I would like to participate in the **lottery**. I agree that my e-mail address will be saved until the winner is drawn. My interview will continue to be anonymous and my email address will not be passed on to third parties.

TEST

Thank you very much for your participation!

We would like to thank you very much for your support.

Your answers have been saved, you can now close the browser window.

[B.Sc. Sören Schöder](#), [B. Eng. Sourabh Zanwar](#), [Prof. Astrid Rosenthal-von der Pütten](#), RWTH Aachen

We would like to know how you generally think about algorithms.

Einleitung Attitudes**AA02****Please indicate how much you agree with the following statements.**

I think..	completely disagree	strongly disagree	somewhat disagree	undecided	somewhat agree	strongly agree	completely agree
Algorithms can relieve people of difficult decisions.	<input type="radio"/>						
Algorithms cannot consider the consequences of a decision.	<input type="radio"/>						
algorithm-based decisions are too impersonal for me.	<input type="radio"/>						
that certain decisions should only be made by people.	<input type="radio"/>						
Algorithms prefer no one.	<input type="radio"/>						
that many people would simply follow algorithm-based recommendations.	<input type="radio"/>						
Algorithms can save decision makers a lot of work.	<input type="radio"/>						
Algorithms have no prejudices.	<input type="radio"/>						
algorithm-based decisions are too uncertain for me.	<input type="radio"/>						
it is problematic that algorithms cannot be held responsible	<input type="radio"/>						
Algorithms have no good and no bad days.	<input type="radio"/>						
Recommendations by an algorithm lead to people thinking less about decisions.	<input type="radio"/>						
Algorithms can process more data than a human.	<input type="radio"/>						
Algorithms can analyze data faster than a human.	<input type="radio"/>						

Algorithms apply the same scale to everyone.	<input type="radio"/>						
Algorithms make decisions more responsibly than humans.	<input type="radio"/>						
Algorithms can not be held responsible.	<input type="radio"/>						
Algorithms should not make morally difficult decisions.	<input type="radio"/>						
Algorithms treat all people equally.	<input type="radio"/>						
algorithm-based decisions are not transparent.	<input type="radio"/>						
People could let themselves be determined by algorithms.	<input type="radio"/>						
Algorithms are not suitable for making personal decisions.	<input type="radio"/>						
Algorithms are less flexible than humans in evaluating decision factors.	<input type="radio"/>						
Algorithms are not aware of the responsibility of a decision.	<input type="radio"/>						
Algorithms can make decisions that no human should have to make.	<input type="radio"/>						
Algorithms cannot be bribed.	<input type="radio"/>						
Algorithms can make more precise decisions than a human.	<input type="radio"/>						
Algorithms are completely rational and therefore comprehensible.	<input type="radio"/>						

This algorithm was developed with the "ABCDE method" in mind. This is a common method for ~~detecting melanoma~~ by visual inspection. It can help both medical laypersons and clinicians to identify features in a skin lesion that could indicate melanoma.

ABCDEs Introduction

Please read the following text carefully.

ABCDE Method

Asymmetry:

A stands for asymmetry. A melanocytic nevus (harmless mole) is usually symmetrical, while a melanoma often has an irregular or asymmetrical shape.

Border Irregularity:

B stands for Border Irregularity, i.e. irregularity at the border. A melanocytic nevus (harmless liver spot) has smooth and even edges, while a melanoma often has irregular and difficult to define edges.

Color Variation:

C stands for Color Variation. A melanocytic nevus (harmless mole) usually has a single shade or two shades, one entering the other or repeating regularly (generally pink, brown, or tan). Melanoma can be brown but can have up to five or six colors (blue, black, brown, tan, grey, pink, and red). These colors are unevenly or irregularly distributed.

Diameter greater than 6mm:

D stands for diameter. Most melanomas have a diameter of more than 6 mm when they are diagnosed.

Evolution:

E stands for Evolution or change. A melanocytic nevus (harmless mole) is usually stable and does not change in size, shape, or color, whereas a melanoma changes over time. Changes in size, color, shape, or structure can become noticeable over months to years.

AC03

1. The following are the diameters measured for spots on 5 different people for the diagnosis of Melanoma. Which is most likely to be a Melanoma based on the diameter?

- 4.0 mm
- 5.1 mm
- 6.9 mm
- 5.5 mm
- 3.6 mm

Please read the following task description carefully.

Introduction

We would like to ask you to put yourself in the role of the treating dermatologist.

Based on the information available to you, you will decide how likely the birthmark could be melanoma and whether you would have a biopsy performed.

Below are some cases where you should make this decision. In the example below you can see that the following information is available to you:

- You can analyze the image of the melanoma using the well-known ABCDE formula.
- You can ask the patient whether the melanoma has changed over time in terms of color, shape or size.
- You can ask the patient whether there has been any itching or bleeding at the birthmark.

Example: Image of a birthmark for analysis with accompanying information



- The spot on the skin has slightly increased in size over the last four months (6.2mm to 6.6mm). It does not itch or bleed.

Demo case

Seite 04

MANIPUL

In addition to the case descriptions you will receive a risk assessment of our algorithm in the following form.

Transparency_All

The risk assessment algorithm predicts that there is a 20% probability that this mole is a melanoma.

This algorithm was developed by us to help dermatologists in the assessment of moles.

Transparency_Low

2. From the information provided on the previous pages, I can say that:

AC02

For the survey...

- I'm not getting a recommendation from an algorithm
- I get a recommendation from a person
- I get a recommendation from an algorithm

Seite 05

EXPEVAL

Please evaluate the information you have received about the algorithm:

BE01

The explanation was...

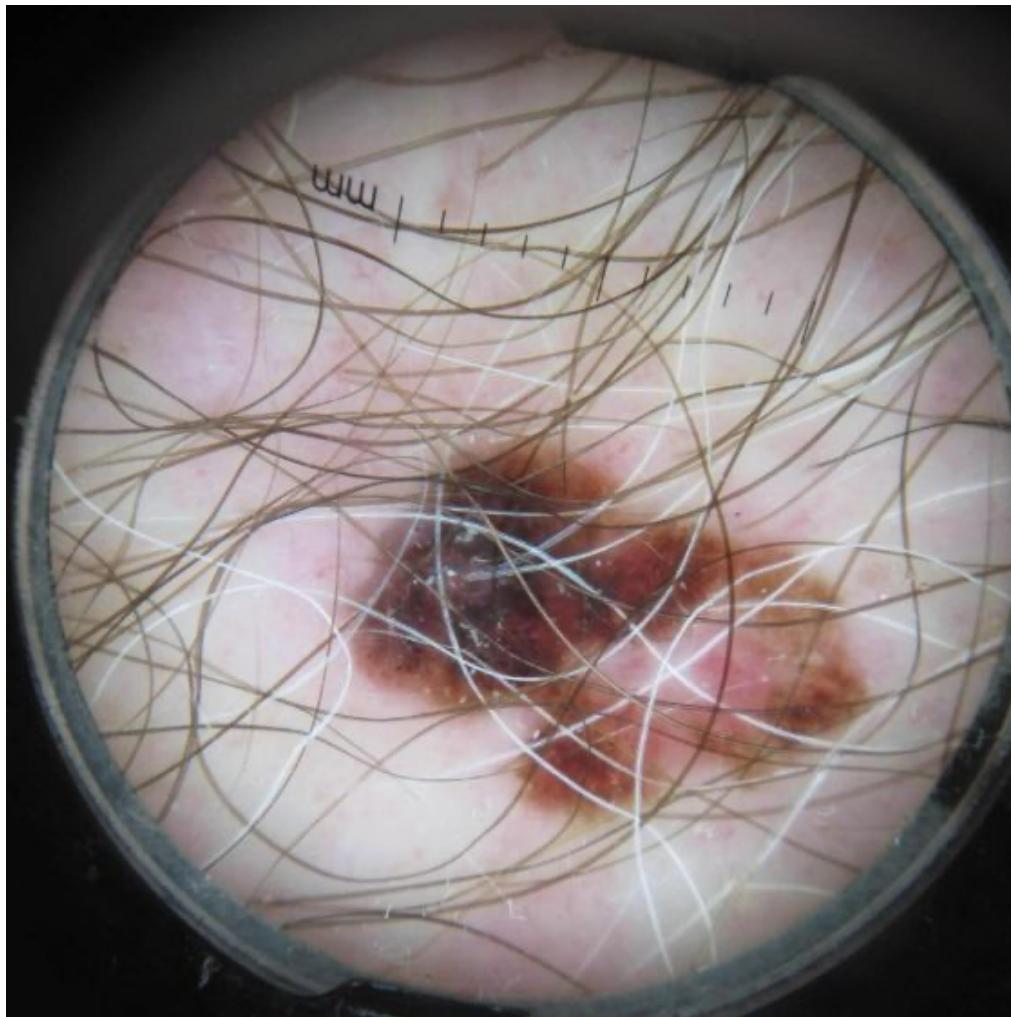
not understandable	<input type="radio"/>	understandable						
hard to understand	<input type="radio"/>	easily understandable						
misleading	<input type="radio"/>	unambiguous						
not useful	<input type="radio"/>	helpful						
uninformative	<input type="radio"/>	informative						
overwhelming	<input type="radio"/>	Underchallenging						
very short	<input type="radio"/>	very long						

Please indicate to what extent you agree with the following statements.

BE02

	completely disagree	strongly disagree	somewhat disagree	undecided	somewhat agree	strongly agree	completely agree
I think I know how to use the algorithm.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I think I know how the algorithm works.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I think I have a good grasp of the algorithm.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Please look at the image and read the information given below carefully. And answer the following questions.



- Prediction from the Algorithmic Decision Making by AI for Melanoma: 84%.
- The spot on the skin has been growing in size and the borders of the spot are irregular.

CP1

Make a prediction

PR01

What, according to you, is the probability that the spot on skin in the image shown above is Melanoma? (in %)

0%

100%

DC01

Make a decision A biopsy should be performed on the patient No biopsy should be performed on the patient.**How sure are you of your decision?****CS01** Not at all
sure/sehr unsicher Not sure/unsicher Moderate/moderat Sure/sicher Very Sure/sehr
sicher**How reliable do you consider the recommendation of the algorithm to be?****CA01** Not at all
reliable/sehr
unzuverlässig Not Reliable/
unzuverlässig Moderate/moderat Reliable/zuverlässig Very Reliable/sehr
zuverlässig

Seite 07

CN1

Please look at the image and read the information given below carefully. And answer the following questions.**00STYLE_Resi Case title**



- Prediction from the Algorithmic Decision Making by AI for Melanoma: 12%.
- The spot on the skin has faded over time.

CN1**Make a prediction****PRO2**

What, according to you, is the probability that the spot on skin in the image shown above is Melanoma?(in %)

0%

100%

Make a decision**DC02** 

A biopsy should be performed on the patient



No biopsy should be performed on the patient.

CS02

How sure are you of your decision?

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Not at all sure/sehr unsicher	Not sure/unsicher	Moderate/moderat	Sure/sicher	Very Sure/ sehr sicher

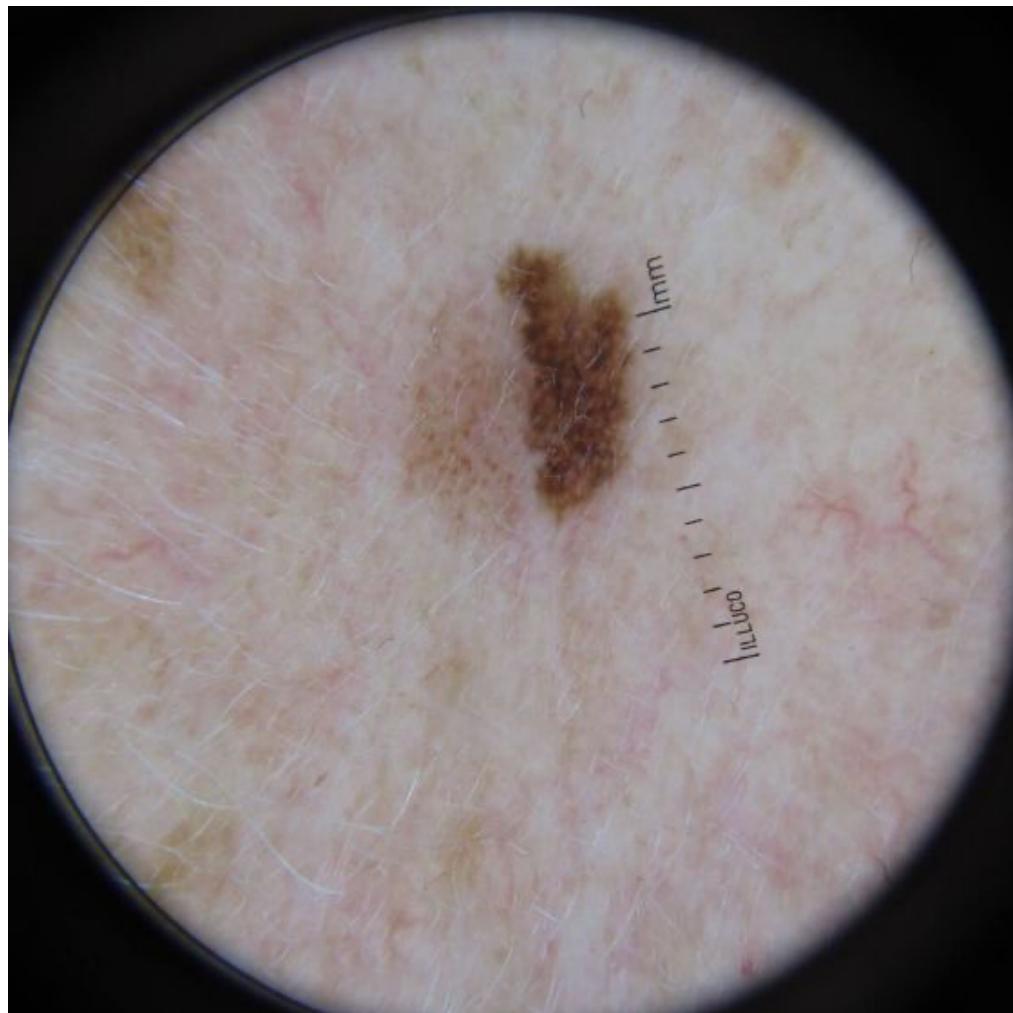
How reliable do you consider the recommendation of the algorithm to be?**CA02**

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Not at all reliable/sehr unzuverlässig	Not Reliable/ unzuverlässig	Moderate/moderat	Reliable/zuverlässig	Very Reliable/ sehr zuverlässig

Seite 08

A1

Please look at the image and read the information given below carefully. And answer the **Case title**.



- Prediction from the Algorithmic Decision Making by AI for Melanoma: 54%
- The spot on the skin has slightly increased in 4 months (6.2mm to 6.6mm).

Amb1**Make a prediction****PRO3**

What, according to you, is the probability that the spot on skin in the image shown above is Melanoma?(in %)

0%

100%

Make a decision**DC03**

A biopsy should be performed on the patient



No biopsy should be performed on the patient.

How sure are you of your decision?**CS03**Not at all
sure/sehr unsicher

Not sure/unsicher



Moderate/moderat



Sure/sicher

Very Sure/sehr
sicher**How reliable do you consider the recommendation of the algorithm to be?****CA03**Not at all
reliable/sehr
unzuverlässigNot Reliable/
unzuverlässig

Moderate/moderat



Reliable/zuverlässig

Very Reliable/sehr
zuverlässig**Seite 09**

FP1

Please look at the image and read the information given below carefully. And answer the following questions.



- Prediction from the Algorithmic Decision Making by AI for Melanoma: 79%.
- The spot has smooth even borders and the pigmented component fades towards outside.

FP1**Make a prediction****PRO4**

What, according to you, is the probability that the spot on skin in the image shown above is Melanoma?(in %)

0%

100%

Make a decision**DC04**

A biopsy should be performed on the patient



No biopsy should be performed on the patient.

CS04

How sure are you of your decision?

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Not at all sure/sehr unsicher	Not sure/unsicher	Moderate/moderat	Sure/sicher	Very Sure/sehr sicher

How reliable do you consider the recommendation of the algorithm to be?**CA04**

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Not at all reliable/sehr unzuverlässig	Not Reliable/ unzuverlässig	Moderate/moderat	Reliable/zuverlässig	Very Reliable/sehr zuverlässig

Seite 10**CN2**

Please look at the image and read the information given below carefully. And answer the **Case title**.



- Prediction from the Algorithmic Decision Making by AI for Melanoma: 6%
- No additional information was available for this case.

CN2**Make a prediction****PRO5**

What, according to you, is the probability that the spot on skin in the image shown above is Melanoma?(in %)

0%

100%

Make a decision**DC05**

A biopsy should be performed on the patient



No biopsy should be performed on the patient.

How sure are you of your decision?**CS05**Not at all
sure/sehr unsicher

Not sure/unsicher



Moderate/moderat



Sure/sicher

Very Sure/sehr
sicher**How reliable do you consider the recommendation of the algorithm to be?****CA05**Not at all
reliable/sehr
unzuverlässigNot Reliable/
unzuverlässig

Moderate/moderat



Reliable/zuverlässig

Very Reliable/sehr
zuverlässig**Seite 11**

FN1

Please look at the image and read the information given below carefully. And answer the following questions.



- Prediction from the Algorithmic Decision Making by AI for Melanoma: 15%.
- The size of the spot is 7mm, and the borders of the spot are irregular.

FN1**Make a prediction****PRO6**

What, according to you, is the probability that the spot on skin in the image shown above is Melanoma?(in %)

0%

100%

Make a decision**DC06**

A biopsy should be performed on the patient

No biopsy should be performed on the patient.

CS06

How sure are you of your decision?

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Not at all sure/sehr unsicher	Not sure/unsicher	Moderate/moderat	Sure/sicher	Very Sure/sehr sicher

How reliable do you consider the recommendation of the algorithm to be?**CA06**

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Not at all reliable/sehr unzuverlässig	Not Reliable/ unzuverlässig	Moderate/moderat	Reliable/zuverlässig	Very Reliable/sehr zuverlässig

Seite 12**A2**

Please look at the image and read the information given below carefully. And answer the **Case title**.



- Prediction from the Algorithmic Decision Making by AI for Melanoma: 48%
- No additional information was available for this case.

Amb2**Make a prediction****PRO7**

What, according to you, is the probability that the spot on skin in the image shown above is Melanoma?(in %)

0%

100%

Make a decision**DC07**

A biopsy should be performed on the patient



No biopsy should be performed on the patient.

How sure are you of your decision?**CS07**Not at all
sure/sehr unsicher

Not sure/unsicher



Moderate/moderat



Sure/sicher

Very Sure/sehr
sicher**How reliable do you consider the recommendation of the algorithm to be?****CA07**Not at all
reliable/sehr
unzuverlässigNot Reliable/
unzuverlässig

Moderate/moderat

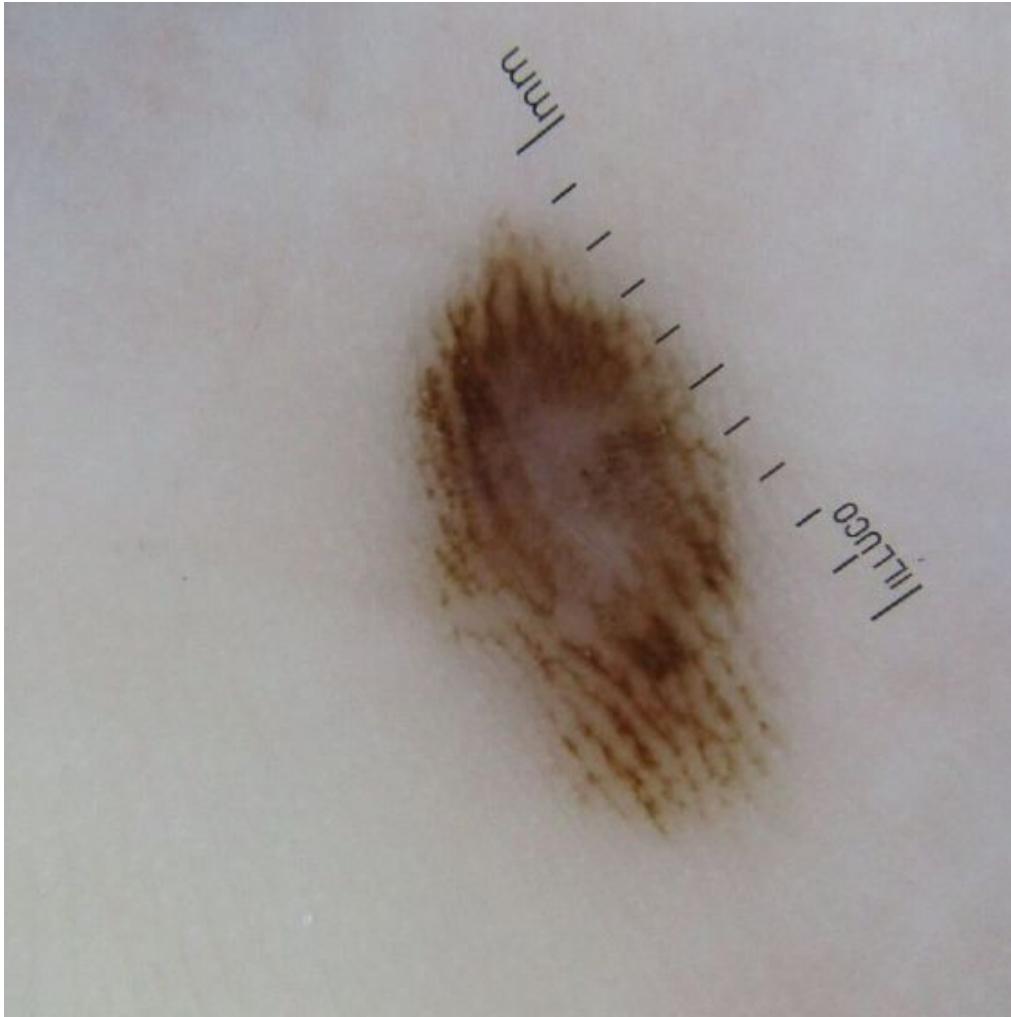


Reliable/zuverlässig

Very Reliable/sehr
zuverlässig**Seite 13**

A3

Please look at the image and read the information given below carefully. And answer the following questions.



- Prediction from the Algorithmic Decision Making by AI for Melanoma: 51%.
- The color of lesion has become darker over a period of 2 years.

Amb3**Make a prediction****PRO8**

What, according to you, is the probability that the spot on skin in the image shown above is Melanoma?(in %)

0%

100%

Make a decision**DC08**

A biopsy should be performed on the patient

No biopsy should be performed on the patient.

CS08

How sure are you of your decision?

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Not at all sure/sehr unsicher	Not sure/unsicher	Moderate/moderat	Sure/sicher	Very Sure/sehr sicher

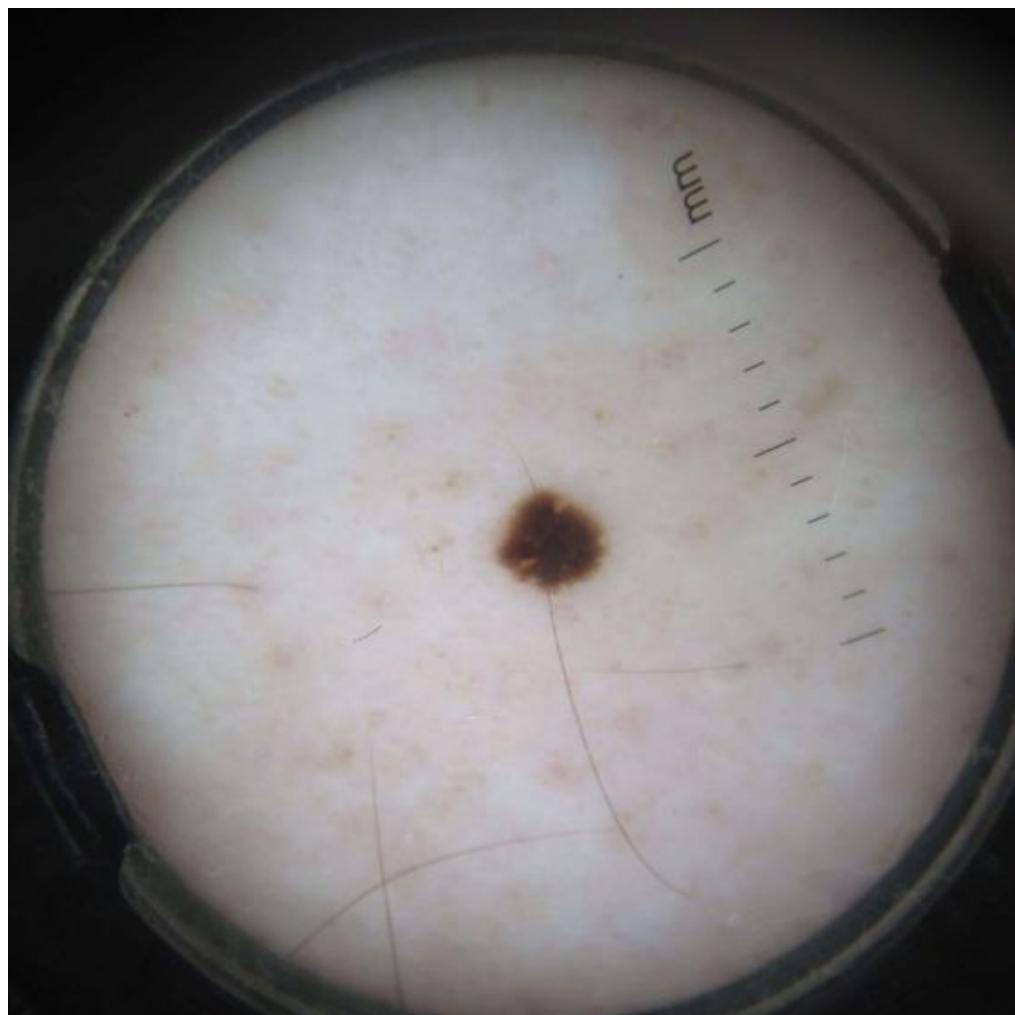
How reliable do you consider the recommendation of the algorithm to be?**CA08**

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Not at all reliable/sehr unzuverlässig	Not Reliable/ unzuverlässig	Moderate/moderat	Reliable/zuverlässig	Very Reliable/sehr zuverlässig

Seite 14

CP3

Please look at the image and read the information given below carefully. And answer the **Case title**.



- Prediction from the Algorithmic Decision Making by AI for Melanoma: 75%.
- The color of the lesion has become darker and the size has slightly grown.

CP3**Make a prediction****PR09**

What, according to you, is the probability that the spot on skin in the image shown above is Melanoma?(in %)

0%

100%

Make a decision**DC09**

A biopsy should be performed on the patient



No biopsy should be performed on the patient.

How sure are you of your decision?**CS09**Not at all
sure/sehr unsicher

Not sure/unsicher



Moderate/moderat



Sure/sicher

Very Sure/sehr
sicher**How reliable do you consider the recommendation of the algorithm to be?****CA09**Not at all
reliable/sehr
unzuverlässigNot Reliable/
unzuverlässig

Moderate/moderat



Reliable/zuverlässig

Very Reliable/sehr
zuverlässig**Seite 15**

FN2

Please look at the image and read the information given below carefully. And answer the following questions.

00STYLE_Res Case title



- Prediction from the Algorithmic Decision Making by AI for Melanoma: 43%.
- The color of the lesion has become darker, also it has become firmer. The size has slightly increased.

FN2**Make a prediction****PR10**

What, according to you, is the probability that the spot on skin in the image shown above is Melanoma?(in %)

0%

100%

Make a decision**DC10**

A biopsy should be performed on the patient



No biopsy should be performed on the patient.

CS10

How sure are you of your decision?

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Not at all sure/sehr unsicher	Not sure/unsicher	Moderate/moderat	Sure/sicher	Very Sure/sehr sicher

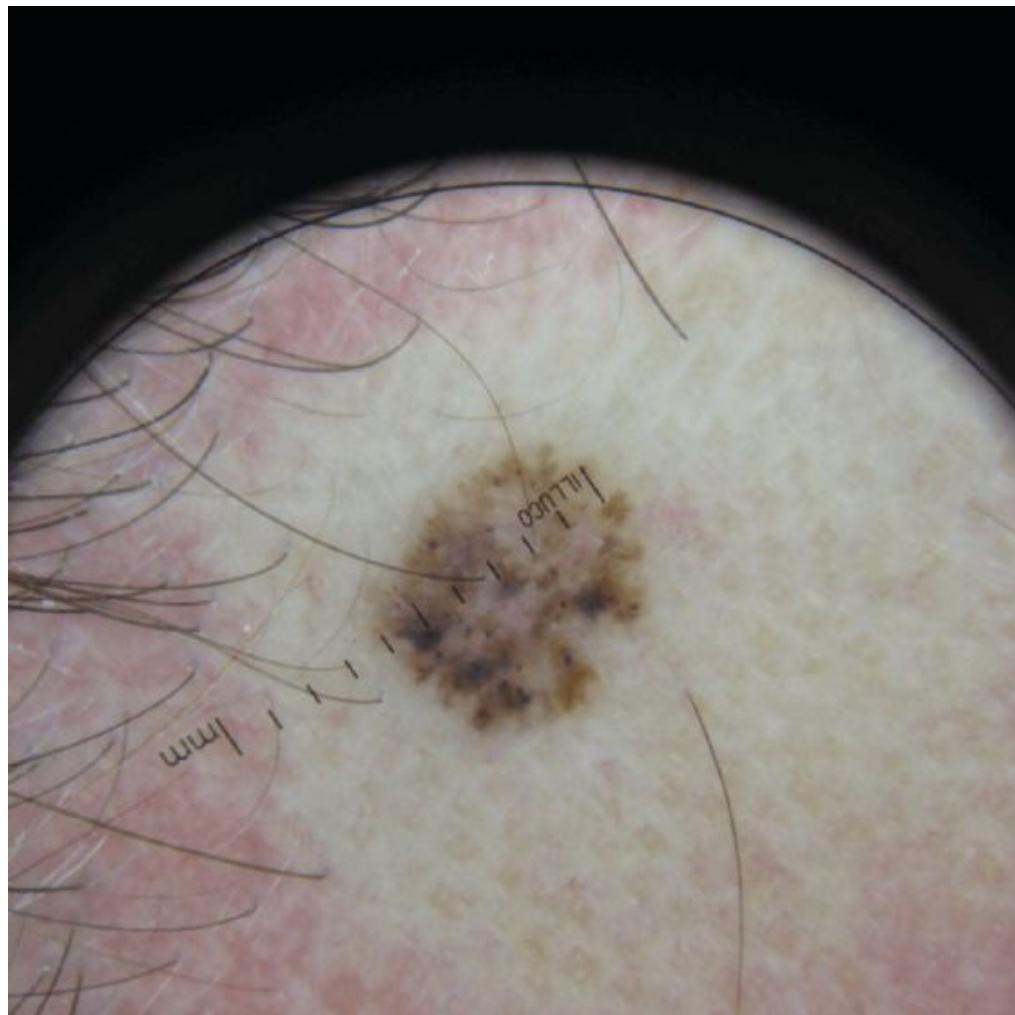
How reliable do you consider the recommendation of the algorithm to be?**CA10**

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Not at all reliable/sehr unzuverlässig	Not Reliable/ unzuverlässig	Moderate/moderat	Reliable/zuverlässig	Very Reliable/sehr zuverlässig

Seite 16

A4

Please look at the image and read the information given below carefully. And answer the **Case title**.



- Prediction from the Algorithmic Decision Making by AI for Melanoma: 47%.
- The spot has an approximate diameter of 5.9mm

Amb4**Make a prediction****PR11**

What, according to you, is the probability that the spot on skin in the image shown above is Melanoma?(in %)

0%

100%

Make a decision**DC11**

A biopsy should be performed on the patient



No biopsy should be performed on the patient.

How sure are you of your decision?**CS11**Not at all
sure/sehr unsicher

Not sure/unsicher



Moderate/moderat



Sure/sicher

Very Sure/sehr
sicher**How reliable do you consider the recommendation of the algorithm to be?****CA11**Not at all
reliable/sehr
unzuverlässigNot Reliable/
unzuverlässig

Moderate/moderat



Reliable/zuverlässig

Very Reliable/sehr
zuverlässig**Seite 17**

A5

Please look at the image and read the information given below carefully. And answer the following questions.



- Prediction from the Algorithmic Decision Making by AI for Melanoma: 49% Amb5
- The size of the spot on the skin hasn't grown for over 2 years, but it has irregular boundaries and dark color.

Make a predictionPR12

What, according to you, is the probability that the spot on skin in the image shown above is Melanoma?(in %)

0%

100%

Make a decisionDC12

A biopsy should be performed on the patient



No biopsy should be performed on the patient.

How sure are you of your decision?**CS12**

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Not at all sure/sehr unsicher	Not sure/unsicher	Moderate/moderat	Sure/sicher	Very Sure/sehr sicher

How reliable do you consider the recommendation of the algorithm to be?**CA12**

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Not at all reliable/sehr unzuverlässig	Not Reliable/ unzuverlässig	Moderate/moderat	Reliable/zuverlässig	Very Reliable/sehr zuverlässig

Seite 18**CP4**

Please look at the image and read the information given below carefully. And answer the **Case title**.



- Prediction from the Algorithmic Decision Making by AI for Melanoma: 96%.
- The spot has become firm and is continuously growing in size.

CP4**Make a prediction****PR13**

What, according to you, is the probability that the spot on skin in the image shown above is Melanoma?(in %)

0%

100%

Make a decision**DC13**

A biopsy should be performed on the patient



No biopsy should be performed on the patient.

How sure are you of your decision?**CS13**Not at all
sure/sehr unsicher

Not sure/unsicher



Moderate/moderat



Sure/sicher

Very Sure/sehr
sicher**How reliable do you consider the recommendation of the algorithm to be?****CA13**Not at all
reliable/sehr
unzuverlässigNot Reliable/
unzuverlässig

Moderate/moderat



Reliable/zuverlässig

Very Reliable/sehr
zuverlässig**Seite 19**

CN5

Please look at the image and read the information given below carefully. And answer the following questions.

00STYLE_Resi Case title



- Prediction from the Algorithmic Decision Making by AI for Melanoma: 4%.
- No change in size, color, shape, or structure noted.

CN5**Make a prediction****PR14**

What, according to you, is the probability that the spot on skin in the image shown above is Melanoma?(in %)

0%

100%

Make a decision**DC14**

A biopsy should be performed on the patient



No biopsy should be performed on the patient.

CS14

How sure are you of your decision?

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Not at all sure/sehr unsicher	Not sure/unsicher	Moderate/moderat	Sure/sicher	Very Sure/sehr sicher

How reliable do you consider the recommendation of the algorithm to be?**CA14**

<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Not at all reliable/sehr unzuverlässig	Not Reliable/ unzuverlässig	Moderate/moderat	Reliable/zuverlässig	Very Reliable/sehr zuverlässig

Seite 20

CP5

Please look at the image and read the information given below carefully. And answer the **Case title**.



- Prediction from the Algorithmic Decision Making by AI for Melanoma: 97%. CP5
- The spot has been itchy and shows no signs of healing. The skin has become darker around the spot, with blood clots forming around the spot.

Make a predictionPR15

What, according to you, is the probability that the spot on skin in the image shown above is Melanoma?(in %)

0%

100%

Make a decisionDC15

A biopsy should be performed on the patient



No biopsy should be performed on the patient.

How sure are you of your decision?CS15Not at all
sure/sehr unsicher

Not sure/unsicher



Moderate/moderat



Sure/sicher

Very Sure/sehr
sicher**How reliable do you consider the recommendation of the algorithm to be?**CA15Not at all
reliable/ sehr
unzuverlässigNot Reliable/
unzuverlässig

Moderate/moderat



Reliable/zuverlässig

Very Reliable/ sehr
zuverlässig**Seite 21****POSTSURVEY**

You have now processed all 15 cases. Please answer the questions below.

Perception IntroBE03

Please indicate to what extent you agree with the following statements.

	completely disagree	strongly disagree	somewhat disagree	undecided	somewhat agree	strongly agree	completely agree
I think I have a good grasp of the algorithm.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I think I know how to use the algorithm.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I think I know how the algorithm works.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	completely disagree	strongly disagree	disagree	undecided	agree	strongly agree	PO04 <input type="checkbox"/> <small>completely disagree</small>
I have largely ignored the algorithm in my decisions	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The algorithm was very helpful in the decision making process.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have incorporated the recommendations of the algorithm into my decision-making process.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I found the recommendations of the algorithm reasonable.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The recommendations of the algorithm were in line with my assessment.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The recommendations of the algorithm were easy to understand.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
In my opinion the algorithm did not give good recommendations.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The algorithm made errors.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The algorithm was unreliable.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Please indicate how much you agree with the following :**PO02** **“The decisions of algorithm influenced my decisions”**

<input type="radio"/>				
Strongly Disagree	Somewhat Disagree	Moderate	Somewhat Agree	Strongly agree

Please indicate how fair you consider the algorithm to be:**PO03**

<input type="radio"/>				
Very Unfair	Somewhat Unfair	Moderate	Somewhat Fair	Strongly Fair

I think it is good when decision-makers in the medical system receive assistance from algorithm-based recommendation systems.

Please indicate how much you agree with the following statement.

Strongly Disagree

Somewhat Disagree

Moderate

Somewhat Agree

Strongly Agree

Seite 22

CON

CV03

Algorithm-based recommendation systems are also used in many other areas. We have listed four of these use cases below.

Please sort them according to their severity. The most serious use case should be placed on rank 1.

There are two ways to sort the terms. Either (a) you drag the cards with the mouse to a free rank or (b) you click on them one after another with a double click.

*Triage = The prioritization of medical aid in case of resource shortage. Example: Who will receive life-sustaining treatments during the Covid-19 pandemic and who will not.

Diagnosis of skin cancer

Selection of applicants in a company

Decisions in the criminal justice system

Recommendations in a dating app

Triage*

1

2

3

4

5

CV04

Please sort the use cases now according to which case you would most likely agree to the use of a recommendation system.

Diagnosis of skin cancer	Selection of applicants in a company	1
Decisions in the criminal justice system	Recommendations in a dating app	2
Triage*		3
		4
		5

Seite 23

DEM

In the following, we would like to collect some demographic information

DE01

3. Please indicate your gender

DE02

4. Please indicate the age group you belong to.

DE06

[Please choose]

5. In what setting did you learn the basics of the ABCDE method?

DE14

For example, course module name, event or training.

6. Please name the institution where you learned the basics of the ABCDE method.

DE15

For example, name of the university, clinic or practice.

DE10

7. Please indicate your field of study.

8. If you are currently studying, please indicate your study semester.**DE11**

9. How would you rate your knowledge in the following areas?**DE09**

Knowledge in the area of	non-existent	minimal	moderate	advanced	very advanced
Computer Science (general)	<input type="radio"/>				
Machine Learning	<input type="radio"/>				
Dermatology	<input type="radio"/>				

10. Please state your general opinion on the use of algorithm-based recommendation systems.**DE12**

I am more **in favor** of the use of algorithm-based recommendation systems.

I am rather **against** the use of algorithm-based recommendation systems.

Please give a short justification for your decision.

DE13

Why:

Thank you very much!

Debriefing lowBad

Thank you for taking the time to participate in this survey.

At the end of the survey, we would like to inform you about the background of our scientific question.

This study investigates to what extent people include the decision template of an algorithm in their own decision making. In this study we investigate the influence of:

- 1) little as opposed to detailed information about the algorithm on the probability that people will include the algorithm's recommendation in their decision making.
- 2) the reliability of the algorithm (algorithm makes no mistakes or makes mistakes) on the probability that people will include the algorithm's recommendation in their decision.

They were shown little information and the algorithm made mistakes.

To ensure that all study participants receive the same cases and recommendations, no real algorithm was used in this study. The cooperation with the Department of Computer Science of the RWTH Aachen University and the Department of Dermatology of the RWTH Aachen University does not exist and served to make the cases appear credible.

Even though no algorithm was developed in the context of this survey, algorithms of this kind exist with a very good prediction probability. Our investigation helps to find out how such systems can be used responsibly by experts in the future.

Should you discover conspicuous skin marks on yourself after this study, please contact your family doctor or a dermatologist.

If you have further questions about the content of this study, please send us an email with the subject "Question about the Derma Algorithm Study" to itec@humtec.rwth-aachen.de.

On the following page, you have the opportunity to participate in the lottery mentioned at the beginning of the survey. Thank you again for taking the time to support us in our research.

- I would like to participate in the **lottery**. I agree that my e-mail address will be saved until the winner is drawn. My interview will continue to be anonymous and my email address will not be passed on to third parties.

LO01

TEST

Letzte Seite

Thank you very much for your participation!

We would like to thank you very much for your support.

Your answers have been saved, you can now close the browser window.

[B.Sc. Sören Schöder](#), [B. Eng. Sourabh Zanwar](#), [Prof. Astrid Rosenthal-von der Pütten](#), RWTH Aachen