

# 1 List of Biases

The biases are categorized and the relevance for PASSION is added to their chapter title in italic, e.g. *high*.

## 1.1 Category: Sampling Bias

Sampling biases occur when the process of collecting data results in samples that are not representative of the broader population. These biases affect the generalisability of machine learning models, especially in medical applications, where population diversity is crucial. According to Mehrabi et al. [13], non-random or selective sampling can lead to serious consequences in terms of fairness and effectiveness of AI systems.

### 1.1.1 Sampling Bias, *high*

- **Definition:** Bias introduced through non-random sampling of subgroups, leading to poor generalisation.
- **Example:** An ML model trained predominantly on patients from urban hospitals may underperform for rural patients.
- **PASSION Relevance:** PASSION aims to address dermatologic sampling bias against highly pigmented skin, but if the included data is not truly representative across populations (e.g., over-representation of certain regions), this could still result in sampling bias [13].
- **Mitigation Strategy:** Ensure a truly random and inclusive sampling strategy across geography, socioeconomic status, and skin types.

### 1.1.2 Selection Bias, *high*

- **Definition:** Bias arising when only a specific subset of the population is used, which is not representative.

- **Example:** Training a model only on adult data, when the target population includes children.
- **PASSION Relevance:** PASSION may suffer from selection bias if only data from severe dermatology cases in hospitals is used [15, 3].
- **Mitigation Strategy:** Include a broad variety of case severities and healthcare settings in the dataset.

### 1.1.3 Systematic Selection Bias, *high*

- **Definition:** A form of selection bias where chosen samples differ systematically from the general population.
- **Example:** Including only hospitalized patients in a dataset, while most cases are treated in outpatient settings.
- **PASSION Relevance:** If PASSION uses data only from dermatology centers treating severe cases, it introduces systematic selection bias [c5, c6, c33, 3].
- **Mitigation Strategy:** Include mild, moderate, and severe cases from various clinical settings.

### 1.1.4 Ascertainment Bias, *high*

- **Definition:** A systematic distortion arising from the method by which participants or data are selected for inclusion.
- **Example:** Studies on STD prevalence conducted only in public clinics may overlook patients from higher-income backgrounds who go to private practitioners.
- **PASSION Relevance:** If PASSION's dataset is composed mostly of patients from certain types of clinics, it may not generalise well to other socioeconomic groups [c5, 3].
- **Mitigation Strategy:** Ensure that data is collected from a diverse range of sources, including both public and private healthcare facilities.

### 1.1.5 Availability Bias, *high*

- **Definition:** Overreliance on easily accessible data rather than the most representative data.
- **Example:** Using only online available datasets for skin conditions may underrepresent rare diseases.
- **PASSION Relevance:** PASSION inherits availability bias by relying on FST scale-labelled datasets, which may not fully reflect global skin tone diversity [c9, c10, 3].
- **Mitigation Strategy:** Actively seek underrepresented data sources, especially for less common or less documented skin types.

### 1.1.6 Survivorship Bias, *medium*

- **Definition:** Only using data from "survivors", i.e., subjects that make it through a certain threshold or are retained in the dataset, ignoring those who were lost earlier.
- **Example:** Evaluating the success of a treatment based only on patients who completed it, ignoring those who dropped out due to side effects.
- **PASSION Relevance:** If certain dermatology diseases are lethal or if the dataset excludes patients unable to attend the centers involved in PASSION, survivorship bias may be present [15].
- **Mitigation Strategy:** Account for dropout rates and include cases from a wide range of medical access points.

## 1.2 Category: Representation Biases

Representation biases occur when a sample used to train or evaluate a machine learning model fails to adequately reflect the diversity of the target population. These biases can lead to underperformance for certain subgroups and may negatively impact the fairness and accuracy of a model in real-world applications. In the context of dermatology, these biases could result in skin diseases being underrepresented or misclassified in specific demographic groups, leading to poorer diagnostic outcomes for those populations.

### 1.2.1 Representation Bias, *high*

- **Definition:** Representation bias arises when the sample used to train a model does not adequately represent all subgroups of the target population, leading to missing or misrepresented characteristics in the data.
- **Example:** If a skin disease detection model is trained predominantly on skin types I-IV, it may struggle to accurately diagnose conditions in individuals with darker skin tones (FST skin types V-VI).
- **PASSION Relevance:** PASSION attempts to mitigate representation bias by including more FST skin types, but challenges may still exist. The dataset could still lack full representation of all diverse skin conditions and demographic factors, leading to potential misdiagnoses or underperformance for specific subgroups.
- **Mitigation Strategy:** A potential mitigation strategy could involve ensuring a more balanced representation of FST skin types, including rare and diverse skin conditions, and periodically reassessing the dataset to ensure comprehensive inclusion of all skin types across various demographics.

### 1.2.2 Population Bias, *medium*

- **Definition:** Population bias occurs when the sample's demographic characteristics (such as age, gender, or ethnicity) do not align with the target population, leading to non-representative data.
- **Example:** If a dataset is predominantly comprised of one ethnic group, a model trained on this data may not generalize well to other ethnic groups, especially if the manifestation of skin diseases varies across ethnicities.
- **PASSION Relevance:** PASSION might be impacted by population bias if it is insufficiently diverse in terms of patient demographics (e.g., ethnicity, age). The dataset needs to ensure that skin diseases are accurately represented across different population groups to avoid skewing results and compromising diagnostic accuracy.
- **Mitigation Strategy:** A mitigation strategy could involve collecting data from diverse populations and ensuring the dataset reflects the target population's demographic diversity, particularly for ethnicities and age groups that may exhibit different disease manifestations.

### 1.2.3 Aggregation Bias, *high*

- **Definition:** Aggregation bias occurs when conclusions drawn from the entire population do not apply to individual subgroups, leading to incorrect or generalized assumptions. This bias arises when significant differences between subgroups (such as gender or ethnicity) are not properly accounted for.
- **Example:** A diagnostic model trained on a heterogeneous dataset might fail to capture how skin diseases manifest differently across genders or ethnic groups, potentially leading to misdiagnosis or unequal treatment recommendations.
- **PASSION Relevance:** Aggregation bias is a significant concern in PASSION, particularly since skin diseases can manifest differently across ethnicities, genders, or genetic backgrounds. The model needs to account for these variations to avoid generalized conclusions that might harm certain subgroups.
- **Mitigation Strategy:** To mitigate aggregation bias, the model should incorporate subgroup-specific data and analysis, ensuring that disease manifestations are correctly accounted for and tailored to different demographic characteristics.

### 1.2.4 Simpson's Paradox, *medium*

- **Definition:** Simpson's Paradox is a form of aggregation bias where trends that appear in aggregated data may reverse when the data is disaggregated into subgroups. This paradox can lead to misleading conclusions if not properly addressed.
- **Example:** A dataset may show that skin disease detection is more accurate overall for a specific demographic group, but when the data is broken down by age or skin type, the trend reverses for certain subgroups.
- **PASSION Relevance:** Simpson's Paradox could be an issue in PASSION if aggregated data from different subgroups results in misleading conclusions. For example, overall accuracy may appear high, but specific skin conditions in certain ethnicities or age groups could have lower accuracy when analyzed separately.

- **Mitigation Strategy:** A mitigation strategy would involve analyzing data at both the aggregated and disaggregated levels, ensuring that subgroup-specific trends are considered to avoid false conclusions or the reversal of apparent associations.

## 1.3 Category: Measurement Biases

Measurement biases occur when the process of choosing, using, or measuring features leads to inaccurate or misleading results. These biases can emerge from various sources such as mismeasured variables, subconscious expectations of researchers, or inconsistencies in human annotation, and they can significantly affect the reliability of the dataset.

### 1.3.1 Measurement Bias, *high*

- **Definition:** Measurement bias occurs when features or variables are inaccurately measured or selected, leading to incorrect interpretations of the outcome.
- **Example:** If a proxy variable, such as country of origin, is used to infer ethnicity or genetic background, it could lead to misinterpretation of the data. For instance, the country of origin may not directly correlate with ethnic background, potentially skewing results in genetic or disease research [13].
- **PASSION Relevance:** In the context of the PASSION dataset, measurement bias could arise if country of origin is misused as a proxy for ethnicity, which is not directly related to genetic predispositions or skin conditions. This could result in misleading conclusions about skin diseases across different demographic groups, potentially amplifying health disparities.
- **Mitigation Strategy:** To mitigate measurement bias in PASSION, careful consideration should be given to the choice of features used in the dataset. Avoiding proxy variables such as country of origin to infer ethnicity and instead focusing on genetically relevant factors could improve the accuracy of the data and its interpretation.

### 1.3.2 Observer Bias, *medium*

- **Definition:** Observer bias occurs when researchers or testers influence the results by projecting their expectations or perceptions onto the data collection process, or when different observers report the same observation differently.
- **Example:** A researcher may subconsciously interpret certain skin disease symptoms differently based on their own expectations or biases, leading to inconsistent data collection or interpretation [15].
- **PASSION Relevance:** In PASSION, observer bias could affect the consistency and reliability of skin disease annotations. For example, a researcher might influence how they categorize or diagnose certain skin diseases based on their personal biases or experience. This could lead to inaccurate classifications, particularly for diseases that are subjective in appearance.
- **Mitigation Strategy:** To address observer bias in PASSION, standardized training for annotators and a clear, objective set of criteria for diagnosis should be implemented. Additionally, using multiple annotators and cross-checking results can help reduce the impact of individual biases.

### 1.3.3 Annotator Bias, *high*

- **Definition:** Annotator bias is a form of observer bias where human annotators are influenced by personal background, expectations, or external factors, which can lead to inconsistent or skewed labeling of data [16].
- **Example:** If an annotator is more likely to label a darker skin tone as "severe" or "critical" due to personal or cultural biases, this can introduce inaccuracies in the dataset, which may not be representative of the actual severity of the condition.
- **PASSION Relevance:** In PASSION, annotator bias could particularly affect the labeling of skin tones, which are highly subjective and dependent on individual perception. This bias could lead to inconsistent classifications of skin conditions across different demographic groups, which is critical when assessing dermatological diseases in a diverse population.

- **Mitigation Strategy:** To reduce annotator bias in PASSION, a diverse team of annotators should be trained to recognize and overcome their personal biases. Additionally, the annotation process should be regularly audited to ensure consistency, and the use of automated tools for initial labeling could provide more objectivity in the process.

### 1.3.4 Recall Bias, *medium*

- **Definition:** Recall bias occurs when individuals do not accurately remember or report information due to selective memory, which can lead to misinterpretations or inaccurate conclusions in data analysis [15, 3].
- **Example:** If patients are asked to recall past skin conditions or treatments, they may forget important details, leading to inaccurate reporting in the dataset. This could affect the analysis of how different skin diseases develop or respond to treatments.
- **PASSION Relevance:** Recall bias may not be directly relevant in the context of PASSION since the dataset appears to rely on clinical observations and annotations rather than patient-reported data. However, if there is any patient input, such as in follow-up surveys or self-reported symptoms, recall bias could still influence the dataset.
- **Mitigation Strategy:** To mitigate recall bias, it would be important to gather more objective data through clinical observations or imaging, and ensure that patient self-reports are validated through corroborating medical records or consistent follow-ups.

**Potential Biases in PASSION** Measurement Bias: Country of origin should not be used as a proxy for ethnicity in the PASSION dataset, as it may not be directly related to genetic or disease factors. Additionally, annotator bias regarding skin tone labeling has been investigated in recent studies and should be addressed in PASSION’s annotation process [16].

## 1.4 Category: Research Biases

Research biases refer to the ways in which researchers’ decisions, intentions, and contexts influence the outcomes of their studies, potentially introducing systematic errors that may affect the validity or generalizability of the findings.



### 1.4.1 Funding / Sponsorship bias, *medium*

- **Definition:** Funding or sponsorship bias occurs when research findings are consciously or unconsciously influenced by the expectations or interests of the study's financial backers. This can lead to findings that favor the sponsor's interests.
- **Example:** A dermatology study funded by a pharmaceutical company that produces skin disease treatment medications may emphasize the effectiveness of the company's products, even if there is no strong evidence supporting their superiority.
- **PASSION Relevance:** Funding bias could affect the PASSION dataset if the research or data collection process were influenced by sponsors or stakeholders with vested interests in certain outcomes. While this bias is not explicitly mentioned in PASSION, it is important for future studies to ensure that funding sources do not shape data interpretation or collection in a way that would lead to skewed or misleading results.
- **Mitigation Strategy:** To mitigate this, independent funding sources or transparent funding disclosure practices should be implemented. Additionally, external audits or independent validation of the findings can help prevent undue influence from sponsors.

### 1.4.2 Data dredging bias, *low*

- **Definition:** Data dredging bias arises when researchers deliberately select statistical methods or models that lead to specific p-values or results, potentially making their hypothesis appear more likely to be true than it actually is.
- **Example:** A researcher testing multiple variables in a dataset might select those combinations that yield the most statistically significant results, even if the relationships between the variables were not hypothesized initially.
- **PASSION Relevance:** Given that PASSION is a large dermatology dataset, it could be vulnerable to data dredging if analysts test many variables or relationships without pre-specified hypotheses. This could lead to spurious findings or models that do not generalize well to new data.

- **Mitigation Strategy:** To avoid data dredging, a clear and well-defined hypothesis should be established before conducting any statistical tests. Additionally, cross-validation techniques and reporting of all tested models can ensure transparency in the research process.

### 1.4.3 Hypothetical bias, *not applicable*

- **Definition:** Hypothetical bias occurs when responses to hypothetical questions do not reflect real-world behavior or preferences.
- **Example:** Asking participants how likely they would be to adopt a particular skincare treatment, without actually testing their behavior in real-world settings.
- **PASSION Relevance:** This bias is not applicable to the PASSION dataset, as the dataset does not involve hypothetical scenarios or self-reported intentions. The dataset primarily contains real-world medical data related to dermatology, which does not rely on participant speculation or hypothetical responses.
- **Mitigation Strategy:** Since this bias is not relevant to PASSION, no specific mitigation strategy is necessary.

**Potential Biases in PASSION** Since the PASSION dataset is already published, the research biases might already be introduced. It is not feasible during the duration of this thesis to make an evaluation on those biases. Instead, I would recommend the PASSION team and researchers in general to check the list above carefully and take measures against them. Maybe, an external evaluation could help to detect and prevent those biases even better.

## 1.5 Category: Feature Representation Biases

Feature representation biases occur when the features or variables used in a model do not adequately capture the complexity of the problem or reflect all relevant aspects of the data, potentially leading to biased or incomplete predictions.

### 1.5.1 Omitted Variable Bias, *high*

- **Definition:** Omitted variable bias arises when key variables are left out of a model, causing the model to be unprepared to account for cer-

tain aspects of the data and potentially leading to biased or inaccurate predictions.

- **Example:** If a dermatology model only includes skin condition data but omits important demographic information such as ethnicity or age, it may fail to identify or misinterpret certain patterns in the data.
- **PASSION Relevance:** The PASSION dataset has an omission of ethnicity as a feature, which could lead to biased results. Certain skin diseases and their manifestation can vary significantly across different ethnic groups. Without this variable, the model may fail to capture important differences in the data, leading to inaccurate predictions or generalizations.
- **Mitigation Strategy:** To address this bias, it is important to include a comprehensive set of features, such as ethnicity, age, gender, and other demographic factors, which could help the model better account for variations in skin conditions across different populations.

### 1.5.2 Collider Bias, *medium*

- **Definition:** Collider bias occurs when two variables influence a common third variable (the collider variable), and the analysis restricts sampling based on this collider, leading to a distorted or biased relationship between the variables.
- **Example:** In the case of skin disease models, if researchers only include patients who seek treatment for a specific skin condition (the collider), this may limit the analysis to a non-representative sample, potentially distorting the relationship between disease characteristics and other factors.
- **PASSION Relevance:** Although no specific collider bias has been identified in PASSION, it is important to consider that factors like patient willingness to seek treatment or the specific type of skin disease could act as collider variables. Restricting the dataset based on these factors might create a biased representation of the population.
- **Mitigation Strategy:** To reduce collider bias, it is important to ensure that the sample is as representative as possible of the broader population. Researchers should avoid restrictions that could inadvertently create a non-representative dataset and be mindful of how their sampling methods may introduce bias.

**Potential Biases in PASSION** The PASSION dataset may suffer from omitted variable bias, particularly with the lack of ethnicity data, which can affect the fairness and accuracy of dermatology models. Collider bias could also emerge depending on how the dataset is sampled or restricted based on treatment-seeking behavior or disease severity. It is important for researchers to monitor for these biases and take steps to mitigate them by ensuring that data collection and sampling strategies are inclusive and comprehensive.

## 1.6 Category: Imaging Biases

Imaging biases refer to the influence that technical variations, environmental factors, and other visual elements have on image-based classification systems. These biases can arise from issues such as the quality of the image, artifacts present in the image, or the field of view captured, which can all influence the performance of machine learning models.

### 1.6.1 Image Quality Bias, *high*

- **Definition:** Image quality bias occurs when the quality of an image—such as the zoom level, focus, or lighting—affects how a machine learning model classifies or diagnoses the image. Poor image quality can lead to misclassification or lower prediction accuracy.
- **Example:** If a dermatologist captures an image with insufficient lighting or poor focus, the model may struggle to identify skin conditions like melanomas, potentially leading to a misdiagnosis.
- **PASSION Relevance:** In the PASSION dataset, variations in image quality could lead to biased predictions. For instance, images captured under different lighting conditions or at varying zoom levels might cause the model to overfit to certain image qualities, mistaking them for certain conditions. This could reduce the model’s generalizability to diverse real-world conditions.
- **Mitigation Strategy:** To mitigate image quality bias, it is essential to standardize image acquisition protocols and pre-process images to normalize variations in quality. Implementing techniques like image enhancement and quality control during data collection could help improve model performance.

### 1.6.2 Visual Artifact Bias, *high*

- **Definition:** Visual artifact bias arises from artifacts in dermatology images, such as hair, surgical ink markings, or other extraneous elements that could interfere with accurate classification of skin diseases.
- **Example:** A photograph of a skin lesion may contain hair or tattoos from previous medical procedures, making it more difficult for the model to identify the skin condition correctly.
- **PASSION Relevance:** The PASSION dataset may include dermatology images with artifacts like surgical markings or hair, which could confuse the model into associating these artifacts with the presence of a skin disease. This could lead to incorrect predictions, especially if the model cannot differentiate between the artifact and the actual lesion.
- **Mitigation Strategy:** To reduce visual artifact bias, it is important to implement preprocessing steps that remove or mask artifacts in images. This could involve techniques such as hair removal or the use of clean, artifact-free image samples for training.

### 1.6.3 Field of View Bias, *high*

- **Definition:** Field of view bias occurs when the portion of the body or skin that is captured in an image is limited, affecting how well a model can classify a skin condition. Different angles, distances, or body parts in the view may lead to different prediction results.
- **Example:** If only a small portion of a skin lesion is captured in the image (e.g., just the edge of a mole), the model may miss critical features needed to correctly identify melanoma or other conditions.
- **PASSION Relevance:** In the PASSION dataset, field of view bias could emerge if certain lesions are captured from angles or in parts of the body that limit the information available for accurate classification. This could result in the model underperforming on images that are not representative of common views of skin conditions.
- **Mitigation Strategy:** To address field of view bias, the dataset should ensure that images are captured from standardized and consistent angles or distances. Augmenting the dataset with a variety of views from multiple angles could help improve the model's ability to generalize to unseen cases.

**Potential Biases in PASSION** The PASSION model could learn to associate unrelated visual effects, hair, body parts, or image quality with a disease, which could impact its performance. Ensuring standardized image acquisition methods and removing artifacts could mitigate some of these biases.

## 1.7 Category: Medical Biases

Medical biases are specific to healthcare-related machine learning applications and can have direct implications for diagnosis, treatment, and patient outcomes. These biases often arise from the healthcare system's structure and can lead to distorted or inaccurate predictions based on incomplete or unrepresentative data.

### 1.7.1 Berksonian Bias, *medium*

- **Definition:** Berksonian bias occurs in hospital-based studies when certain factors (such as disease severity or risk factors) influence whether patients seek treatment or are hospitalized. This can distort the relationship between variables due to the study population being unrepresentative of the general population.
- **Example:** In a study focusing on skin diseases, if only patients who sought care for severe conditions are included, the relationship between disease severity and other factors could be overstated, leading to inaccurate conclusions.
- **PASSION Relevance:** The PASSION dataset could be influenced by Berksonian bias if the images are sourced only from patients who visited certain hospitals or dermatologists, potentially skewing the representation of less severe or untreated conditions. This could limit the model's generalization to populations with different healthcare access.
- **Mitigation Strategy:** To mitigate Berksonian bias, it is important to include a diverse set of patients from multiple sources, including both hospital and non-hospital populations, ensuring a more representative dataset.

### 1.7.2 Informed Presence Bias, *medium*

- **Definition:** Informed presence bias occurs when individuals who seek medical care are more likely to be screened for other diseases. This bias can result in misleading interpretations of the relationships between diseases.
- **Example:** A person who is already being treated for one skin condition might also be screened for other conditions, leading to a misinterpretation of comorbidities or a false relationship between conditions.
- **PASSION Relevance:** In the PASSION context, informed presence bias could affect correlations between different skin diseases. If patients with certain conditions are more likely to seek treatment, the model might overestimate the likelihood of co-occurrence between those conditions.
- **Mitigation Strategy:** To reduce informed presence bias, the model should account for patients with varying levels of care-seeking behavior and ensure that both treated and untreated conditions are represented in the dataset.

### 1.7.3 Diagnostic Access Bias, *medium*

- **Definition:** Diagnostic access bias occurs when individuals in certain geographical locations have better access to medical care, leading to earlier diagnosis and potentially higher disease prevalence in those regions.
- **Example:** Patients in urban areas with better healthcare access may receive earlier diagnoses of skin conditions like melanoma, while those in rural or underserved areas may have their conditions diagnosed at a later stage.
- **PASSION Relevance:** PASSION attempts to address diagnostic access bias by including samples from later stages of diseases. However, it could still be relevant if the dataset over-represents well-diagnosed cases from areas with better healthcare access, skewing the distribution of disease stages.
- **Mitigation Strategy:** To address diagnostic access bias, it is important to ensure that the dataset includes a diverse range of geographical locations and healthcare access levels, including both early and late-stage conditions.

### 1.7.4 Diagnostic Reference Test Bias, *medium*

- **Definition:** Diagnostic reference test bias occurs when not all individuals in a study receive the same reference test, leading to discrepancies in diagnoses.
- **Example:** Inconsistent use of reference tests across different hospitals or dermatologists may result in different diagnoses for the same patient, causing confusion and inconsistency in the results.
- **PASSION Relevance:** Depending on how dermatologists work in the PASSION dataset, diagnostic reference test bias could be present. If different diagnostic methods or reference tests are used, the model may learn to associate certain diagnostic practices with specific diseases, rather than the diseases themselves.
- **Mitigation Strategy:** To mitigate diagnostic reference test bias, it is important to standardize the diagnostic processes across different healthcare settings and ensure consistent use of reference tests when collecting data.

**Potential Biases in PASSION** Some of the medical biases that could impact PASSION include Berksonian bias, informed presence bias, diagnostic access bias, and diagnostic reference test bias. Each of these could influence how the model generalizes to real-world populations. Addressing these biases requires careful consideration of the dataset’s diversity and the standardization of diagnostic practices across different settings.

## 1.8 Category: Temporal Biases

Temporal biases arise due to differences in populations and their behavior over time. These biases can manifest when studying the progression of diseases or tracking changes in populations over time. In studies where data are collected over extended periods, temporal biases can affect the accuracy and generalizability of the results.

### 1.8.1 Longitudinal Data Fallacy, *not applicable*

- **Definition:** Longitudinal data fallacy refers to the misinterpretation or improper use of data collected over time, often caused by overlooking important variables or assuming temporal relationships without proper evidence.



- **Example:** A study might incorrectly assume that a disease progression observed over a period directly results from the treatment being applied, while other confounding factors may also play a role.
- **PASSION Relevance:** Temporal biases such as longitudinal data fallacy do not apply to the PASSION dataset, as it does not track disease progression over time but instead consists of static images that are not connected to temporal data.
- **Mitigation Strategy:** Since the PASSION dataset does not involve longitudinal data, no mitigation strategy is necessary for this particular bias.

### 1.8.2 Chronological Bias, *not applicable*

- **Definition:** Chronological bias occurs when the timing of data collection influences the results or introduces errors, often due to the use of data collected at different time points that may not be representative of the population or phenomenon being studied.
- **Example:** If a medical dataset includes only images collected from patients in a certain time period where a specific treatment was more commonly used, the findings might be skewed to reflect outcomes that are not generalizable to other time periods.
- **PASSION Relevance:** Chronological bias is irrelevant to the PASSION dataset as it consists of static images of skin diseases, without any temporal association or tracking of disease progression.
- **Mitigation Strategy:** Since PASSION does not contain temporal data, no mitigation strategy is needed for chronological bias in this case.

### 1.8.3 Immortal Time Bias, *not applicable*

- **Definition:** Immortal time bias refers to a situation where the period of time during which an event could have occurred is misclassified, leading to incorrect conclusions, typically when patients are erroneously considered "at risk" for an event for a period in which the event could not have occurred.
- **Example:** A study that tracks patients who have received a specific treatment might misclassify the time between treatment and disease

progression as time "at risk," even though the patients were not at risk during the follow-up period.

- **PASSION Relevance:** Immortal time bias does not apply to PASSION, as the dataset does not track time or disease progression and focuses on static images of skin diseases, eliminating the possibility of immortal time bias.
- **Mitigation Strategy:** No mitigation strategy is necessary for immortal time bias in PASSION, as it is not a relevant concern for the dataset.

## 1.9 Category: Algorithmic Biases

When an algorithm adds biases to unbiased input data, it is referred to as **Algorithmic Bias** [1]. This can arise due to various algorithmic design choices such as optimization functions, regularizations, and statistically biased estimators [6].

### 1.9.1 User Algorithm Interaction Biases, *high*

- **Definition:** User interaction biases arise when the user interface or user behavior influences the way an algorithm behaves, potentially introducing bias. This can occur when the user interface encourages specific actions or when users impose their own biases during interaction. [1]
- **Example:** A user interacting with a teledermatology system might over-rely on certain image features, skewing the algorithm's assessment or recommendation of treatment. For instance, if a teledermatology app visually emphasizes certain markers that are less important clinically, users may begin to prioritize those markers, which could distort the results the algorithm provides. Lerman and Hogg [12] and Mehrabi et al. [13]
- **PASSION Relevance:** In the PASSION project, user interaction biases could emerge as teledermatology platforms become more publicly available. As users interact with the system, they may unintentionally influence the algorithm's output, leading to biased diagnosis or treatment recommendations, particularly if the user interface highlights or prioritizes certain image features over others.

- **Mitigation Strategy:** To mitigate this bias, a careful evaluation of the user interface design is crucial. Ensuring that no unintended prioritization of image features occurs and that the interface does not suggest biases in how users should interact with the system would help. Additionally, the algorithm should be tested with diverse user interactions to ensure its robustness.

### 1.9.2 Emergent Bias, *high*

- **Definition:** Emergent bias occurs when changes in the population interacting with an algorithm cause shifts in how the algorithm behaves over time. These changes are not anticipated during the design phase and may appear after the algorithm is deployed. Emergent bias is especially common in user interfaces as they evolve with user behavior [8].
- **Example:** If a teledermatology system starts with a limited dataset and is deployed for a specific demographic group, users from other demographics may cause the system to make inaccurate or biased decisions, as the system was not trained to account for their skin types or conditions.
- **PASSION Relevance:** Emergent bias could be a significant concern in PASSION, especially as the system expands to a larger, more diverse user base. If the platform's initial training data predominantly comes from one demographic, the system may perform less effectively for other skin types or conditions, leading to biased diagnosis or treatment recommendations.
- **Mitigation Strategy:** Continuous monitoring of how the system interacts with different demographic groups is essential. Ensuring that new data from diverse populations is incorporated into the training set periodically can help counteract emergent biases.

## 1.10 Category: External Influence Biases

External influence biases are introduced by external factors such as inappropriate benchmarks, reference tests, or popularity metrics. These factors can distort model predictions or evaluations, leading to biases in the system's decision-making process.

### 1.10.1 Evaluation Bias, *medium*

- **Definition:** Evaluation bias occurs when inappropriate or disproportionate benchmarks are used to assess the performance of a model. This can introduce external biases into the system by measuring it against benchmarks that don't fully represent the target data or user population [21, 2].
- **Example:** If PASSION's dermatological model is evaluated using a benchmark set that overrepresents certain types of skin diseases, it may lead to the underperformance of the model for conditions that are less frequently represented in the benchmark.
- **PASSION Relevance:** This bias is relevant to PASSION because the dermatological datasets used to train and evaluate the model must be diverse and representative of the broader population. An evaluation benchmark skewed toward common conditions could impair the model's ability to accurately diagnose rare or underrepresented skin diseases.
- **Mitigation Strategy:** PASSION should implement diverse and representative benchmarks to evaluate model performance, ensuring that rare or less common conditions are also included in the evaluation dataset. Regular updates to the evaluation set as the dataset grows will help mitigate evaluation bias.

### 1.10.2 Incorporation Bias, *low*

- **Definition:** Incorporation bias arises when index tests in diagnostic accuracy studies are part of the reference tests, leading to artificially elevated sensitivity for the index tests [c21, c25, c26, 3, 23].
- **Example:** If PASSION uses diagnostic tests that are part of its reference set for evaluating accuracy, this could result in an overestimation of the model's sensitivity because the model is essentially being compared to itself, skewing results.
- **PASSION Relevance:** Incorporation bias is less relevant for PASSION since the platform likely relies on independent diagnostic benchmarks and tests to validate its dermatological models, reducing the chance of this type of bias affecting its evaluations.
- **Mitigation Strategy:** Ensuring that the reference tests used for validation are distinct and independent from the model's diagnostic tests can mitigate incorporation bias.

### 1.10.3 Popularity Bias, *low*

- **Definition:** Popularity bias occurs when more popular items or data points are exposed more often in the training dataset or evaluation process. This can lead to a model that overemphasizes popular features or outcomes, disregarding less common but potentially important cases [4, 13].
- **Example:** In the context of PASSION, if the training data is overly focused on commonly encountered dermatological conditions or frequently observed features, the model may struggle to correctly diagnose rarer skin diseases that are underrepresented.
- **PASSION Relevance:** Popularity bias is relevant for PASSION, particularly if the training dataset includes a disproportionate number of common skin conditions, thereby reducing the effectiveness of the model for rarer conditions.
- **Mitigation Strategy:** To mitigate popularity bias, it is important for PASSION to ensure that the training dataset includes a balance of both common and rare skin conditions, offering a comprehensive representation of dermatological diseases.

## 1.11 Category: Cognitive Biases

Cognitive biases refer to systematic patterns of deviation from norm or rationality in judgment, whereby inferences about other people and situations may be drawn in an illogical fashion. These biases can impact how data is presented and interpreted [14].

### 1.11.1 Confirmation Bias, *high*

- **Definition:** Confirmation bias occurs when individuals favor information that confirms their preconceptions, leading them to ignore or dismiss evidence that contradicts their beliefs [14].
- **Example:** In healthcare, patients may interpret their symptoms based on information they find on the internet, confirming their own beliefs about a condition, even if this information is not medically accurate [c15, c14, 3].

- **PASSION Relevance:** For PASSION, confirmation bias could affect the initial diagnoses of dermatological conditions, resulting in biased labeling of skin diseases. If a medical professional has preconceived notions about a condition, they may incorrectly diagnose or label skin diseases, influencing the quality and accuracy of data.
- **Mitigation Strategy:** To reduce confirmation bias, diagnostic labels in PASSION could be cross-checked by multiple independent experts, ensuring diverse viewpoints and reducing the impact of pre-existing biases on data labeling.

### 1.11.2 Belief Bias, *high*

- **Definition:** Belief bias occurs when an individual's judgment is unduly influenced by their pre-existing beliefs or intuitions, leading them to accept conclusions that fit those beliefs without critically evaluating the evidence [14].
- **Example:** A researcher may ignore contradictory data in favor of results that support their hypothesis, even when the data doesn't robustly support their claim [14].
- **PASSION Relevance:** In the context of PASSION, belief bias could lead to inaccurate diagnosis and labeling if experts rely too heavily on their subjective interpretation of the data rather than objectively evaluating it. This could skew the dataset, impacting model training and accuracy.
- **Mitigation Strategy:** Implementing blind labeling processes, where experts are unaware of previous diagnoses, could help reduce belief bias. Additionally, training experts to focus on evidence-based diagnostic criteria would help mitigate the impact of this bias.

### 1.11.3 Previous Opinion Bias, *medium*

- **Definition:** Previous opinion bias occurs when the knowledge of prior results or diagnoses influences the interpretation of new data, leading to biased conclusions [3].
- **Example:** A dermatology expert who knows the result of a previous diagnosis might let this knowledge influence their interpretation of subsequent test results, leading to potential bias in the diagnosis process [3].

- **PASSION Relevance:** In PASSION, this bias could affect the consistency and accuracy of dermatological diagnoses. If experts are aware of previous diagnoses, they might be influenced by them, which could compromise the reliability of data in the system.
- **Mitigation Strategy:** To reduce this bias, PASSION could ensure that labeling experts independently diagnose cases without access to previous diagnoses, promoting impartiality in each evaluation.

#### 1.11.4 Cause-Effect Bias, *low*

- **Definition:** Cause-effect bias arises when correlations between two variables are incorrectly interpreted as indicating a causal relationship, even when no such relationship exists [14].
- **Example:** An increase in the occurrence of skin rashes may be correlated with a particular season, but mistakenly concluding that the season is the cause of the rashes, rather than other factors, would be an example of cause-effect bias [14].
- **PASSION Relevance:** Cause-effect bias is less of an issue in PASSION, since the dataset primarily deals with diagnoses and symptoms without analyzing the underlying causes of diseases. However, if the algorithm were to be trained to predict causes, there could be a risk of misinterpreting correlations as causal relationships.
- **Mitigation Strategy:** To prevent cause-effect bias, any future development in PASSION's algorithm should focus on clear differentiations between correlation and causation, ensuring that predictions are based on robust, validated data.

#### 1.11.5 Historical Bias, *high*

- **Definition:** Historical bias refers to biases that exist in the world or society, which can influence data collection and generation processes. These biases are often a reflection of past societal inequities [21].
- **Example:** A dataset that primarily includes images of skin conditions from a specific demographic (e.g., primarily white individuals) may not accurately represent skin diseases in other populations [13].
- **PASSION Relevance:** Historical biases in the dermatology field, such as underrepresentation of certain skin types in clinical studies,

could affect the quality of the PASSION dataset. This could lead to algorithms that perform poorly for underrepresented groups.

- **Mitigation Strategy:** Ensuring diversity in the dataset by collecting data from a wide range of demographic groups (age, gender, race, etc.) is essential to reduce historical bias in PASSION’s dataset. Efforts should be made to balance the dataset and account for historically marginalized groups.

### 1.11.6 Content Production Bias, *medium*

- **Definition:** Content production bias occurs when biases are introduced during the creation of user-generated content, influenced by the creators’ backgrounds, contexts, or perspectives [18].
- **Example:** In a study, images of skin diseases may be taken by health-care professionals in settings that differ from those where the disease is most prevalent, leading to a potential misrepresentation of the condition’s typical appearance [18].
- **PASSION Relevance:** In the context of PASSION, content production bias could arise in how images of skin diseases are taken. Variations in lighting, angle, or the quality of images could lead to inconsistencies, which may affect the training and performance of machine learning models.
- **Mitigation Strategy:** To reduce content production bias, standardization of image collection protocols could be implemented, ensuring consistent lighting, angles, and image quality. Additionally, training experts to adhere to these standards would help minimize bias in the data collection process.

## 1.12 Category: Behavioral Biases

Behavioral biases occur due to the actions and judgments of individuals, which are influenced by cultural, contextual, and platform-related factors. These biases can affect data collection, interpretation, and conclusions [18].

### 1.12.1 Behavioral Bias, *medium*

- **Definition:** Behavioral bias refers to how individuals’ behavior can be influenced by the platforms they interact with, their cultural back-



ground, or their personal context [18].

- **Example:** Patients from different countries may present different behaviors when seeking medical advice for skin conditions, influenced by their cultural background and understanding of healthcare [18].
- **PASSION Relevance:** For PASSION, behavioral biases could influence who seeks dermatological care and why. Differences in healthcare-seeking behavior across cultures or countries may lead to an unrepresentative sample in the dataset. Therefore, including data from various countries could help account for these differences and improve the generalizability of the model.
- **Mitigation Strategy:** To mitigate behavioral bias, PASSION should aim to include a diverse set of data points from various geographical and cultural backgrounds. This would help ensure that the model is representative of different healthcare-seeking behaviors.

### 1.12.2 Self-Selection Bias, *high*

- **Definition:** Self-selection bias occurs when participants in a study are allowed to choose whether to participate, leading to an unrepresentative sample where certain groups are over- or underrepresented [15, 13].
- **Example:** In PASSION, only patients who seek dermatological care at hospitals would be included, which could exclude individuals with skin conditions who do not seek medical help, leading to skewed data [15, 13].
- **PASSION Relevance:** Self-selection bias is a significant issue for PASSION since the dataset relies on patients who visit hospitals, meaning those who do not seek treatment or who do not have access to healthcare will be underrepresented in the dataset.
- **Mitigation Strategy:** To mitigate self-selection bias, PASSION could look into alternative data sources, such as surveys or community outreach programs, to gather information from individuals who may not seek formal dermatological care.

## 1.13 Category: Publication Biases

Publication biases are introduced when research outcomes are selectively reported or published based on certain characteristics such as positive results

or trending topics. These biases can distort the scientific record and lead to misinterpretation or overemphasis on particular findings.

### 1.13.1 Publication Bias, *high*

- **Definition:** Publication bias occurs when studies with significant or positive results are more likely to be published than studies with non-significant or negative results. This leads to a skewed representation of the effectiveness of an intervention or treatment.
- **Example:** If studies showing positive results of a dermatological treatment for skin diseases are more likely to be published than studies with neutral or negative findings, this creates a publication bias in the medical literature.
- **PASSION Relevance:** In the context of the PASSION dermatology dataset, publication bias may manifest if studies based on the dataset predominantly focus on successful diagnoses or treatments, leaving out less effective or inconclusive results. This can lead to an overestimation of the dataset's utility and effectiveness in detecting skin diseases.
- **Mitigation Strategy:** A strategy to mitigate publication bias is the promotion of open access to all research outcomes, including negative or neutral results. Encouraging the publication of replication studies and meta-analyses that incorporate a wide range of findings, not just the most positive ones, can help counteract this bias.

### 1.13.2 Hot Stuff Bias, *medium*

- **Definition:** Hot stuff bias refers to the tendency for journals to be less critical of research related to trending or highly popular topics, leading to the disproportionate publication of these studies.
- **Example:** In dermatology, if there is a sudden interest in a new skin disease detection technology, studies related to this technology may be published more frequently, regardless of their quality, because they align with the current hot topic.
- **PASSION Relevance:** In the context of PASSION, if a certain skin disease detection method is trending, there could be a tendency for studies utilizing the PASSION dataset to be published more often, potentially overshadowing other important findings or datasets that may also contribute valuable insights.

- **Mitigation Strategy:** To mitigate hot stuff bias, it is important to prioritize the quality and robustness of the research rather than focusing on its alignment with current trends. Peer reviewers should be vigilant and ensure that the novelty of a topic does not overshadow the scientific rigor of the study.

### 1.13.3 All is Well Bias, *low*

- **Definition:** All is well bias occurs when theories that align with the majority or dominant views are more likely to be published than those that challenge the consensus.
- **Example:** In the field of dermatology, if the majority of researchers agree on a specific method for diagnosing skin diseases, studies questioning the effectiveness of this method may be less likely to be published, even if they provide valid and critical insights.
- **PASSION Relevance:** This bias is less directly relevant to PASSION as the dataset itself is focused on real-world data collection, which may be less influenced by theoretical debates. However, if there is widespread consensus on a particular model or diagnostic approach using PASSION data, studies presenting opposing findings could be underrepresented.
- **Mitigation Strategy:** Encouraging diversity in research perspectives and methodologies can help counteract the all is well bias. Peer reviewers should actively look for and support research that challenges the majority view, ensuring that minority perspectives are also given a platform.

### 1.13.4 Rhetoric Bias, *medium*

- **Definition:** Rhetoric bias occurs when the way in which research findings are presented, such as through charismatic writing or media coverage, influences the perception of those findings more than the actual data.
- **Example:** If a particular dermatology treatment is presented with highly persuasive language and strong media support, it may be perceived as more effective than it truly is, regardless of the underlying data.

- **PASSION Relevance:** In the context of PASSION, if studies using the dataset are written in a particularly persuasive or charismatic manner, they may attract more attention, regardless of their scientific rigor. This could lead to an overemphasis on certain findings, overshadowing others that may be equally valuable but presented less dramatically.
- **Mitigation Strategy:** Researchers should focus on presenting data clearly and objectively, avoiding exaggerated claims or overly persuasive language. Journals and reviewers should also ensure that rhetoric does not overshadow the actual scientific contribution of a paper.

### 1.13.5 Novelty Bias, *high*

- **Definition:** Novelty bias refers to the tendency to favor new interventions, treatments, or findings over established ones, often because newer approaches are perceived as being better, even if the evidence does not support this.
- **Example:** A new method for detecting skin diseases, such as a machine learning model, may be hailed as a breakthrough, even if it has not been proven to be more effective than traditional diagnostic methods.
- **PASSION Relevance:** Novelty bias can be particularly relevant for PASSION as researchers may place disproportionate emphasis on the latest machine learning techniques or models, potentially overlooking more established methods that are still effective in skin disease detection. This bias can lead to a focus on novelty at the cost of reliability.
- **Mitigation Strategy:** To mitigate novelty bias, researchers should compare new approaches to established methods in rigorous, controlled studies. Peer reviewers should ensure that novelty does not overshadow the importance of replicability and robustness in research findings.

**Potential Biases in PASSION** These biases are relevant for all researchers working with datasets like PASSION. They should be kept in mind when interpreting, publishing, and peer-reviewing papers. Ensuring that these biases are recognized and addressed will help maintain the integrity and utility of the PASSION dermatology dataset in advancing skin disease detection and treatment.

## 1.14 Category: Medical Biases

Medical biases refer to distortions in healthcare and diagnostic practices that can arise due to various external factors, leading to an inaccurate representation of diseases or patient conditions. These biases can affect both clinical assessments and datasets, potentially skewing the analysis and treatment approaches.

### 1.14.1 Popularity Bias, *high*

- **Definition:** Popularity bias occurs when more well-known or stigmatized diseases are over-represented in healthcare settings compared to less common diseases. This can result in a distorted view of the prevalence and severity of different conditions.
- **Example:** A hospital may see more cases of diseases that are widely known or stigmatized, such as skin cancers, leading to an overemphasis on these conditions in research and treatment, while conditions like rare dermatological disorders are underrepresented.
- **PASSION Relevance:** Popularity bias is highly relevant to the PASSION dataset, as it may include more common or well-known skin diseases due to hospital reporting trends. This could lead to an overrepresentation of these conditions in the dataset, skewing the model's ability to detect less prevalent dermatological diseases accurately.
- **Mitigation Strategy:** To mitigate popularity bias, PASSION should ensure a balanced representation of diseases in the dataset. This could involve actively seeking data from hospitals or clinics that treat a wider variety of dermatological conditions, including rare ones.

### 1.14.2 Apprehension Bias, *low*

- **Definition:** Apprehension bias arises when patients exhibit anxiety or fear about upcoming medical procedures, which can influence physiological measurements or diagnostic results, leading to inaccuracies.
- **Example:** A patient may have elevated blood pressure readings due to anxiety before a dermatological procedure, leading to an inaccurate diagnosis or assessment.

- **PASSION Relevance:** While apprehension bias may be relevant in clinical settings, its direct impact on the PASSION dataset is likely low. The dataset focuses on dermatological conditions, where apprehension bias is less likely to affect diagnostic images or disease annotations.
- **Mitigation Strategy:** Although not a primary concern for PASSION, ensuring that patients are comfortable during the data collection process and minimizing procedural anxiety could improve the accuracy of any associated clinical measurements.

### 1.14.3 Hawthorne Bias, *medium*

- **Definition:** Hawthorne bias refers to changes in behavior by subjects when they know they are being observed, which can influence study results and clinical assessments.
- **Example:** If clinicians or patients are aware that their cases are being monitored for a dermatology study, they might alter their behavior, such as reporting symptoms differently or providing more detailed information than they normally would.
- **PASSION Relevance:** The Hawthorne bias could be relevant in the PASSION dataset, particularly if annotators or clinicians are aware that their diagnostic decisions are being evaluated. This could lead to over-reporting or under-reporting certain disease characteristics.
- **Mitigation Strategy:** PASSION could utilize regular follow-ups to observe natural trends in disease progression and symptom reporting. By minimizing the knowledge of when they are being observed, PASSION can reduce the impact of this bias on the data quality.

### 1.14.4 Centripetal Bias, *medium*

- **Definition:** Centripetal bias occurs when patients tend to seek care from well-known or highly reputable specialists or institutions, which may skew the cases seen by those professionals towards more complex or specialized conditions.
- **Example:** In dermatology, patients with severe or uncommon conditions may prefer to see a renowned specialist, while more routine cases are handled by general practitioners or less well-known dermatologists.

- **PASSION Relevance:** Centripetal bias is relevant to PASSION as it may affect which hospitals or specialists contribute data to the dataset. If more well-known institutions are over-represented, this could lead to a dataset that is biased towards more severe or complicated cases of skin diseases.
- **Mitigation Strategy:** PASSION can mitigate centripetal bias by ensuring diversity in the data sources. It should include data from both specialized and general dermatology practices, as well as a mix of both urban and rural hospitals, to provide a comprehensive view of skin diseases across different settings.

**Potential Biases in PASSION** PASSION must be careful in interpreting the metadata. Since the data is from hospitals, there could be an over-representation of more popular or severe diseases, leading to a distorted dataset. Additionally, PASSION could leverage the Hawthorne bias to improve the consistency and quality of the annotations, using follow-ups to observe how annotators' behavior may shift over time. Furthermore, centripetal bias can be considered when selecting partners to work with, ensuring that a variety of specialists and institutions contribute to the dataset.

## 1.15 initial sources

### 1.15.1 Bias Sources

The general Machine Learning (ML) lifecycle consists of data gathering, training the algorithm and the user interaction with the trained model. Now, while data gathering, biases can arise either through the collection process or it is already inherited in the available data. Further, depending on the algorithm design, during training, the existing bias in the data can be amplified and new bias can be introduced. Lastly, the result of the algorithm can affect the user experience on inference which can lead to further bias amplification. This generates a feedback loop between the biases in each step of the ML lifecycle which can make it hard to identify the original bias source. The feedback loop is illustrated in Figure 1.1, which also shows first bias definitions, which were categorized according to this feedback loop [13].

check  
that all  
stuff  
above  
matches  
the stuff  
below

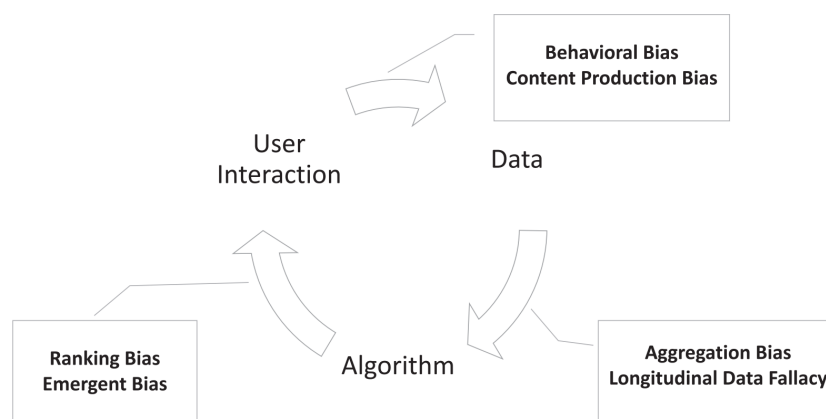


Figure 1.1: Bias definitions in a ML lifecycle [13].

- two potential sources of unfairness in machine learning outcomes - those that arise from biases in the data and those that arise from the algorithms ... we observe that biased algorithmic outcomes might impact user experience, thus generating a feedback loop between data, algorithms and users that can perpetuate and even amplify existing sources of bias [13].
- The loop capturing this feedback between biases in data, algorithms, and user interaction is illustrated in Figure 1. We use this loop to categorize definitions of bias in the section below [13]

### 1.15.2 Bias Types

The Table 1.1 aims to provide an overview over what kind of biases exist according to research. The more detailed categories listed in the table try to capture similar kind of biases. This thesis follows roughly the categorization of Mehrabi et al. [13]. Some biases might actually fit in multiple categories. The definition of the categories including examples of specific biases follows.

check  
the c  
itations  
in the  
following  
chapters



Bias	Mentioned in Context of	
	ML	Dermatology
<b>Data Biases</b>		
Sampling Biases	X <sup>1,2,3</sup>	X <sup>4</sup>
Representation Biases	X <sup>1</sup>	X <sup>5,6</sup>
Measurement Biases	X <sup>1,3</sup>	X <sup>4,6</sup>
Research Biases	X <sup>7</sup>	X <sup>4</sup>
Feature Representation Biases	X <sup>1,3</sup>	X <sup>4</sup>
Imaging Biases		X <sup>5</sup>
Medical Biases	X <sup>8</sup>	X <sup>4</sup>
Temporal Data Biases	X <sup>1</sup>	X <sup>4</sup>
<b>Algorithmic Biases</b>		
User-Algorithm Interaction Biases	X <sup>1</sup>	
External Influence Biases	X <sup>1</sup>	X <sup>4</sup>
<b>User Biases</b>		
Cognitive Biases	X <sup>1,7</sup>	X <sup>4</sup>
Behavioral Biases	X <sup>1,3</sup>	X <sup>4,5</sup>
Publication Biases		X <sup>4</sup>
Medical Biases	X	X <sup>4</sup>

<sup>1</sup> [13]

<sup>2</sup> [11]

<sup>3</sup> [15]

<sup>4</sup> [3]

<sup>5</sup> [23]

<sup>6</sup> [16]

<sup>7</sup> [14]

<sup>8</sup> [7]

Table 1.1: Bias categories - grouped according the ML lifecycle of Mehrabi et al. [13]

Data Biases

**Sampling Biases** When gathering data, it's usually not possible to gather the data of a whole population. Instead, the data is gathered by sampling. A sample is a subgroup of individuals from the population. To get unbiased results, this sampling process should represent the true population, with a low sampling error [11]. This is often achieved with randomized samples. With non-random sampling processes, sampling bias arises. The consequence is, that the insights of one sampled population may not generalize with insights on another sampled popluation [13].

Those biases can be introduced with a flawed sampling process:

- **Sampling bias**, due to nonrandom sampling of subgroups, leading to poor generalization [13]

Feedback  
Astrid:  
zu viele  
Un-  
terkapi-  
tel ->  
anders  
struk-  
turieren

- **Selection bias**, working only on specific subset of the population which is not representative [15, 3]
- **Systematic selection bias**, chosen samples differ dramatically from the representative populations; e.g. in dermatology, when only the most severe patient data gets included [c5, c6, c33, 3]
- **Ascertainment bias**, tendency to exclude segments from the population due to e.g. cultural differences, such as which patient segment goes to government clinics vs. private clinics (usually influenced by socioeconomic status) [c5, 3]
- **Availability bias**, focus on widely available data instead of most representative data [c9, c10, 3, <empty citation>]
- **Survivorship bias**, focus only on pre-selected data, ignoring the initial data-points which got filtered out [15].

**Potential Biases in PASSION** PASSION tries to reduce sampling bias in dermatology against high pigmented skin. PASSION might introduce (systematic) selection bias or Ascertainment bias, if in the dermatology centers only sickest / more severe patients are seen as indicated by Chakraborty [3] PASSION inherits availability bias as it is using Fitzpatrick skin type (FST) scale. Survivorship bias could be relevant for PASSION, if dermatology diseases could be lethal. Further, all patients which are not able to go to one of the dermatology centers which were used in PASSION could be considered to left out by survivorship bias.

used

- **Sampling Bias.** Sampling bias is similar to representation bias, and it arises due to nonrandom sampling of subgroups. As a consequence of sampling bias, the trends estimated for one population may not generalize to data collected from a new population. [13]. This is what the PASSION dataset tries to improve
- **Selection bias - wrong sampling method**, working on a specific subset of audience; usually by working only with data that is easy to access [15, 14] - statistical bias
- **Selection bias:** Since it is not possible to work with large populations, for most dermatological studies, samples are chosen that are said to be representative of the original population. In selection bias, the selected subgroups are not representative of their original population. A

variation of this is systematic selection bias, where samples chosen differ dramatically from their representative populations. Our experience suggests, such selection bias occurs more commonly in studies conducted in regional referral centers where only the sickest or more severe patients are usually seen. For example, a study compared the efficacy of thalidomide vs. prednisolone in hospitalised patients of erythema nodosum leprosum. It derived that thalidomide was more efficacious than steroids in erythema nodosum leprosum. Such findings cannot be generalised to all erythema nodosum leprosum since patients admitted to a regional referral center will likely have more severe disease.<sup>5,6,33</sup> [3]

- Availability bias: More emphasis is placed on widely available data than scantily available data. A classic example is the use of antihistamines in pregnancy dermatoses, where nearly all standard books recommend first-generation antihistamine chlorpheniramine because more data is available.<sup>9 10.</sup> [3] - dermatology
- Survivorship bias [15, 14] - statistical bias
- Ascertainment Bias: This bias is commonly encountered in venereology practice. It is defined as a bias due to the tendency of some segments of the target population to get excluded due to cultural and other differences. For example, in most venereology clinics in government setups, studies show that venereal diseases are commoner in lower socioeconomic status. One reason might be that the higher socioeconomic status people tend to go to private practitioners and thereby get excluded from government-run clinics.<sup>9,10</sup> Allocation concealment and blinding are good ways to avoid this. <sup>5.</sup> [3] - healthcare

even more extensive

- Selection bias is again divided into two types endogenous selection bias and exogenous selection bias. The best example of endogenous selection bias in dermatology is the inclusion of non-response. If a trial tests the efficacy of a particular biologic in psoriasis, the response is usually collected from trial participants via postal services. Certain participants will not respond, although they might have substantially improved. Their exclusion will result in significant differences in efficacy evaluation.<sup>33</sup> Exogenous selection bias results when both treatment and outcome result from dependency on an external variable that is not controlled. For example, if sunlight exposure is not controlled, it

will influence both the intervention and control groups since psoriasis is a photosensitive (and photoexacerbated) dermatosis. [3] - dermatology

- survivorship bias - World War II planes [Silfwer\_\_2017] - <https://doctorspin.org/media-psychology/psychology/survivorship-bias/>

## Representation Biases

Those biases can be introduced :

still describe this category

- **Representation bias**, non-representative sample lead to missing subgroups or other representation anomalies, which can be harmful to downstream applications. Popular ML datasets suffer from representation bias [13, 20]
- **Population Bias**. Population bias arise when statistics, demographics and characteristics in the sample differ from the target population [18]. The data it creates is non-representative for the target population [13].
- **Aggregation bias** occurs, when "false conclusions are drawn about individuals from observing the entire population". It doesn't matter, whether the subgroups are represented equally in the training set, any generalized assumptions can result in aggregation bias [13]. In medicine, diseases can present themselves differently across genders and ethnicities [21]. Therefore, diagnostic models need to incorporate those differences to mitigate aggregation bias [13].
- **Simpson's Paradox** is a type of aggregation bias, which arises in heterogeneous data analysis. Observed associations disappear or reverses in the subgroup data [13].

**Potential Biases in PASSION** PASSION tries to mitigate representation bias, by including more FST skin types - however, it could introduce other representation biases Aggregation bias and Simpson's Paradox could potentially be an issue when the analyzed skin diseases present themselves differently in patients based on their genetics

used

- **Representation Bias**. Representation bias arises from how we sample from a population during data collection process [21]. Non-representative samples lack the diversity of the population, with missing subgroups and other anomalies [13].

- Popular machine-learning datasets that serve as a base for most of the developed algorithms and tools can also be biased—which can be harmful to the downstream applications that are based on these datasets. ... In [20], researchers showed that these datasets suffer from representation bias and advocate for the need to incorporate geographic diversity and inclusion while creating such datasets. [13]
- Population Bias. Population bias arises when statistics, demographics, representatives, and user characteristics are different in the user population of the platform from the original target population [18]. Population bias creates non-representative data. ... More such examples and statistics related to social media use among young adults according to gender, race, ethnicity, and parental educational background can be found in [10]. [13]
- Aggregation Bias. Aggregation bias (or ecological fallacy) arises when false conclusions are drawn about individuals from observing the entire population. An example of this type of bias can be seen in clinical aid tools. Consider diabetes patients who have apparent morbidity differences across ethnicities and genders. Specifically, HbA1c levels, that are widely used to diagnose and monitor diabetes, differ in complex ways across genders and ethnicities. Therefore, a model that ignores individual differences will likely not be well-suited for all ethnic and gender groups in the population [21]. This is true even when they are represented equally in the training data. Any general assumptions about subgroups within the population can result in aggregation bias. [13]. -> could also be important for dermatology issues!!!
  - Simpson’s Paradox. Simpson’s paradox is a type of aggregation bias that arises in the analysis of heterogeneous data [18]. The paradox arises when an association observed in aggregated data disappears or reverses when the same data is disaggregated into its underlying subgroups (Fig. 2(a)). ... After analyzing graduate school admissions data, it seemed like there was bias toward women, a smaller fraction of whom were being admitted to graduate programs compared to their male counterparts. However, when admissions data was separated and analyzed over the departments, women applicants had equality and in some cases even a small advantage over men. The paradox happened as women tended to apply to departments with lower admission rates for both genders. Simpson’s paradox has been observed in a variety

of domains, including biology [37], psychology [81], astronomy [109], and computational social science [91]. [13].

**Measurement Biases** How features are chosen, used and measured can lead to biases [13, 21].

Examples for such biases are:

- **Measurement bias** in general, e.g. using mismeasured proxy variables lead to misinterpretations of the outcome [13]
- **Observer bias** is a subconscious bias which can occur in different forms. Either, researchers projects their own expectations on the research and influence the testers accordingly [15]. In other cases, different observers report the same observation differently [c29, c26, 3]
- **Annotator bias** is a special form of observer bias. The labeling process of human annotators can be influenced by lots of factors (e.g. personal background, social context) and even minor design choices (e.g. scale order, image context). This can introduce inconsistencies when labeling the data [16]
- **Recall bias**. This bias occurs when queried individuals do not remember things correctly, due to humans selective memory. This can cause misinterpretation, for example when analyzing causes and effects of behaviour on certain diseases in medicine [15, c3-6, c2, 3].

**Potential Biases in PASSION** Measurement Bias (proxy var) - Country of Origin in PASSION depending on the interpretation - should not be used for ethnicity, as this is not linked directly to the genes, see example <https://medium.com/bcggamma/practice-ai-responsibly-with-proxy-variable-detection-42c2156ad986>

Annotator bias regarding skin tone labeling has been investigated in [16]. PASSION should evaluate its process.

used

- **Measurement Bias**. Measurement, or reporting, bias arises from how we choose, utilize, and measure particular features [21] (e.g. mismeasured proxy variables) [13]. (= e.g. someone who lives at that postal code probably has this ethnicity ); -> could that be an issue with the country of origin feature?

- This study found that while using skin tone instead of race for fairness evaluations in computer vision seems objective, the annotation process remains biased by human annotators. Untested scales, unclear procedures, and a lack of awareness about annotator backgrounds and social context significantly influence skin tone labeling. This study exposes how even minor design choices in the annotation process, like scale order (dark to light instead of light to dark) or image context (face or no face, skin lesion presence), can sway agreement and introduce uncertainty in skin tone assessments. ... The researchers emphasize the need for greater transparency, standardized procedures, and careful consideration of annotator biases to mitigate these challenges and ensure fairer and more robust evaluations in computer vision. [16] - demographic dermatology bias
- Observer bias - projecting expectations onto the research [15, 14] - statistical bias
- Observer bias: When different observers view the same observation, they report it differently e.g., different observers may give differing descriptions about subtle features in the histopathology report of a skin biopsy.<sup>29 26</sup>. [3] - dermatology
- Recall bias - respondent doesn't remember things correctly; Recall bias is another common error of interview/survey situations. It happens when the respondent doesn't remember things correctly. It's not about bad or good memory – humans have selective memory by default. After a few years (or even a few days), certain things stay and others fade. It's normal, but it makes research much more difficult. [15, 14] - statistical bias
- Memory or recall bias: This is a type of bias where sufferers of a disease, often termed cases, have a greater tendency to recall a particular habit than non-sufferers, viz controls. This results in an uneven distribution of risk factors between the cases and controls. An example of this would be a case-control study to evaluate the association between dental amalgam use and the development of oral lichen planus. Those with lichen planus are more likely to recall a history of dental amalgam use than those who do not have the disease. This difference in recall between a diseased cohort and control has resulted in difficulties in assessing the association between diet and many dermatological diseases – like milk and chocolate consumption and acne, fatty meals and psoriasis, sugary meals and psoriasis, agricultural exposure to insecticides

keep  
an eye  
on this  
when  
recall-  
ing evi-  
dences!!

and pemphigus and so on.3–6 2. [3] - dermatology

**Research Biases** Researchers and their processes can also be biased in multiple ways:

- **Funding / Sponsorship bias**, when a study is deliberately supporting those findings, which the sponsor expects [c22, 3, 14]
- **Data dredging bias**. The statistical methods and model are chosen to provide a certain p-value, to improve the probability of the research hypothesis being true. [3]
- **Hypothetical bias**. Hypothetical questions lead to responses that do not reflect, what interviewees would do in real life. [c31, c28, 3]

consider to move at beginning / out of data biases

consider to move this to an own reporting section

isn't this a user bias instead?

**Potential Biases in PASSION** Since the PASSION dataset is already published, the research biases might already be introduced. It is not feasible during the duration of this thesis to make an evaluation on those biases. Instead, I would recommend the PASSION team and researcher in general, to check the list above carefully and take measures against them. Maybe, an external evaluation could help to detect and prevent those biases even better.

used

- Funding bias [15, 14] - statistical bias
- Industry sponsorship bias: This has now been reclassified as conflict-of-interest bias. In short, the study deliberately supports the findings expected from it by its sponsors. 22.[3] - dermatology

Reporting biases

- **Data dredging bias**: It is an entirely avoidable bias. This is subdivided into two types – Fishing type and “P-value hacking” type. It involves using multiple statistical methods to get the desired p-value and selecting the statistical model that gives the p-value the author wants. This is “lamentably common” in dermatological research.<sup>16</sup> To detect data dredging bias, always perform a “p-curve analysis” while performing a meta-analysis.<sup>17,18</sup> Much emphasis is nowadays given to the confidence interval instead of the p-value, which gives an approximate idea of the range in which one can be 95% (or 90%, depending on the confidence interval chosen) sure that the result is correct. The confidence interval remains unaffected by p-value dredging. This subject has been reviewed in depth in recent works.<sup>18,19</sup> 15.[3]



- **Hypothetical bias:** Many dermatological researches (and some life quality questionnaires like vitiQoL) use hypothetical questions – like “What would you do when some stranger asks you about your lesion?”. The responses to these questions by the study participants often do not tally with what they would do in real life. This is called hypothetical bias and is avoided by adopting the ex-ante approach.<sup>31 28</sup>. [3] - dermatology

**Feature Representation Biases** Some of those biases are:

- **Omitted Variable Bias** arises when variables are not included in the model, which leads to situations for which the model is not ready for [13, 15][5, 19][17].
- **Collider Bias** Two variables can influence a common third variable, the collider variable. When sampling is restricted by this collider variable, it could lead to a distortion [c4, c8, c9, 3].

**Potential Biases in PASSION** The ethnicity is omitted in the PASSION dataset which could lead to issues See the medical section for more specific collider bias, maybe there could be others

used

- **Omitted Variable Bias.** Omitted variable bias<sup>4</sup> occurs when one or more important variables are left out of the model [5, 19][17]. Something that the model was not ready for[13]. did not take into account [13]
- **Omitted variable bias** [15, 14] - statistical bias
- **Collider Bias:** This is an under-appreciated bias, and often confused with a confounder. This is especially seen in observational studies where it is defined as a distortion produced by the restriction of sampling by a collider variable. A collider variable is defined as one that has an independent effect on the outcome studied apart from the studied variable. In simpler terms, collider bias occurs when exposure and development influence a common third variable. That variable or collider is controlled by study design or in the analysis. An example is the observation that psoriasis patients tend to have more depression and anxiety disorders. Since severe psoriasis patients tend to get hospitalised and also get screened for mental health issues, a spurious association between them could have been obtained due to collider bias.

The two variables viz psoriasis and depression converged, i.e., collided, into a single outcome – hospitalization.8,9 4. [3] - dermatology

**Imaging Biases** Dealing with images can lead to a whole other set of challenges, which can lead to biases. The challenges are for example technical variations in hardware and software but also differences in how images are gathered or what is in it [23].

Those biases can be introduced :

- **Image Quality Bias.** The quality of an image (zoom level, focus, lightning) could be associated with the classification [23]
- **Visual Artifact Bias.** Other artifacts, such as presence of hair or surgical ink markings on dermatology images, can decrease classification performance [Winkler et al., 2019 & Bisla et al., 2019 (from Young\_2020)]
- **Field of View Bias.** What view is captured in the image can interfere with prediction quality what is it, consequence [Mishra et al., 2019 from Young\_2020]

**Potential Biases in PASSION** The PASSION model could learn to associate unrelated visual effects, hair, body parts or image quality with a disease.

used

- **Image quality.** Several barriers to Artificial Intelligence (AI) implementation in the clinic need to be overcome with regards to imaging (Figure 1). These include technical variations (e.g., camera hardware and software) and differences in image acquisition and quality (e.g., zoom level, focus, lighting, and presence of hair). For example, the presence of surgical ink markings is associated with decreased specificity (Winkler et al., 2019), field of view can significantly affect prediction quality (Mishra et al., 2019), and classification performance improves when hair and rulers are removed (Bisla et al., 2019). We have developed a method to measure how model predictions might be biased by the presence of a visual artifact (e.g., ink) and proposed methods to reduce such biases (Pfau et al., 2019). Poor quality images are often excluded from studies, but the problem of what makes an image adequate is not well studied. Ideally, models need to be able to express a level of confidence in a prediction as a function of image quality and appropriately direct a user to retake photos if needed. [23] - dermatology

**Medical Biases** In ML for health care, there are special medical versions of the mentioned biases as well as completely new biases. They require special attention, since they directly influence the diagnosis or treatment of a disease.

Those biases can be introduced:

- **Berksonian bias** occurs in hospital-based studies when two variables influence hospital or clinical attendance independently. This can lead to a distorted estimation of the relationship between those variables because the study population of hospitalized patients is not representative of the whole population [c3, c7, 3]
- **Informed presence bias**, the probability to get screened for other diseases is higher for people who seek medical care. Like Berksonian bias, this can lead to misleading interpretations of relationships between two diseases [c27, c23, 3]
- **Diagnostic access bias**, depending on the geographical location, individuals have better access to medical care. Therefore, their disease prevalence could appear to be higher and diseases could be diagnosed earlier. [c19-c21, 3]
- **Diagnostic reference test bias** is a **verification bias**, where not all individuals receive the same reference test for the diagnostic process, potentially leading to different diagnoses. [c21, 3]
- **Mimicry bias**, exposures to treatment options can cause a disease which presents itself similar to the study disease, which potentially creates misleading data [c28, c25, 3]
- **Unacceptable Disease bias**. When a disease is socially unacceptable, it can result in under-reporting of the same disease [c30, c27, 3]
- **Healthy volunteer selection bias**, is a type of self-selection bias where the volunteers are in general healthier than the population due to more interest in health [7]

**Potential Biases in PASSION** Berksonian bias depending on the chosen hospitals Informed presence bias regarding correlation between impetigo and the other diseases Diagnostic access bias can somewhat be addressed by PASSION, since its dataset includes samples of later states of diseases. However, in the PASSION context itself, this bias could still be

relevant. Diagnostic reference test bias could be inherited in the PASSION dataset, depending on how the dermatologists work. Mimicry bias is not relevant regarding the exposures since PASSION does not hold any exposure data. However, diseases which mimicry others could lead to issues if they are not detected.

used

- **Berksonian Bias:** Named after Dr. Joseph Berkson, this bias reflects the variation in rates of hospital admission or clinic attendance for different diseases. For example, if a study is conducted to know the effect of pregnancy on syphilis in an antenatal clinic, we are likely to get biased data since the two conditions, viz pregnancy and syphilis, are both likely to affect clinic attendance and all observations related to the relationship between pregnancy and syphilis.<sup>7 3</sup>. [3] - dermatology
- **Informed presence bias:** Simply, a person attending a health center is more likely to get screened for other unrelated comorbidities than those not attending a health center e.g., the finding psoriasis is associated with depression has now been criticised because those having psoriasis also have a greater chance to be screened for depression since they are already attending a health center.<sup>27 23</sup>. [3] - dermatology
- **Diagnostic Access Bias:** Individuals in certain geographical localities have better access to medical care and, hence, may appear to have higher disease prevalence. For example, atopic dermatitis is believed to be commoner in the West – this could be due to better and earlier diagnostic facilities available than in India.<sup>19,20 17</sup>. [3]
- **Diagnostic reference test bias:** These bias results when all individuals do not receive the same reference test. e.g., direct immunofluorescence studies may not be done for all patients with pemphigus vulgaris some patients may receive only a skin biopsy-based diagnosis. It is a subtype of verification bias. Another variation of this type of bias is partial reference bias, where only some of the study participants receive the index and the reference tests.<sup>21</sup>[3]
- **Mimicry bias:** When an exposure causes a disease that resembles the study disease, mimicry bias can result. For example, certain drugs are known to cause a pityriasis rosea-like reaction, which, although looks like pityriasis rosea, differs from it.<sup>28 25</sup>. [3] - dermatology
- **Unacceptable disease bias:** This occurs in socially unacceptable diseases like leprosy and STDs, which result in under-reporting.<sup>30 27</sup>. [3]

- dermatology

- ther such studies were conducted in [[9]] which states that UK Biobank, a large and widely used genetic dataset, may not represent the sampling population. Researchers found evidence of a “healthy volunteer” selection bias. [150] has other examples of studies on existing biases in the data used in the medical domain. [157] also looks at machine-learning algorithms and data utilized in medical fields, and writes about how artificial intelligence in health care has not impacted all patients equally.[13] -> [150] also provides an overview over the impact of social determinants on health, such as Economic stability, neighborhood and physical environment, education, food, community and scial context, access to healthcare and quality
- The healthy volunteer effect is a particular case: when the participants are healthier than the general population. [7]

**Temporal Biases** Differences in populations and their behaviour over time can lead to temporal biases [18]. Certain studies require to track temporal data, to learn about their behaviour over time. Disease progression is also a factor measured over time [13]. For PASSION, temporal biases are currently irrelevant, since PASSION contains images independently of time and is not tracking the disease progression. Therefore, the listed biases in this chapter are not explained in detail, refer to the sources for further information.

Examples for temporal data biases are:

- **Longitudinal Data Fallacy** [13]
- **Chronological bias** [c9, c13, 3]
- **Immortal time bias** [c24, c20, 3]

added  
until  
here

### 1.15.3 Algorithmic Biases

When an algorithm adds biases to unbiased input data one speaks of **Algorithmic Bias** [1]. This could occur due to algorithmic design choices like optimization functions, regularizations and statistically biased estimators [6].

used

- **Algorithmic Bias.** Algorithmic bias is when the bias is not present in the input data and is added purely by the algorithm [1]. The algorithmic design choices, such as use of certain optimization functions,

regularizations, choices in applying regression models on the data as a whole or considering subgroups, and the general use of statistically biased estimators in algorithms [6], can all contribute to biased algorithmic decisions that can bias the outcome of the algorithms.[13].

### User Algorithm Interaction Biases

- **User Interaction Bias.** This biases can be triggered by the user interface or the user themselves. The user interface influences the user to behave in a certain way, which could introduce bias in the user behaviour. Users impose this (or their own) biased behavior through interaction on the algorithm [1]. **Presentation bias** and **Ranking bias** are further subtypes mentioned by Lerman and Hogg [12] and Mehrabi et al. [13].
- **Emergent Bias.** When real users interact with an algorithm, this bias arises some time after the design was completed due to changes in population. It appears more likely in user interfaces [8].

**Potential Biases in PASSION** The user interaction biases, especially the emergent bias could potentially become an issue for PASSION, when the project starts to become publicly available teledermatology. Also, the interface design should be evaluated, so that no presentation or ranking bias gets introduced.

used

- **Emergent Bias.** Emergent bias occurs as a result of use and interaction with real users. This bias arises as a result of change in population, cultural values, or societal knowledge usually some time after the completion of design [8]. This type of bias is more likely to be observed in user interfaces, ... This type of bias can itself be divided into more subtypes, as discussed in detail in [8]. [13]. probably less relevant at the first stage
- **User Interaction Bias.** User Interaction bias is a type of bias that can not only be observant on the Web but also get triggered from two sources—the user interface and through the user itself by imposing his/her self-selected biased behavior and interaction [1]. This type of bias can be influenced by other types and subtypes, such as presentation and ranking biases. [13]. – more relevant for later, when the application would become bigger

- **Presentation Bias.** Presentation bias is a result of how information is presented [1] (can only click on content they see, could be the case that user does not see all info on web) [13].
- **Ranking Bias.** The idea that top-ranked results are the most relevant and important will result in attraction of more clicks than others. This bias affects search engines [1] and crowdsourcing applications [12].[13].

**External Influence Biases** Those biases can be introduced :

- **Evaluation Bias.** When inappropriate or disproportionate benchmarks are used in model evaluation, they can introduce the benchmarks biases into the model. [21, 2]
- **Incorporation bias.** When index tests in diagnostic accuracy studies are part of the reference tests, this results in elevated sensitivity for the index tests [c21, c25, c26, 3, 23].
- **Popularity Bias.** More popular items tend to be exposed more. Popularity metrics can be manipulated though or not reflecting good quality, this can lead to bias [4, 13].
- **Generalization Issues.** [<empty citation>]

#### Potential Biases in PASSION

used

- **Evaluation Bias.** Evaluation bias happens during model evaluation [21]. This includes the use of inappropriate and disproportionate benchmarks for evaluation of applications such as Adience and IJB-A benchmarks. These benchmarks are used in the evaluation of facial recognition systems that were biased toward skin color and gender [2], and can serve as examples for this type of bias [21]. [13]. – important for this thesis
- **Incorporation bias:** This is principally relevant for diagnostic accuracy studies when the index test forms a part of the reference test, resulting in elevated sensitivity e.g., if one wants to compare the grattage test vs. dermoscopy in psoriasis and does dermoscopy only from areas of grattage positivity, one would get a very high sensitivity for the grattage test because it was incorporated into the reference test, i.e., dermoscopy.25,26 21.[3]

add  
those  
from  
young

add

- **Popularity Bias.** Items that are more popular tend to be exposed more. However, popularity metrics are subject to manipulation—for example, by fake reviews or social bots [4]. ... this presentation may not be a result of good quality; instead, it may be due to other biased factors. [13].

#### 1.15.4 User Biases

**Cognitive Biases** Biases which are related to human perception belong to the category of cognitive biases. They are affecting how data should be presented and interpreted [14]

Those biases can be introduced :

- **Confirmation Bias.** When people have pre-conceptions, they will only listen to the part of presented information which reinforce those "facts", regardless whether the facts are true or not [14]. In health-care, this can be observed when patients report increases in diseases due to potentially nonfactual information they found on the internet [c15, c14, 3].
- **Belief Bias.** A stronger version of the confirmation bias: Someone who is affected by this bias is so sure about their own gut feelings that they will ignore results of a data research project [14].
- **Previous Opinion Bias.** When performing multiple tests, the knowledge about the outcome of the previous tests probably influences the results [3]
- **Cause-Effect Bias.** The famous sentence "correlation does not imply causation" can be used here - when correlation between two variables is misinterpreted as a cause-effect in the wrong direction, this bias applies [14]
- **Historical Bias.** Preexisting biases in the world can affect the data generation process [21]. Even if they reflect the current reality, it is worth to consider whether those biases should affect the algorithms in question [13].
- **Content Production Bias.** User generated contents can introduce biases by systematical differences in the production process, stucture and appearance, which might stem from the users background [18].



**Potential Biases in PASSION** For PASSION, confirmation bias could lead to issues in the initial diagnosis and could therefore lead to biased data labeling. Same with the previous opinion bias. The later can be reduced when it is ensured, that the labeling experts are diagnosing the diseases independently of each other, so that they do not know the previous opinions. Cause-Effect bias is lesser an issue for PASSION, since the causes of the diseases are not analyzed. It could more be an inherit problem, that the algorithm learns wrong causes for diseases, such as appearing hair Historical bias can affect PASSIONs process in various ways. In PASSION context, Content Production Bias could have an impact on how the images are taken.

used

- 
- Cognitive bias [14] - statistical bias
- Previous opinion bias: In performing a second diagnostic test, if the result of a previous test is known, it is likely to influence the result. An extension of this is the Greenwald's law of lupus: the Sontheimer amendment – anything and everything that happens to a lupus erythematosus patient is correctly or incorrectly attributed to lupus.<sup>32 29</sup>. [3] - dermatology
- Confirmation bias: This bias occurs when study participants have a preconceived notion of their disease that may not be based on facts. For example, we have observed that in North India many tinea patients report an increase in their disease due to taking meat, fish, and other so-called “hot foods”. They may also present information they have collected from the internet which reinforces their beliefs.<sup>15 14</sup>. [3] - dermatology
- Cause-effect bias [15, 14] - statistical bias
- Historical Bias. Historical bias is the already existing bias and socio-technical issues in the world and can seep into from the data generation process even given a perfect sampling and feature selection [21]. ... search results were of course reflecting the reality, but whether or not the search algorithms should reflect this reality is an issue worth considering [13] - maybe relevant
- Content Production Bias. Content Production bias arises from structural, lexical, semantic, and syntactic differences in the contents generated by users [18]. [13] – could the quality of the pictures been related to this as well?

**Behavioral Biases** Those biases can be introduced :

- **Behavioral Bias.** User behaviour can differ depending on the platforms, contexts, cultures, or datasets [18].
- **Self-Selection Bias.** This subtype of selection bias occurs when study participants can select themselves. Less proactive people, people with less time or interest will be excluded or underrepresented [15, 13]. **Non-Responder bias** is a subtype, where part of the population is not responding e.g. to fill out a survey or post-study responses queried by postal services [3].
- **Social Bias.** When the actions of others affect our judgment, it is called social bias. For example ratings in juries can be affected by this [1].

**Potential Biases in PASSION** For PASSION the behavioral biases can affect who is going to the dermatologists for what reasons. Therefore, the approach to use data from different countries may be beneficial, since potentially the cultural differences could differ. Self-selection is an issue, since only those patients can be included in the database which go to the hospitals.

used

- **Self-Selection Bias.** Self-selection bias<sup>4</sup> is a subtype of the selection or sampling bias in which subjects of the research select themselves. [13]
- **Self-selection bias** - when you let the subjects of the analyses select themselves, less proactive people will be excluded [15, 14]- statistical bias A variation of this is non-responder bias, where non-responders to a questionnaire differ significantly from responders.<sup>9</sup> [3] - dermatology
- **Social Bias.** Social bias happens when others' actions affect our judgment [1]. (case where we want to rate or review an item with a low score, but when influenced by other high ratings, we change our scoring thinking that perhaps we are being too harsh [[1], [22].) [13]
- **Behavioral Bias.** Behavioral bias arises from different user behavior across platforms, contexts, or different datasets [18]. [13] maybe, people from different countries go to the dermatologist for different diseases, based on cultural differences?

maybe categorize this in the data biases or the healthy volunteer bias here

could be an issue as well for PASSION, couldn't it? since the doctors probably ask the clients. One way to go is to default should be to provide access to the data. but is it ethical?

**Publication Biases** Those biases can be introduced :

- **Publication Bias**
- **Hot stuff bias** is a subtype of publication bias, where Journals are less critical about trending topics, which lead to more frequent publishing of those topics. This in turn can lead to flawed meta-analyses regarding those topics [c22, c23, c19, 3].
- **All is Well Bias.** This bias is a different view on the hot stuff bias. Theories which align with the view of the majority are more likely to be published than an opposing view [c7, c10-12, 3].
- **Rethoric Bias.** Charismatic writing or when the press is more vocal about findings can lead to greater influence over individuals than other available facts [3].
- **Novelty Bias.** Newer interventions appear to be better. Over time, this effect decreases [3].

**Potential Biases in PASSION** These biases are relevant for all researchers. They should kept in mind when interpreting, publishing and peer-reviewing papers.

used

- Hot stuff bias: Editors of journals may be less critical about topics that are “fashionable” or currently in vogue and consequently end up publishing them more frequently, resulting in publication bias as well as hot stuff bias. It can result in flawed meta-analyses based on these studies. An example is how cutaneous manifestations of COVID-19 were published. Indian Journal of Dermatology Venereology and Leprosy stood out by choosing not to publish anything and everything related to COVID-19, thus reducing hot stuff bias.22,23 19. [3]
- All is well bias: It is a subjective bias where theories supported by the majority tend to get more easily published than the opposing view supported by the minority. For example, ideas on the origin of endemic pemphigus supporting autoimmunity are more likely to be published than theories exploring an infectious trigger. According to some authors, this bias is very difficult to eliminate and is a variant of publication bias.10-12 7.[3] - dermatology

- **Rhetoric bias:** A more charismatic piece of writing has a greater influence on the study participants than other available literature. An example is the wider use of sunscreen for polymorphous light eruption over photoprotective strategies like umbrellas, broadbrimmed hats, etc, because the lay press is more vocal about sunscreens.<sup>14</sup> [3] - dermatology
- **Novelty bias:** The newer an intervention, the better it appears, and with time, its efficacy seems to decrease. When ligelizumab, an IgE antagonist was first discovered, ligelizumab was believed to be better than omalizumab; however, evidence soon pointed to the contrary. 16.[3] - dermatology

**Medical Biases** Those biases can be introduced :

- **Popularity Bias.** In medicine, when more popular diseases (usually well-known or stigmatized ones) get compared with less popular diseases, clinic rates can show a distorted view. The more popular diseases appear to be over-represented over more commoner ones [c9, c6, 3].
- **Apprehension Bias.** Fear related to an upcoming procedure can lead to false evaluations, e.g. when measuring blood pressure [c13, 3].
- **Hawthorne bias.** Subjects might modify their behaviour when they know they are being watched. This bias can be practically utilized by introducing regular follow-ups [c8, 3].
- **Centripetal Bias.** Better reputations affect to which physicians or hospitals patients tend to go to. Famous specialists probably see more cases in regards of their specialty than others [c12, 3].

**Potential Biases in PASSION** PASSION must be careful in interpreting the metadata. Since the data is from hospitals, they could be biased towards more popular diseases. PASSION can potentially use Hawthorne bias to improve the work of the annotators. Centripetal bias can also be used when selecting the partners to work with.

used

- **Popularity Bias:** This bias arises when a particular disease is more popular (i.e. either more well-known or more stigmatised) among the participants than the disease with which it is compared. For example,

if a study compares clinic attendance rates among various dermatological disorders, one would see vitiligo patients are over-represented over melasma. While melasma is commoner in the normal population, vitiligo, due to its popularity because of media publicity and other factors, tends to present earlier.<sup>9 6.</sup> [3] - dermatology

- Apprehension bias: This results from fear and apprehensions related to an impending procedure. The classic example is the false elevation of blood pressure because the person is apprehensive of his or her blood pressure being measured.<sup>13</sup> A variant of this is the Hawthorne bias, where subjects modify their behavior, such as regularly taking a prescribed drug or exercising, simply because they know they are being watched, but not due to any apprehensions. Hawthorne bias is practically utilised in many leprosy clinics since regular follow-up has been shown to improve adherence to therapy based on Hawthorne bias. <sup>8.</sup> [3] - dermatology
- Centripetal bias: Patients tend to go to more reputed physicians and hospitals than others. For example, a famous or better-known cosmetologist with a good reputation tends to see more cases than other cosmetologists. <sup>12.</sup>[3] - dermatology

## Appendix Bibliography

- [1] Ricardo Baeza-Yates. “Bias on the web”. In: *Commun. ACM* 61.6 (May 23, 2018), pp. 54–61. ISSN: 0001-0782. DOI: 10.1145/3209581. URL: <https://dl.acm.org/doi/10.1145/3209581> (visited on 03/16/2025).
- [2] Joy Buolamwini and Timnit Gebru. “Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification”. In: *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*. Conference on Fairness, Accountability and Transparency. ISSN: 2640-3498. PMLR, Jan. 21, 2018, pp. 77–91. URL: <https://proceedings.mlr.press/v81/buolamwini18a.html> (visited on 03/16/2025).
- [3] Atreyo Chakraborty. “Biases in dermatology: A primer”. In: *Indian J Dermatol Venereol Leprol* 90.2 (Feb. 20, 2024). Publisher: Scientific Scholar, pp. 250–254. ISSN: 0378-6323. DOI: 10.25259/IJDVL\_126\_2023. URL: <https://ijdv1.com/biases-in-dermatology-a-primer/> (visited on 03/10/2025).
- [4] Giovanni Luca Ciampaglia et al. “How algorithmic popularity bias hinders or promotes quality”. In: *Sci Rep* 8.1 (Oct. 29, 2018). Publisher: Nature Publishing Group, p. 15951. ISSN: 2045-2322. DOI: 10.1038/s41598-018-34203-2. URL: <https://www.nature.com/articles/s41598-018-34203-2> (visited on 03/16/2025).
- [5] Kevin A. Clarke. “The Phantom Menace: Omitted Variable Bias in Econometric Research”. In: *Conflict Management and Peace Science* 22.4 (Sept. 1, 2005). Publisher: SAGE Publications Ltd, pp. 341–352. ISSN: 0738-8942. DOI: 10.1080/07388940500339183. URL: <https://doi.org/10.1080/07388940500339183> (visited on 03/16/2025).
- [6] David Danks and Alex John London. “Algorithmic Bias in Autonomous Systems”. In: *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*. Twenty-Sixth International Joint Conference on Artificial Intelligence. Melbourne, Australia: International

- Joint Conferences on Artificial Intelligence Organization, Aug. 2017, pp. 4691–4697. ISBN: 978-0-9992411-0-3. DOI: 10.24963/ijcai.2017/654. URL: <https://www.ijcai.org/proceedings/2017/654> (visited on 03/16/2025).
- [7] M. Delgado-Rodríguez and J. Llorca. “Bias”. In: *Journal of Epidemiology & Community Health* 58.8 (Aug. 1, 2004). Publisher: BMJ Publishing Group Ltd Section: Continuing professional education, pp. 635–641. ISSN: 0143-005X, 1470-2738. DOI: 10.1136/jech.2003.008466. URL: <https://jech.bmj.com/content/58/8/635> (visited on 03/29/2025).
  - [8] Batya Friedman and Helen Nissenbaum. “Bias in computer systems”. In: *ACM Trans. Inf. Syst.* 14.3 (July 1, 1996), pp. 330–347. ISSN: 1046-8188. DOI: 10.1145/230538.230561. URL: <https://dl.acm.org/doi/10.1145/230538.230561> (visited on 03/16/2025).
  - [9] Anna Fry et al. “Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population”. In: *American Journal of Epidemiology* 186.9 (Nov. 1, 2017), pp. 1026–1034. ISSN: 0002-9262. DOI: 10.1093/aje/kwx246. URL: <https://doi.org/10.1093/aje/kwx246> (visited on 04/03/2025).
  - [10] Eszter Hargittai. “Whose Space? Differences among Users and Non-Users of Social Network Sites”. In: *Journal of Computer-Mediated Communication* 13.1 (Oct. 1, 2007), pp. 276–297. ISSN: 1083-6101. DOI: 10.1111/j.1083-6101.2007.00396.x. URL: <https://doi.org/10.1111/j.1083-6101.2007.00396.x> (visited on 03/16/2025).
  - [11] Suresha HP. *Sampling — Statistical approach in Machine learning*. Analytics Vidhya. Nov. 1, 2022. URL: <https://medium.com/analytics-vidhya/sampling-statistical-approach-in-machine-learning-4903c40ebf86> (visited on 03/28/2025).
  - [12] Kristina Lerman and Tad Hogg. “Leveraging Position Bias to Improve Peer Recommendation”. In: *PLOS ONE* 9.6 (June 11, 2014). Publisher: Public Library of Science, e98914. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0098914. URL: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0098914> (visited on 03/16/2025).
  - [13] Ninareh Mehrabi et al. “A Survey on Bias and Fairness in Machine Learning”. In: *ACM Computing Surveys (CSUR)* (July 13, 2021). Publisher: ACM/PUB27New York, NY, USA. DOI: 10.1145/3457607. URL: <https://dl.acm.org/doi/10.1145/3457607> (visited on 02/28/2025).

- [14] Tomi Mester. *Statistical Bias Types explained - part2 (with examples)*. Data36. Aug. 28, 2017. URL: <https://data36.com/statistical-bias-types-examples-part2/> (visited on 03/22/2025).
- [15] Tomi Mester. *Statistical Bias Types explained (with examples) - part1*. Data36. May 16, 2022. URL: <https://data36.com/statistical-bias-types-explained/> (visited on 03/08/2025).
- [16] Laura N. Montoya, Jennafer Shae Roberts, and Belén Sánchez Hidalgo. “Towards Fairness in AI for Melanoma Detection: Systemic Review and Recommendations”. In: *Advances in Information and Communication*. Ed. by Kohei Arai. Cham: Springer Nature Switzerland, 2025, pp. 320–341. ISBN: 978-3-031-84460-7. DOI: 10.1007/978-3-031-84460-7\_21.
- [17] David B. Mustard. “Reexamining Criminal Behavior: The Importance of Omitted Variable Bias”. In: *The Review of Economics and Statistics* 85.1 (Feb. 1, 2003), pp. 205–211. ISSN: 0034-6535. DOI: 10.1162/rest.2003.85.1.205. URL: <https://doi.org/10.1162/rest.2003.85.1.205> (visited on 03/16/2025).
- [18] Alexandra Olteanu et al. “Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries”. In: *Front. Big Data* 2 (July 11, 2019). Publisher: Frontiers. ISSN: 2624-909X. DOI: 10.3389/fdata.2019.00013. URL: <https://www.frontiersin.org/journals/big-data/articles/10.3389/fdata.2019.00013/full> (visited on 03/16/2025).
- [19] Stephanie K. Riegg. “Causal Inference and Omitted Variable Bias in Financial Aid Research: Assessing Solutions”. In: *The Review of Higher Education* 31.3 (2008). Publisher: Johns Hopkins University Press, pp. 329–354. ISSN: 1090-7009. URL: <https://muse.jhu.edu/pub/1/article/232773> (visited on 03/16/2025).
- [20] Shreya Shankar et al. *No Classification without Representation: Assessing Geodiversity Issues in Open Data Sets for the Developing World*. Nov. 22, 2017. DOI: 10.48550/arXiv.1711.08536. arXiv: 1711.08536[stat]. URL: <http://arxiv.org/abs/1711.08536> (visited on 04/03/2025).
- [21] Harini Suresh and John Guttag. “A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle”. In: *Proceedings of the 1st ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*. EAAMO ’21. New York, NY, USA: Association for Computing Machinery, Nov. 4, 2021, pp. 1–9. ISBN: 978-1-4503-8553-4. DOI: 10.1145/3465416.3483305. URL:



- <https://dl.acm.org/doi/10.1145/3465416.3483305> (visited on 03/16/2025).
- [22] Ting Wang and Dashun Wang. “Why Amazon’s Ratings Might Mislead You: The Story of Herding Effects”. In: *Big Data* 2.4 (Dec. 2014). Publisher: Mary Ann Liebert, Inc., publishers, pp. 196–204. ISSN: 2167-6461. DOI: 10.1089/big.2014.0063. URL: <https://www.liebertpub.com/doi/full/10.1089/big.2014.0063> (visited on 03/16/2025).
- [23] Albert T. Young et al. “Artificial Intelligence in Dermatology: A Primer”. In: *Journal of Investigative Dermatology* 140.8 (Aug. 1, 2020), pp. 1504–1512. ISSN: 0022-202X. DOI: 10.1016/j.jid.2020.02.026. URL: <https://www.sciencedirect.com/science/article/pii/S0022202X2031201X> (visited on 03/23/2025).