Contextual Bias Score - A Quantitative Metric for Gender Bias Detection in Visual Datasets

|  |  |  |  |
| --- | --- | --- | --- |
| Amandi Ekanayake *Faculty of Computing* *Sri Lanka Institute of Information Technology* Malabe, Sri Lanka amandi.mithila@gmail.com | Tharushi Mudalige *Faculty of Computing* *Sri Lanka Institute of Information Technology* Malabe, Sri Lanka tharushimudalige01@gmail.com | Imesha Dhanawardhana  *Faculty of Computing* *Sri Lanka Institute of Information Technology* Malabe, Sri Lanka imesha.kaviz99@gmail.com | Sriwanthi Jayawardena *Faculty of Computing* *Sri Lanka Institute of Information Technology* Malabe, Sri Lanka sriwanthij@gmail.com |
|  |  |  |  |
|  | Prasanna S. Haddela *Faculty of Computing* *Sri Lanka Institute of Information Technology* Malabe, Sri Lanka prasanna.s@sliit.lk | Thisara Shyamalee  *Faculty of Computing* *Sri Lanka Institute of Information Technology* Malabe, Sri Lanka thisara.s@sliit.lk |  |

*Abstract*— Artificial Intelligence (AI) systems in computer vision often inherit contextual biases from training datasets, resulting in unintended gender associations during visual recognition. This paper introduces the Contextual Bias Score (CBS), a novel metric for quantifying gender bias in image datasets by evaluating object prominence, 3D spatial relationships, and scene-level semantics. Unlike conventional approaches that rely on co-occurrence statistics, CBS integrates object-person spatial features, relative size, depth, and distance with CLIP-based scene embeddings to assess how visual context influences bias. The framework is evaluated across nine experimental approaches using eight COCO-derived datasets, leveraging SHAP-based interpretability, PCA-based feature weighting, and Ridge regression modeling. Approach 9 (SHAP + Ridge) achieves 85% agreement with human-annotated misclassification labels and strong score stability. Validation includes statistical significance testing, human perception alignment, and cross-approach correlation analysis. CBS offers a dataset-independent, interpretable tool for auditing contextual bias in high-impact domains. Limitations include binary gender assumptions and computational overhead (~90 seconds per image), with future work aimed at extending the framework to non-binary and culturally diverse contexts.

Keywords— Contextual bias, gender bias, fairness in AI, model interpretability, object-person relationships, scene embeddings.

# Introduction

Artificial Intelligence (AI) and computer vision have significantly advanced object recognition, scene understanding, and automated decision-making. However, AI models often inherit contextual biases from training datasets, leading to skewed and unintended predictions. Contextual bias occurs when an AI system’s decision is influenced not just by the main object, but also by its size, surrounding elements, spatial positioning, and background scene. For example, an AI model may associate sports equipment with men and household items with women, reinforcing societal stereotypes rather than making neutral classifications [1]. These biases become particularly problematic in hiring algorithms, surveillance, and AI-generated media, where biased model decisions can perpetuate discrimination and reinforce gender disparities [2].

Existing bias evaluation frameworks, such as REVISE and Fairness AI tools, primarily focus on co-occurrence statistics and demographic imbalances but fail to assess how spatial positioning, object size, 3D relationships, and depth contribute to biased predictions [3]. Research has shown that AI vision models can develop biases based on spatial dependencies and object-background interactions, which existing fairness assessments do not fully quantify [4]. While attribution-based methods such as Grad-CAM and Score-CAM highlight bias visibility, they lack a structured scoring mechanism to quantify contextual bias in AI models [5].

To address this limitation, this research introduces a Contextual Bias Score a metric for measuring gender bias in AI-driven image datasets based on object prominence, spatial positioning, 3D distance, depth perception, and scene context. Unlike conventional methods that focus solely on who appears in an image, CBS evaluates how objects are represented in relation to their surroundings, allowing AI researchers to measure whether an object exhibits male, female, or neutral bias. By integrating bounding box dimensions, depth maps, and scene-embedding, this approach provides a structured, dataset-independent framework for bias measurement.

This study does not aim to mitigate bias but rather to quantify it, providing AI practitioners with a metric that helps assess the extent of contextual bias in image datasets. By offering a systematic approach to evaluating contextual dependencies, this research contributes to the broader effort of understanding and diagnosing bias in AI-driven visual systems.

# literature review

## Contextual Bias in AI-Driven Image Classification

Several studies have investigated how AI-driven image classification models exhibit contextual bias, where predictions are influenced by surrounding elements rather than the primary object. Wang et al. [1] introduced REVISE, a tool designed to analyze dataset imbalances by evaluating biases across object-based, person-based, and geography- based dimensions. Their approach highlights the extent of dataset-driven biases but lacks an explicit mechanism to quantify how object positioning and spatial elements influence model predictions.

Sabir and Padró [2] conducted an in-depth study on how object size and distance contribute to gender bias in AI models. Their research found that models often associate larger objects with male identities and smaller objects with female identities, reinforcing gender stereotypes in classification tasks. Additionally, they demonstrated that the spatial distance between a person and an object can influence gender-based predictions, meaning that AI models rely on relative object prominence rather than neutral classification mechanisms. However, their approach focuses on statistical co-occurrence trends, leaving room for a more structured numerical bias measurement that accounts for spatial positioning and depth.

Schaaf et al. [3] conducted an analysis on bias visibility in AI models using attribution maps like Grad-CAM and Score-CAM, revealing that models often rely on irrelevant background features instead of meaningful object cues. Their findings confirm that contextual bias extends beyond dataset composition and is deeply embedded in how AI models interpret images. Liu et al. [4] advanced this research by proposing Causal Context Debiasing (CCD), which utilizes structural causal models (SCMs) to mitigate biases caused by spurious contextual correlations. While their work effectively intervenes in bias, it does not introduce a quantitative metric to measure the degree of bias in contextual settings.

Recent work by Meister et al. [6] introduced the concept of gender artifacts—visual features such as background scenes, lighting, and object co-occurrence that implicitly influence gender predictions. They showed that models trained on visual datasets like COCO and OpenImages often learn gender associations from non-human elements. Frey et al. [7] further supported this by demonstrating that CNNs underperform when interpreting semantically incongruent scenes, unlike humans who adjust more flexibly. These findings underscore the importance of modeling scene-level bias cues, as implemented in the Scene Similarity Bias (SSB) component of CBS.

## Existing Bias Measurement Approaches

Mauro et al. [8] examined bias in dermatological AI datasets, demonstrating that factors like viewpoint variation, lighting, and class distribution imbalances significantly impact bias perception. Their research emphasizes the need for a structured bias quantification method beyond dataset corrections. Similarly, Al Sahili et al. [5] analyzed vision-language models, revealing that scaling dataset size does not necessarily reduce bias in some cases, larger datasets amplify existing biases.

In contrast, Govil et al. [9] introduced COBIAS, a context-aware bias assessment framework for language models. Their work demonstrated that context plays a key role in bias perception, but no comparable metric exists for image-based bias quantification. Siddik et al. [10] extended this research into healthcare AI, proposing a bias documentation framework to increase transparency in dataset curation and model validation. These studies collectively reinforce the need for a numerical bias metric tailored to visual AI systems.

Herranz et al. [11]studied scale and spatial biases in CNN-based scene recognition systems, concluding that object size and layout significantly affect classification outcomes. Their findings support the inclusion of size-normalized components in fairness metrics like CBS. Similarly, French and DeAngelis [12] showed that object motion relative to background affects depth perception in humans, paralleling the idea that CNNs may interpret contextual prominence as a cue for classification. These studies highlight the importance of incorporating depth and scale into bias quantification.

## Research Gaps in Bias Measurements

Despite advancements in bias detection, existing methodologies do not fully address the influence of contextual factors on AI model predictions. Many studies primarily rely on explainability techniques such as attribution maps, which highlight bias visibility but do not provide a structured numerical bias score. Additionally, current models do not fully account for 3D contextual analysis, as they often ignore object depth, spatial relationships, and scene interactions. Furthermore, most fairness evaluation tools are designed for specific datasets and applications, making them difficult to generalize across different AI models, particularly those used in facial recognition, automated hiring, and surveillance technologies. These limitations indicate a need for a dataset-independent approach that quantifies contextual bias beyond qualitative assessments.

## Contribution of this Research

To explore these challenges, this study introduces the Contextual Bias Score (CBS), a metric aimed at quantifying gender bias in AI-driven image datasets by considering object prominence, 3D spatial relationships, depth-based context, and co-occurrence dependencies. Unlike previous methods that primarily assess who appears in an image, CBS shifts the focus to how objects are represented within their surroundings and the degree to which contextual factors influence classification outcomes. As an initial investigation, CBS is evaluated on a selected dataset, with the goal of assessing its feasibility in bias measurement. Further research will be required to determine its applicability across multiple datasets and refine the metric based on observed limitations. This study serves as a starting point for developing structured contextual bias evaluation methods, with potential applications in facial recognition, automated hiring, and surveillance AI systems.

# METHODOLOGY

## Overview of the Research Design

This study adopts a quantitative, model-driven approach to detect and measure gender bias in image datasets. The proposed methodology evaluates object-person spatial relationships and scene-level context to assess their influence on gender classification. A multi-stage pipeline is developed to extract object attributes, depth information, and semantic scene features. These components are integrated into a unified measurement framework referred to as the Contextual Bias Score (CBS), which quantifies the potential bias present in object-gender associations. Visual analytics and embedding-based similarity scores are used to support the robustness of the findings.

## Data Collection

A filtered subset of the Microsoft COCO dataset [13] was used, comprising 852 images containing persons co-occurring with objects commonly associated with gender stereotypes. The selected object categories include handbag, hair drier, umbrella, cup, sports ball, baseball bat, bicycle, and skateboard. Each image was manually verified to contain at least one person and one target object.

Object detection and segmentation were performed using the YOLOv8 model, followed by refinement with the Segment Anything Model (SAM) [14] for enhanced mask precision. These pre-trained models were adapted from their original open-source implementations and repurposed to support contextual bias analysis in this work. All annotations, including bounding boxes, masks, and object labels, were retained for downstream analysis.

|  |  |
| --- | --- |
|  |  |

Fig. 1. Object detection and segmentation using YOLOv8 and SAM. The left panel shows the original test image. The right panel illustrates bounding boxes around the detected person and object (umbrella), along with mask overlays used for bias feature extraction.

## Feature Extraction

To support contextual bias measurement, multiple features were extracted from each image:

### Relative Size

The size of each object is measured relative to the detected person:

### 3d Distance Computation

Using DepthAnything v2 [15], grayscale depth maps were generated to estimate the spatial separation between persons and objects.



Fig. 2. Depth estimation for person and object using DepthAnything. The left image shows the original input, while the right displays the estimated grayscale depth map. These outputs are used to calculate Z-axis separation and combined 3D distance between person and object for contextual bias computation.

* 2D Euclidean distance in the image plane:
* Depth (Z-axis) separation:
* Combined 3D distance:

### Scene Similarity Bias

To evaluate contextual alignment with gendered environments, scene embeddings were extracted using a pre-trained CLIP model [16]. The CLIP model was adapted for scene-level bias quantification by comparing embeddings against curated male- and female-dominant reference scenes.

Two reference vectors were created:

Each scene vector is compared to these references via cosine similarity:

The final Scene Similarity Bias (SSB) score is calculated as:

A positive SSB score suggests the scene is more aligned with male-associated settings, while a negative score indicates a female-dominant context.

## Bias Metric Design and Approaches

This section presents the design of the Contextual Bias Score (CBS), a novel metric for quantifying gender bias in image datasets. CBS integrates object-level spatial features and high-level semantic scene context. Eight experimental approaches are proposed to evaluate the robustness of this metric using different learning and weighting strategies.

### Contextual Bias Score (CBS) Design

The Contextual Bias Score (CBS) is designed to capture how strongly certain objects and scenes influence gender associations in visual data. It combines two core components:

* **Object Influence Score (OIS):** Captures spatial prominence of objects relative to persons using three normalized features:
* Relative Size (RS): Area ratio of the object to the person.
* 3D Distance (D3): Normalized Euclidean distance between object and person calculated using depth maps.
* Depth Difference (DZ): Absolute difference in average depth between object and person.

These are combined with weighted contributions:

Weights are computed using dimensionality reduction (PCA), explainability techniques (SHAP), or regression models, depending on the approach.

* **Scene Similarity Bias (SSB):** Reflects the alignment of a scene’s semantics with gender-associated contexts. Scene embeddings are extracted using a pretrained CLIP model, and cosine similarity is computed against reference vectors:

A positive value suggests male-oriented context; a negative value indicates female-oriented context.

* **Final Bias Score:** The unified bias score is then computed as:

Where and are weights determined through PCA loading, SHAP analysis or regression analysis.

### Bias Score Computational and Categorization

To assess directional gender bias for each object class, the **Bias Score (BS)** is computed from group-wise CBS averages:

The small constant is added to prevent division by zero or instability when denominator values are near zero.

A data-driven thresholding strategy is used to classify object classes:

* Compute the mean () and standard deviation () of all BS values.
* Define thresholds as:
  + **Male-Biased:**
  + **Female-Biased:**
  + **Neutral:**

This adaptive method accounts for the actual distribution of bias scores and avoids reliance on arbitrary thresholds.

### Experimental Approaches for Bias Analysis

# To test how well the Contextual Bias Score (CBS) works under different settings, nine experimental approaches were developed. Each approach uses a different way of calculating feature importance, training models, or balancing data. The goal is to see how these changes affect bias detection results.

#### Approach 1: PCA-Based Unified Bias Metric(UBM)

This approach uses Principal Component Analysis (PCA) to get weights for the spatial features: Relative Size (RS), 3D Distance (D3), and Depth Difference (DZ). These weights are used to calculate the Object Influence Score (OIS), which is then combined with the Scene Similarity Bias (SSB) to compute the CBS.

#### Approach 2: SHAP-Based Feature Attribution

A Random Forest model is trained to predict gender. SHAP values are used to find out how much each feature contributes. SHAP explainability is adapted from [17] to quantify the contribution of spatial and scene features in gender bias scoring. These values are used as weights to calculate OIS. CBS is then computed by combining this OIS with SSB.

#### Approach 3: Hybrid SHAP + PCA Weights

This method mixes the weights from SHAP and PCA by averaging them. The combined weights are used to compute OIS. Then, OIS and SSB are combined to get CBS.

#### Approach 4: XGBoost with SHAP Explanation

XGBoost is used as the model to predict gender. SHAP values are calculated to find the importance of each feature. These are used to compute OIS, and then combined with SSB to form the final CBS.

#### Approach 5: Balanced XGBoost with SMOTE

SMOTE is used to balance the number of male and female examples in the dataset. Then, XGBoost is trained on this balanced data. SHAP values are used to get feature weights, which are used to compute OIS and then CBS.

#### Approach 6: SHAP + PCA on Balanced Data

This method is similar to Approach 3 but uses the balanced dataset from Approach 5. SHAP and PCA weights are combined, and used to compute OIS and CBS.

#### Approach 7: Co-Occurrence-Aware Correction

This approach adjusts the CBS using how often each object appears with male or female labels. This helps reduce bias caused by uneven object-gender pairings in the dataset.

#### Approach 8: Ridge Regression-Based Optimization

A Ridge Regression model is trained to learn the best weights for the spatial features. These learned weights are used to calculate OIS, which is then combined with SSB to get CBS.

#### Approach 9: Optuna-Optimized XGBoost with SHAP

This approach uses Optuna to automatically find the best settings for the XGBoost model. SHAP values are taken from the optimized model and used to calculate OIS. OIS is then combined with SSB to produce the final CBS

A graph of a graph

AI-generated content may be incorrect.

Fig. 3: Bias scores by object class using the final unified bias detection approach (Approach 9). Bars are colored based on categorized bias: male-biased (blue), female-biased (red), and neutral (gray).

Fig. 3 visualizes the object-level bias scores produced by the final unified bias detection method (Approach 9). It categorizes each object as male-biased, female-biased, or neutral based on their contextual bias scores. Strong male bias is observed in objects like *baseball bat*, while *hair drier* and *handbag* show significant female bias.

All nine approaches output object-level CBS values, bias scores, and supporting visualizations for interpretation. The differences between approaches are explained above.

# EVALUATION AND VALIDATION

## Experimental Setup

The proposed Contextual Bias Score (CBS) framework was evaluated across nine scoring approaches on eight person-centric subsets of the Microsoft COCO dataset. All analyses used real-world COCO images; no simulated data was involved. Each dataset was filtered to include images with at least one person and co-occurring objects that commonly exhibit gender associations (e.g., handbag, sports ball). The framework produced object-level bias scores, predicted bias categories, contextual feature weights, and interpretability outputs (e.g., SHAP values, PCA loadings, Ridge coefficients) for comprehensive evaluation.

## Metrics and Protocols

A multi-metric validation protocol was employed to assess robustness, fairness, and contextual alignment using eight evaluation metrics:

### Misclassification Agreement

Measures the extent to which predicted bias categories align with observed gender misclassifications in object-person pairs.

### Bias Category Distribution

Evaluates the proportion of objects labeled as male-biased, female-biased, and neutral, to assess prediction balance and interpretability.

### Cross-Approach Correlation

Computes Pearson and Spearman coefficients between object-level bias scores across approaches to assess methodological consistency.

### Score Divergence (Object-Level)

Identifies high-variance objects where bias scores significantly disagree across approaches.

### Score Stability (Approach-Level)

Quantifies the standard deviation of bias scores per approach across all datasets to identify stable versus volatile methods.

### Accuracy vs. Variability Correlation

Analyzes the relationship between standard deviation and agreement accuracy to assess whether scoring variability reflects improved predictive alignment.

### Composite Ranking

Aggregates normalized metrics (agreement rate, stability, coverage, divergence, correlation) to rank the overall effectiveness of each approach.

Additionally,

### Human Perception Survey

A human-centric validation was initiated via a Google Form, presenting 30 objects to annotators to assess perceived gender associations.

## Results Analysis

### Misclassification Agreement

Approaches were compared against gender classification mislabels. Approaches 1, 3, and 9 achieved the highest agreement scores averaging around **70% alignment** with real-world misclassification patterns on the original dataset. This indicates strong contextual alignment between predicted bias and observed model behavior.

### Bias Category Distribution

The distribution of bias categories across the nine evaluated approaches revealed important behavioral differences. Approaches 1, 3, and 9 maintained a relatively balanced output across male-biased, female-biased, and neutral classifications, indicating stable and interpretable scoring. In contrast, Approaches 5 and 6 produced a disproportionately high number of neutral labels, suggesting reduced sensitivity to contextual patterns or underlying model uncertainty. These findings reinforce that interpretability and balanced feature use support meaningful bias differentiation.

### Cross-Approach Correlation

A Spearman correlation analysis was conducted between object-level bias scores across all approaches. The results revealed a strong correlation among Approaches 1, 3, 8, and 9, indicating that these methods produce consistent and contextually aligned bias estimates. In contrast, Approaches 4 and 6 exhibited lower correlation with others, suggesting divergence in scoring behavior or reliance on differing feature extraction pipelines.

Notably, the highest pairwise correlation was observed between Approaches 8 and 9 ( 0.99), while Approaches 4 and 6 showed minimal alignment with most other methods, with correlation coefficients falling below 0.20.

### Score Divergence (Object-Level)

Analysis of score disagreements showed inconsistent bias predictions for key objects (e.g., umbrella, sports ball) in Approaches 4, 5, and 6, suggesting instability and reduced interpretability.

### Score Stability (Approach-Level)

Approach 2 exhibited the lowest global standard deviation (), indicating highly consistent scoring. In contrast, Approach 4 produced the highest variability (), reducing its reliability.

### Accuracy vs. Variability

A weak positive correlation (Pearson = 0.15) was observed between agreement accuracy and score variability, suggesting that score volatility alone is not indicative of bias alignment.

### Composite Ranking

Normalized scores across all metrics were aggregated to form a composite ranking. The top-performing approaches were:

* Approach 1: PCA + Random Forest
* Approach 3: SHAP + PCA
* Approach 9: SHAP + Ridge Regression These pipelines consistently demonstrated agreement with real-world signals, score stability, and interpretability.

### Human Perception Survey

A set of 30 survey questions was designed using high-bias and neutral objects identified by Approach 9, evenly divided into male-biased, female-biased, and neutral categories. Fifty participants were asked to select perceived gender associations for each object via a Google Form. The results revealed strong agreement between human judgments and CBS predictions, particularly for stereotypical objects such as baseball bat and sports ball (male-associated), and handbag, hair dryer, and teddy bear (female-associated). Neutral objects like bottle and cell phone were consistently perceived as unbiased. Overall, CBS predictions aligned with human perception in over 85% of the evaluated cases, reinforcing the contextual validity and interpretability of the model.

## Limitations

Despite comprehensive evaluation, the CBS framework faces several limitations:

* Binary Gender Assumptions: Only male and female classes were analyzed.
* Labeling Noise: Ground truth annotations for gender may contain classification inaccuracies.
* Data Sparsity: Low-frequency object-gender co-occurrences can reduce prediction confidence.
* Performance Overhead: Average processing time was 90 seconds per image, limiting real-time deployment.
* Subjectivity in Human Survey: Annotator perception may vary due to cultural or social factors.

Nonetheless, the CBS framework demonstrates robust, interpretable, and statistically validated performance across diverse scoring strategies and contextual conditions. Future work will extend CBS to non-binary gender contexts and optimize runtime.

# DISCUSSION

The evaluation highlights the importance of contextual understanding in detecting gender bias within visual datasets. The Contextual Bias Score (CBS) framework integrates object prominence, spatial relationships, and scene-level semantics to provide a structured and interpretable bias quantification method. Among nine approaches evaluated, hybrid models with interpretability significantly outperformed others. Approach 9 (SHAP + Ridge Regression) achieved the highest agreement, score stability, object coverage, and transparency. Approaches 1 (PCA + Random Forest) and 3 (SHAP + PCA) also performed reliably, while Approaches 4–6 showed inconsistent behaviour, emphasizing the role of model transparency and balanced feature use.

A weak correlation between score variability and accuracy (Pearson0.15) suggests that consistency alone is not a reliable indicator of bias alignment. Co-occurrence validation confirmed CBS's ability to capture contextual rather than statistical associations. While CBS provides detailed and interpretable outputs, its 90-second runtime per image may limit real-time deployment. Planned optimizations include vector caching, model distillation, and lightweight embedding techniques.

Future iterations will extend CBS to support non-binary gender representation and culturally diverse perception models. Ongoing human perception surveys will further validate the model's alignment with societal judgments. To support real-world adoption, a user-accessible toolkit or web interface is envisioned for automated bias diagnostics from uploaded CSVs.

# CONCLUSION

This research introduced the Contextual Bias Score (CBS), a structured metric for quantifying gender bias by integrating spatial, semantic, and scene features. Evaluation across nine approaches and eight datasets demonstrated that interpretable, hybrid models outperform co-occurrence-based methods.

Approach 9 (SHAP + Ridge Regression) emerged as the most effective, offering robust agreement, score stability, and transparency. CBS’s adaptability positions it well for AI fairness auditing in applications such as surveillance and hiring.

Although current limitations include binary gender assumptions and annotator subjectivity, future work will address these gaps by incorporating non-binary representation and cultural diversity. CBS provides a scalable, interpretable method for contextual bias detection, advancing the pursuit of ethical AI systems.

# References

|  |  |
| --- | --- |
| [1] | A. Wang, A. Narayanan, and O. Russakovsky, "REVISE: A Tool for Measuring and Mitigating Bias in Visual Datasets," European Conference on Computer Vision, 2020. |
| [2] | A. Sabir and L. Padró, "Women Wearing Lipstick: Measuring the Bias Between an Object and Its Related Gender," 2023. |
| [3] | Nina Schaaf, Omar de Mitri, Hang Beom Kim, Alexander Windberger, Marco F. Huber, "Towards Measuring Bias in Image Classification," 2021. |
| [4] | R. Liu, H. Liu, G. Li, H. Hou, T. Yu and T. Yang, "Contextual Debiasing for Visual Recognition with Causal Mechanisms," IEEE, 2022. |
| [5] | Zahraa Al Sahili, Ioannis Patras, Matthew Purver, "Scaling for Fairness? Analyzing Model Size, Data Composition, and Multilinguality in Vision-Language Bias," 2025. |
| [6] | Nicole Meister, Dora Zhao, Angelina Wang, Vikram V. Ramaswamy, Ruth Fong, Olga Russakovsky, "Gender Artifacts in Visual Datasets," 2023. |
| [7] | Haley G. Frey; Hojin Jang; Hui-Yuan Miao; Frank Tong, "The role of scene context in object recognition by humans and convolutional neural networks," 2023. |
| [8] | Francisco Filho, Emanoel Santos, Rodrigo Mota, Kelvin Cunha, Fabio Papais, Amanda Arruda, Mateus Baltazar, Camila Vieira, José Gabriel Tavares, Rafael Barros, Othon Souza, Thales Bezerra, Natalia Lopes, Érico Moutinho, Jéssica Guido, Shirley Cruz, Paulo B, "An analysis of data variation and bias in image-based dermatological datasets for machine learning classification," 2025. |
| [9] | Priyanshul Govil, Hemang Jain, Vamshi Krishna Bonagiri, Aman Chadha, Ponnurangam Kumaraguru, Manas Gaur, Sanorita Dey, "COBIAS: Contextual Reliability in Bias Assessment," 2024. |
| [10] | Marjia Siddik, Harshvardhan J. Pandit, "Datasheets for Healthcare AI: A Framework for Transparency and Bias Mitigation," 2025. |
| [11] | Luis Herranz, Shuqiang Jiang, Xiangyang Li, "Scene recognition with CNNs: objects, scales and dataset bias," 2018. |
| [12] | French, R.L., DeAngelis, G.C., "Scene-relative object motion biases depth percepts," 2022. |
| [13] | Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, Piotr Dollár, "Microsoft COCO: Common Objects in Context," 2014. |
| [14] | Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, Ross Girshick, "Segment Anything," 2023. |
| [15] | Lihe Yang, Bingyi Kang, Zilong Huang, Xiaogang Xu, Jiashi Feng, Hengshuang Zhao, "Depth Anything: Unleashing the Power of Large-Scale Unlabeled Data," 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024. |
| [16] | Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, Ilya Sutskever, "Learning Transferable Visual Models From Natural Language Supervision," 2021. |
| [17] | Scott Lundberg, Su-In Lee, "A Unified Approach to Interpreting Model Predictions," 2017. |