Digital Society (2024) 3:48
https://doi.org/10.1007/s44206-024-00142-x

REVIEW PAPER

Insights into Artificial Intelligence Bias: Implications for Agriculture

Mathuranathan Mayuravaani[1] · Amirthalingam Ramanan[1] · Maneesha Perera[2] · Damith Asanka Senanayake[2] · Rajith Vidanaarachchi[2]

Received: 20 February 2024 / Accepted: 10 September 2024 / Published online: 2 October 2024

Abstract
The integration of Artificial Intelligence (AI) has catalysed a paradigm shift in agricultural practices, revolutionising critical aspects of contemporary farming methodologies. In recent years, discrimination has become a significant topic of controversy in the use of AI-based systems. As AI becomes increasingly integrated into agriculture, the potential for bias within these systems raises significant concerns, as biases can affect farming practices, crop management, resource allocation, and the broader agricultural economy. Mitigating these issues is vital to ensure equitable access to technology, enhance farming practices for all, and foster sustainable and inclusive food production systems. This manuscript provides an extensive review of the literature on biases and fairness in AI focusing on the field of agriculture and the need for a holistic and inclusive approach to mitigate AI biases, ensuring that AI advancements benefit a broad spectrum of farming communities while promoting sustainability and equity.

Keywords Artificial Intelligence · AI biases · Precision agriculture · Bias mitigation · Fairness in AI · Ethical AI

1 Introduction

Artificial Intelligence (AI) is an attempt to recapitulate the essence of human intelligence within a computer system. While the truly mystifying aspects of human intelligence such as consciousness and self-awareness are still far from having a universally agreed upon model, some aspects such as memory, learning, and cognition are fairly well understood (Davenport & Ronanki, 2018). From statistical analyses to data-driven modelling, there have been successful attempts at capturing various aspects of intelligent systems. In these scenarios, the three essential components are

Extended author information available on the last page of the article

the data (that is the source of experience and knowledge which trains the intelligent system), the model (that is the structure with which we take in, process and output the data), and the learning algorithm (that is the method with which we find the model parameters to produce the most accurate outputs). In the modern context, data is abundant, it comes in various sizes and formats, and experts increasingly rely on AI systems to derive knowledge, interpretations and summaries from these data.

In the context of AI, bias refers to any prejudice, preconceived notions, or discriminatory tendencies embedded within the data or algorithms of the model. Most modern AI algorithms require large amount of data to train models. However, with this requirement and reliance on data, biases are inevitable in data modelling because algorithms are created by humans, these algorithms inherently and often unintentionally incorporate societal biases, values, and discriminatory tendencies into data modeling (Vieth & Bronowicka, 2017). However, not all of the biases are harmful: some facilitate learning and increase training effectiveness and they are not problematic if they do not interfere with ideal equity. In certain scenarios, intentionally introducing biases may become necessary. For example, if data from underserved populations, such as economically disadvantaged groups, is oversampled to address their prior lack of representation, this deliberate bias in data collection serves to meaningfully include a previously marginalised group. However, the lack of understanding/acknowledgement of the biases can lead to catastrophic outcomes (Pot et al., 2021). Wirth and Hipp (2000) identify that biases can be present at various junctures of an AI pipeline/life-cycle (see Fig. 1) from problem understanding to deployment and maintenance in an iterative manner. Most of the time, human involvement is heavy in data curation which involves data gathering, preprocessing and controlling data quality. It is reasonable to acknowledge that the biases present in the real world can leak into the data through this involvement lending substance to the saying "machines are only as fair as the people who make them" (Dastin, 2018; Lagioia et al., 2020; Sjoding et al., 2020; van Giffen et al., 2022).

Agriculture stands as a vital industry globally, particularly due to the escalating global population. Farms generate an extensive array of data daily, including information on temperature, soil quality, water resources, weather patterns, and more. This wealth of data underscores the significance of agriculture as a data-intensive sector on a global scale. These data are utilised in real-time by AI and ML models to derive insightful knowledge, including determining optimal seed planting times, selecting appropriate crops, and choosing hybrid seeds for higher yields. Therefore, the need for AI in agriculture is essential. Digital agriculture may suffer from harm and unex-

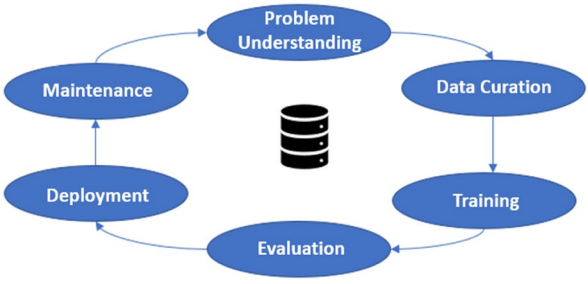

Fig. 1 Stages of an AI pipeline

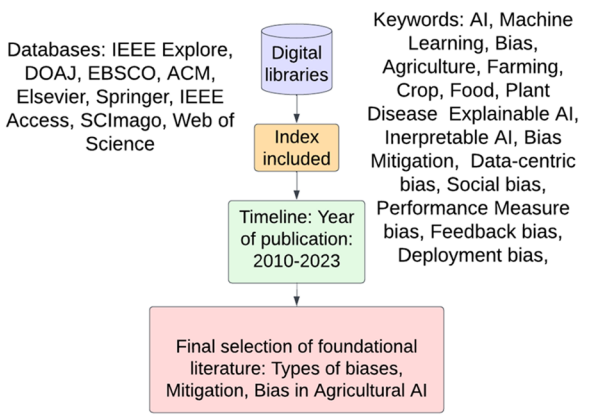

Fig. 2 Selection criteria used in the literature search

pected consequences of poor design and setup of intelligent systems. For instance, AI models predominantly trained on data from large, industrial farms may not be suitable for small-scale or diverse farming operations. This can lead to recommendations that are inadequate or harmful for smaller farms. Similarly, soil nutrient or disease detection models might overlook local soil variations or specific crop needs, resulting in ineffective interventions. Biased crop selection or market analysis that does not consider social, economic, or infrastructure barriers can further disadvantage farmers, limiting their access to resources and fair profits over time. Addressing biases in AI in agriculture is crucial to ensure fair and equitable access to technology, improve agricultural practices for all farmers, and promote sustainable and inclusive food production systems. Despite the significance of biases in AI systems specifically within agriculture, there exists a noticeable scarcity of review articles addressing this critical area. Therefore, in this manuscript we shed light on biases that are present in modern AI systems focusing on the field of agriculture. The process of our literature search is visualised in Fig. 2. The literature search was conducted encompassing publications from 2010 to 2023 using relevant keywords related to agricultural applications and biases in AI systems. Approximately 40 papers were selected for inclusion based on their relevance to the specific focus on agricultural applications and biases.

In Sect. 2, we first provide an overview of various types of biases commonly encountered in the broader field of AI. In Sect. 3, we delve into the strategies and techniques for mitigating biases within AI pipelines. Section 4 narrows the focus to examine specific instances of AI biases as they relate to agriculture. Finally, the manuscript concludes with a conclusion.

2 Background—Types of Biases in AI

We primarily categorise the biases into six sub-categories, some of which are not mutually exclusive: Data-centric bias, social bias, algorithmic bias, performance measure bias, feedback bias and deployment bias. Figure 3 provides a summary of these sub-categories.

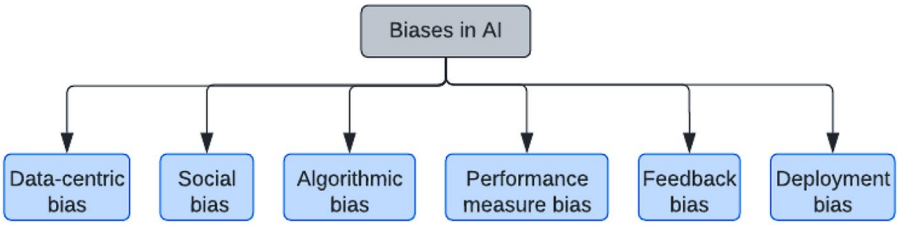

Fig. 3 Biases in AI

2.1 Data-Centric Bias

Algorithms used in machine learning (ML) typically learn using large datasets. This reliance on data makes modern AI/ML techniques highly susceptible to biases being transferred from the datasets (Emmanuel et al., 2021; O'donovan et al., 2015; Robinson et al., 2014; Syerov et al., 2020). We describe these data-centric biases namely selection bias (Du & Wu, 2021), attrition bias (Nunan et al., 2018), group attribution bias (Tarrant & North, 2004), and measurement bias (Hellström et al., 2020) in the following.

Selection bias occurs if samples for the training are chosen in a way that is not reflective of the real distribution of the data or the environment in which a model will be used (Du & Wu, 2021). For example, recruitment methods in a pharmaceutical clinical trial primarily attracted younger and healthier participants, resulting in selection bias. Consequently, the study's sample is skewed towards this demographic, potentially affecting the generalisability of the medication's effectiveness and safety to older or medically complex populations when it is prescribed in real-world settings. Selection bias can take different forms: *Sampling bias* occurs when data collection lacks proper randomization (Heckman, 1979). *Coverage bias* emerges the data is not selected in a representative fashion, while *non-response bias* results from gaps in data due to participation gaps in the data-collection process.

Attrition bias occurs if there are systematic changes in how individuals drop out from the study during data collection, as different rates of dropping out in the sampled groups may change these features/characteristics of groups (Nunan et al., 2018). For example, in a longitudinal study examining the efficacy of a smoking cessation program, an initially diverse cohort of smokers was enrolled. However, as time passed, a significant number of participants discontinued their involvement in the program, resulting in a smaller, distinct subgroup of successful completers. This introduces attrition bias because those who dropped out may have possessed different motivations, resources, or smoking behaviours compared to those who persevered.

Group attribution bias occurs when generalising individuals' behaviour to the entire group to which they belong (Tarrant & North, 2004). For example, in a tech firm, a hiring manager's observation of consistent high performance among employees from a specific university results in a group attribution bias. This bias influences the hiring manager to favor candidates from that university during recruitment, potentially neglecting highly qualified individuals from other institutions and diminishing the opportunity for diverse perspectives and talents within the organisation.

In-group attribution bias is to classify an individual primarily by considering the group it belongs or by considering the characteristics it shares with other individuals, while generalising the individuals of a group or classifying an individual who fits a characteristic as not having that characteristic because of the generalised nature of that group is known as *out-group attribution bias*.

Measurement bias occurs when data is labelled inconsistently or study variables are collected or measured inaccurately (Hellström et al., 2020). For example, measurement bias is evident in healthcare surveys when data is collected through surveys available only in English, potentially leading to inaccurate information from non-English-speaking patients and affecting the study's conclusions and healthcare recommendations. It can be further divided into recall and observer bias. *Recall bias* usually occurs during the data annotation phase of a project. This happens when similar data are inconsistently labelled, leading to low reliability. *Observer bias* happens when the methods or processes used to observe and record information for the study lead to a systematic deviation from the ground truth due to poor practices or a lack of training in using measuring tools or data sources.

2.2 Social Bias

Social bias in AI can be one of the hardest to identify and mitigate as it can originate from long-standing social prejudices that may be subtle and are reinforced by AI and massive datasets. Sengupta and Srivastava (2022) highlighted the presence of inherent biases within datasets and discussed the correlation between racial prejudice and ML algorithms. For instance, they explored how biased systems and human-AI interactions influence an individual's decision-making process. Specifically, when ML systems are trained on racially discriminatory datasets, they might inadvertently absorb historical biases, leading to a carry-over effect that impacts subsequent decisions. As a result, when people engage with AI systems with biases, they frequently deviate from their intended path of action.

Social bias can take different forms such as *pre-existing bias* (Basolo, 1995) which is a preconceived opinion that is not based on reason or actual experience, and *social inequity* (Howard & Borenstein, 2018) which may result in non-equitable outcomes for certain groups of the human population. It also encompasses other forms, such as *gender bias* and bias against *marginalised communities*. Gender bias, for example, can manifest in AI systems that perpetuate stereotypes, such as associating certain professions with one gender over another (Bolukbasi et al., 2016). Similarly, bias against marginalised communities can result in discriminatory outcomes, as seen in AI predictive algorithms that disproportionately focus on minority and economically disadvantaged communities, reinforcing negative perceptions (Haque et al., 2024).

2.3 Algorithmic Bias

Algorithmic bias (Stinson, 2022) occurs during the model development stage of the AI pipeline, which can be attributed to the lack of ability of the model to adapt to the presented data. It is a component of the model error, along with variance, which represents the sensitivity of the model to the differences in the presented data, and

noise, an outcome of the system's stochastic nature and hence is irreducible (Hastie et al., 2009). This model error is minimised in order to fit a model to the data. During the training, the model may start to learn the random noise in the training data rather than representative features to find the solution (Roelofs et al., 2019), resulting in significant variation in predictions. This may lead to *overfitting*. In the context of loan approval algorithms, overfitting can occur during model development. For instance, the AI model may start making decisions based on random noise in the training data, resulting in biased loan approvals. It may mistakenly learn that loan applicants who live on streets with specific names that are purely coincidental in the training data are more likely to repay loans. This bias can lead to unfair lending practices, emphasizing the importance of addressing overfitting to ensure equitable decision-making. Thus, it is important to assess performance on the training and validation sets to identify overfitting. Overfitting can be observed when the performance on the training data improves but the performance on the validation data declines (Paleyes et al., 2022). Sometimes, the model is not or partially learning helpful information to solve the problem which means that the model cannot learn from the signal presented in the data (Cunningham & Delany, 2021). This occurs mostly because the model is unfit for the task; either the model capacity is too low or the process is incorrectly configured (Paleyes et al., 2022). This may lead to algorithmic bias. The loss function, which is a formulation of the error, measures the difference between the output produced by the ML algorithm and the ground truth. Therefore, an `inappropriate loss function` may contribute to the model having a high bias, resulting in poor fit to the data (Sypherd et al., 2021).

Unsupervised learning techniques autonomously identify patterns and structures within data without explicit supervision. However, their deployment without careful consideration of fairness metrics can lead to unintended consequences, particularly in terms of algorithmic bias. For instance, discrepancies in average costs, such as reconstruction errors or distances to centroids, have been observed among different groups (Buet-Golfouse & Utyagulov, 2022). In reinforcement learning (RL), where agents learn optimal behaviours through trial and error by receiving rewards or penalties, biases can stem from the reward structure or the environment in which learning occurs. The bias can occur when algorithms, designed to maximize cumulative rewards, inadvertently favour certain actions or policies over others, potentially perpetuating inequalities across decision-making episodes (Jabbari et al., 2017). Ensuring fairness in RL requires balancing the objectives of optimizing long-term rewards with considerations for fairness to prevent unintentional discrimination.

AI methods often generate continuous prediction scores and necessitate a cut-off threshold to dichotomize the predicted outcomes. Determining the optimal threshold value typically involves post-hoc data-driven analysis, utilising previously analysed information on the same data but for different objectives (Chen et al., 2018). Such practices may lead to *post-hoc confirmation bias.* In some instances, algorithms are intentionally designed for profitability. For example, in the pursuit of profit, some online platforms intentionally design their rating systems to limit informativeness. This calculated approach can inadvertently favor sellers of lower quality, as the restricted information makes it challenging for buyers to discern product or service distinctions accurately. Conversely, sellers offering high-quality products may face

difficulties as their superior offerings struggle to stand out in the constrained information landscape. Such practices prioritise profit maximisation over the transparency of information (Belleflamme & Peitz, 2018).

2.4 Performance Measure Bias

Performance measure bias occurs because of an inappropriate performance matrix or non-representative testing population used to test/evaluate the model. For example, in standardized educational testing, a singular focus on maximizing the average test score as the sole performance measure can lead to performance measure bias, as this measure may not account for various factors, such as students' socioeconomic backgrounds or their access to resources (Bastedo et al., 2023). If the AI algorithm is trained on underrepresented data and tested with similar data, the bias will remain unrecognised due to the `inappropriateness of test data` (Arazo et al., 2020; Suresh & Guttag, 2021). Another reason for performance measure bias is the use of `unsuitable measuring methods` (Tarrant & North, 2004) where the performance of the outcome is evaluated using inappropriate measures. For example, accuracy is not suitable for an imbalanced dataset as a performance measure (Howard & Borenstein, 2018).

2.5 Feedback Bias

Feedback bias arises when the results of the AI model are used as new inputs to training data in such a way that allows a minor bias to be amplified via a feedback loop. It can be divided into algorithm feedback-loop bias and user feedback-loop bias. *Algorithm feedback-loop bias* occurs when the algorithm is used to predict the input itself as labels and if those predicted values are biased, then that will be reinforced to the system (Arazo et al., 2020; R. Wang et al., 2022) (e.g. semi-supervised learning). *User feedback-loop bias* occurs by the user's interaction with the system. For example, decisions of recommendation systems may alter user perceptions and preferences, which may then have an impact on the input received by the system (Mehrabi et al., 2021).

2.6 Deployment Bias

Deployment bias arises when the system is used or interpreted in inappropriate ways, even if none of the above-mentioned biases are present. Deployment bias can be further divided into application and environment bias. *Application bias* takes place when a model is utilised in a way that differs from how it is intended to be used (Paleyes et al., 2022), and *environment bias* occurs when a system was designed for a certain environment but is deployed in a different environment (Chouldechova, 2017). For example, an autonomous vehicle manufacturer initially creates self-driving cars optimised for city traffic, equipped with advanced sensors and algorithms. However, when these vehicles are later deployed in rural or off-road environments, like farms or remote areas, it leads to environment bias. The AI system, originally designed for

urban settings, may struggle to adapt to the different challenges of rural areas, potentially raising safety concerns and causing unexpected behaviors.

3 Background—Bias Mitigation in AI

Understanding biases in AI systems and mitigating them is imperative in circumstances where AI could eventually replace humans in complicated decision-making tasks (Claudy et al., 2022). The development and application of AI tools take place inside a socio-technical framework that takes into account human intellect, personal data, organisational structures, and conventions, as well as technological advancements. Therefore, frameworks such as General Data Protection Regulation (GDPR) are in place to provide regulatory support for the creation of AI applications along with ethical and legal principles to protect and advance both personal interests and social welfare (Lagioia et al., 2020). In recent years, bias mitigation has been an emerging field of research in AI. As we discussed above, there are several factors that affect the fairness of AI in different phases of an AI pipeline. In this manuscript, we consider bias mitigation techniques in different phases including problem understanding, data curation, modelling, evaluation, deployment and maintenance.

3.1 Problem Understanding

In a typical AI pipeline, the first phase includes understanding the problem and involves deciding the objective of the project and forming the development team. Since AI is perceived as capable of increased impartiality compared to humans by many, they may opt for AI in decision-making processes that favour impartiality, such as the allocation of scarce resources (Claudy et al., 2022). However, biases may originate from humans, and leak into the AI, therefore the tackling of biases should also begin at the human participation by considering the technical and social consequences of the project at an early stage (van Giffen et al., 2022). For instance, involving highly diverse communities in the design of AI can help address issues of social inequity, facilitating broader knowledge and data sharing among a more extensive audience (Norori et al., 2021). Thus involving individuals from different backgrounds in particular from underrepresented communities in the research is important as they can identify biases against their communities and contribute to improve the AI design (Tzovaras et al., 2019). Furthermore, biases due to socioeconomic factors or gender or sexual orientation can be explicitly gathered and included as metadata. Comprehensive metadata descriptions improve the discoverability of communities and concepts (Bolam et al., 2018). These descriptions include various attributes, keywords, or information that help categorise and organise content effectively. When metadata is detailed and comprehensive, it assists in improving the discoverability of specific communities and concepts within a dataset or platform.

3.2 Data Curation

Obtaining an unbiased sample is crucial to train AI models as even a minor bias in the data can be reinforced during the training process. Therefore, collecting adequately representative samples from the population is important. For example, Gjoka et al. (2010) proposed two approaches that are found to perform well in obtaining unbiased samples of Facebook users by crawling its social graph: Metropolis-Hasting random walk (MHRW) and a re-weighted random walk (RWRW). Moreover, it is undesirable to use sensitive information or personal characteristics such as gender, age, etc. which leads to a discriminative outcome in the learning process. However, it is hard to guarantee even if known protected attributes are omitted from the analysis. Datta et al. (2017) explained the example for the bank loan application where the non-protected attribute 'zip-code' was found to be a proxy for the protected attribute 'race'. But there are cases when race and zip-code are unrelated which may lead to proxy discrimination (Datta et al., 2017; Prince & Schwarcz, 2019). Therefore, the selection of attributes in an analysis task needs to consider such possible cases of proxy representations of protected attributes, and the analyst must strive to avoid them as much as possible. Missing values are another form that creates biases in data curation. In the literature, there are various strategies used for handling missing values (see Table 1), each with its own downstream consequences (Emmanuel et al., 2021; Rios et al., 2022).

Another important aspect to consider is the prevalence of social prejudice and partiality in data. Making the data openly available to the public will be useful to identify the issues due to social prejudice and partiality (Greshake Tzovaras & Tzovara, 2019). While open science (Mirowski, 2018), where data contributions are public, may contribute to solving social bias, it is challenging to ensure anonymity for public contributions to such open data. Furthermore, it is important to understand the

Table 1 Strategies of handling missing values

Strategies	Consequences
Delete rows with missing values	–Loss of information –Works poorly if the dataset has comparatively large amount of missing data
Replace missing values with mean/median	–Suitable only for numerical data –Data leakage
Replace missing values with a constant	–More appropriate for data where missingness doesn't convey crucial information
Replace missing values with most frequent data/category	–Suitable for categorical data –May reinforce imbalance of the more frequent categories
Use algorithms that support missing values (E.g. K-Nearest Neighbor (k-NN), Naive Bayes, Linear Regression, Random Forest) and deep learning libraries (E.g. DataWig)	–Only some of the AI algorithms handle missing data
Prediction of missing values	–Considered only as a substitute for the true value –Reliant on the accuracy of the prediction model
Leverage expert knowledge to fill the missing values	–May introduce subjectivity due to differing opinions –Need more information to handle complex datasets

dataset before diving into the decision-making process. citele2022survey provide an exploratory analysis using a bayesian network (Holmes & Jain, 2008) to identify the relationships among attributes. Furthermore, there are tools (Atemezing & Troncy, 2013) available to optimise and analyse ML datasets such as Facet Overview (James, 2017), Facet Dive (James, 2017), Facet Map (Smith et al., 2006), etc.

3.3 Model Training

Recently, the potential for discrimination in algorithmic outcomes is becoming more widely recognised (Stinson, 2022). Such concerns about potential biases in model training could be divided into three phases: preprocessing, inprocessing, and postprocessing.

3.3.1 Preprocessing Bias Mitigation Methods

In supervised learning, handling biases before training often involves carefully selecting and processing data. Data batches can be selected for model training in an adaptive manner to improve model fairness. For example, Roh et al. (2020) proposed a batch selection algorithm called FairBatch which implements optimisation and supports prominent fairness measures such as equal opportunity, equalised odds, and demographic parity. The advantage of this FairBatch is that it does not require any changes to data preprocessing or primary model training. In unsupervised learning, preprocessing bias mitigation can involve clustering or dimensionality reduction techniques that account for fairness. Chierichetti et al. (2017) addressed the challenge of fair clustering by transforming it into a classical clustering problem through a novel preprocessing approach. They introduced the concept of "fairlets", which are minimal sets designed to ensure fair representation while closely maintaining the clustering objective. This method underpins the development of fair clustering algorithms that achieve demographic parity across clusters, thereby ensuring equitable representation among diverse demographic groups within the data. Even in RL, preprocessing step is critical as it ensures the quality and fairness of the data used to train the RL models. The missing values can cause the agent to learn incorrect policies if not handled properly. Using population median imputation helps maintain the integrity of the data, ensuring that the DRL model has a consistent and representative dataset to learn (Yang et al., 2023).

Data augmentation is a technique to handle data imbalance and data scarcity. Existing data can be transformed to create fake data, or deep learning methods such as Generative Adversarial Networks (GAN) (Goodfellow et al., 2020), Neural Style Transfer (Jing et al., 2019), and Synthetic Minority Over-sampling Technique (SMOTE) (Chawla et al., 2002) can be used to create fake data. Geometric transformation, colour space transformations and feature space augmentation are some techniques used to tackle the scarcity of data (Shorten & Khoshgoftaar, 2019). These methods can introduce noise if not carefully managed, as they replicate the existing noise in the data. Therefore, noise or synthetic data with the same distribution may cause the model to become misleading.

Ensemble learning is a common strategy in debiasing which combines various individual models to improve better generalisation performance (Ganaie et al., 2022). For example, Rayana et al. (2016) proposed a sequential ensemble method for eliminating outliers from the original dataset to build a better data model from the dataset for the training.

3.3.2 Inprocessing Bias Mitigation Methods

Bias and variance must be balanced in a way that reduces overall error to build an effective model. Both bias and variance can decrease as the number of parameters grows in Neural Networks (Neal et al., 2018). We can optimise the performance of the model by using the validation set and perform an early stopping if the validation matrix decreases or validation loss increases over a few steps. One approach to improve the model performance is to use dropout technique which can retain the complexity of the model (Zhang et al., 2022). Srivastava et al. (2014) show that dropout improves the performance of neural networks on supervised learning tasks in vision, speech recognition, document classification and computational biology. While dropout helps the model for better performance, a high dropout rate might lead to an excessive reduction in model capacity and may lead to underfitting (Bulò et al., 2016). There is a proportional relationship between the model features and the arrival to the overfitting of the susceptible model as not all the characteristics are always important. For example, some features could only cause the data to be noisier. Thus dimensionality reduction could reduce the overfitting while maintaining the performance of the model as good as or better than the original (Salam et al., 2021).

Sometimes, the model cannot learn from the signal presented in the data which leads to underfitting. Domain experts can suggest additional input features to increase the complexity and performance (van Giffen et al., 2022). Jabbar and Khan (2015) conducted a comparative study on two methods: Early stopping and penalty methods to avoid overfitting and underfitting in supervised machine learning and discovered that the early stopping technique, which can prevent overfitting and underfitting while being considerate of the validation time, is better than the penalty method. Self-training algorithms use both labeled and unlabeled samples for the training and they often choose labels for the unlabelled samples that score higher than a threshold (Gao et al., 2020). It is recommended to choose the labels based on the degree of confidence rather than thresholding or maximum probability selection in the prediction; this will increase the accuracy of the pseudo labels' prediction and improve the performance of the system (Mayuravaani & Manivannan, 2021).

An obstacle to the fairness of the system is the absence of human interpretability and manual configuration. To overcome this issue, Halgamuge (2021) proposed Fair, Accessible, Interpretable, Reproducible (FAIR) AI where new techniques enable the automatic creation of interpretable neural network models with the less assistance from AI specialists. Automated machine learning (AutoML) has become a popular topic to make machine learning techniques easier to implement and eliminate the necessity for skilled human specialists (Yao et al., 2018). Self-configuration methods could automatically set the model parameters in accordance with the changes in the dataset. A fast training algorithm could be developed with the ability to self-configure

and set parameters automatically and build a fully trained deep neural network starting with nothing more than data (Wong, 2018). One notable example is Auto-sklearn, which encompasses a broad array of classification algorithms and preprocessing methods. It automates the search for optimal models and the corresponding hyperparameters, thus streamlining the model development process without requiring user interference (Feurer et al., 2015). Similarly, GAMA (Genetic Automated Machine Learning Assistant) is another prominent AutoML framework that employs genetic programming to evolve and optimize machine learning pipelines tailored to specific datasets (Gijsbers & Vanschoren, 2019). By dynamically generating and evaluating different pipeline configurations, GAMA efficiently explores the search space of possible solutions, leading to high-performing models with minimal human oversight. In addition to these AutoML tools, there is a growing need to address fairness and bias in machine learning models. Tools like AI Fairness 360 (Bellamy et al., 2019) and Fairkit-learn (Johnson & Brun, 2022) are designed to detect and mitigate biases in AI systems. AI Fairness 360 offers a comprehensive suite of metrics and algorithms for assessing and enhancing the fairness of machine learning models.

In machine learning, handling class imbalance is crucial for training models that perform well across all classes. One approach to address this issue is by adjusting the sample weights to give more importance to the underrepresented classes. For example, PyTorch provides a weighted random sampler which facilitates this by enabling the model to sample data points with different probabilities, thereby giving higher weight to minority classes during the training process (Paszke et al., 2017). This method helps in mitigating the bias towards the majority class and improves the model's ability to correctly predict instances of the minority class.

In unsupervised learning, addressing biases during the learning process is essential to ensure fair and equitable outcomes. Several innovative approaches have been developed for this purpose. For instance, Fair Principal Component Analysis (Fair-PCA) (Samadi et al., 2018) ensures fairness by balancing the reconstruction errors across different demographic groups, preventing one group from being disproportionately represented in the reduced-dimensional space. In RL, methods from causal inference and constrained optimization are used to develop algorithms that optimize standard objectives, like minimizing regret or maximizing rewards, while also enforcing fairness constraints. These constraints ensure that the outcomes are equitable across different demographic groups, leading to policies that are both effective and fair (Nabi et al., 2019).

Algorithms are generally optimized based on average loss due to the computational complexity involved. However, this approach can introduce fairness concerns, particularly when dealing with imbalanced datasets or protected attributes. Optimizing for average loss alone can lead to the algorithm favouring overrepresented groups, thereby marginalising minority classes. This highlights the importance of incorporating fairness considerations into the loss function. For example, Agarwal et al. (2019) addressed this issue by proposing general schemes for fair regression that incorporate fairness directly into the optimization process. They introduced two notions of fairness: statistical parity and bounded group loss. Statistical parity ensures that predictions are statistically independent of protected attributes. Bounded group loss, on the other hand, requires that the loss function is defined separately for each protected

group and is restricted to remain below a certain threshold. It can guarantee that the algorithm misclassified no group.

Ensemble methods, such as voting, combine predictions from multiple models to improve overall performance. Voting typically involves selecting the class with the highest frequency or weight among the predictions. However, this approach can inadvertently perpetuate biases present in the individual models, leading to overfitting (Barber, 2012). To mitigate these biases, Liu and Cocea (2017) proposed a probabilistic approach to voting within the framework of granular computing. This method improves the accuracy and fairness of the final classification by considering the probabilities associated with each class rather than just their frequencies or weights.

3.3.3 Postprocessing Bias Mitigation Methods

It has been found that even after the protected attributes have been eliminated, non-sensitive features may still function as a proxy for them and produce biased outcomes after training (Kusner et al., 2017). Therefore, a postprocessing debiasing strategy helps in identification of such deficiencies. Another benefit of such a strategy is that any black-box model may be used with the postprocessing debiasing procedure. However, we have to consider the trade-off between accuracy and the degree of resultant fairness while postprocessing (Kamishima et al., 2012). For example, Lohia (2021) suggested a priority-based post-processing bias reduction with the idea being that similar individuals should receive similar results regardless of socio-economic factors. Many effective AI algorithms are called a "black box" approach, making it challenging or impossible to comprehend how the results were reached (Roscher et al., 2020). One efficient method for visualising bias representation is using explainable and interpretable models, where the link between input features and output is transparent and comprehensible (van Giffen et al., 2022). There are some interpretable and explainable methods in the field of computer vision such as class activation map (CAP) (Sun et al., 2020), saliency map (Mundhenk et al., 2019), GRAD-CAM (Joo & Kim, 2019), etc. Gorski et al. (2020) presented an approach to use the image processing technique, Grad-CAM to show the explainability for legal texts. Lundberg and Lee (2017) proposed SHapley Additive exPlanations (SHAP) which is a standardised methodology for interpreting predictions. Another model-independent technique is Local Interpretable Model-agnostic Explanations (LIME) (Ribeiro et al., 2016), which works by approximately simulating the model's behaviour close to a given prediction. In agriculture, explainable AI (XAI) techniques, such as those used for plant disease classification (Ranasinghe et al., 2022), help farmers understand AI predictions. This transparency is essential for ensuring confidence in AI models, making them more trustworthy and reliable for agricultural applications.

As the field of XAI continues to evolve, integrating argumentation with ML represents a promising direction to address the demand for explainable, accountable, and reliable AI systems (Vassiliades et al., 2021). It provides a structured way to elucidate the reasoning behind model decisions, making them more comprehensible and trustworthy. Argumentation frameworks can explain how different pieces of data support or attack each other, thereby clarifying the model's decision-making process. Moreover, incorporating Argumentation-Based Dialogues (ABD) into AI explana-

tions offers a natural and intuitive method for presenting reasoning processes. ABD structures the conversation by following strict protocols, determining factors such as turn-taking, the use of specific knowledge, and the conclusion of the dialogue. While effective, these protocols may need to be adapted when dealing with domain-specific information to ensure they remain applicable and effective (Vassiliades et al., 2021).

The challenge of XAI is if the existing AI method lacks in explainability then it may not create trustworthiness. Saranya and Subhashini (2023) discussed the properties and challenges of XAI such as developing models that are simpler to explain, constructing explanation interfaces, and understanding the psychological circumstances required for a convincing explanation. Crabbé and van der Schaar (2022) introduced label-free extensions of post-hoc explanation methods for unsupervised learning. Moreover, recent advancements have extended these efforts to the domain of RL through explainable reinforcement learning (XRL). XRL specifically addresses the unique challenges of understanding the decision-making processes within RL systems (Puiutta & Veith, 2020).

The objective of AI interpretability is to explain a system's underlying structure in a way that is clear to people which depends on the cognition, expertise, and biases of the user. An interpretable system should generate descriptions that are straightforward enough for a person to comprehend using a language that is relevant to the user (Gilpin et al., 2018). For example, Z.J. Wang et al. (2021) proposed GAM Changer, an open-source interactive system to enable domain experts and data scientists to easily and responsibly modify their Generalized Additive Models (GAMs). Programmatically Interpretable Reinforcement Learning (PIRL) offers a framework for enhancing model transparency by using a high-level, human-readable programming language to represent policies. This approach simplifies the policy structure through predefined grammar, making the decision-making process more interpretable and understandable. By enforcing these constraints, PIRL ensures that reinforcement learning policies remain effective while being accessible and clear to humans (Verma et al., 2018).

Finding the bias and debiasing AI algorithm can be improved by making the source code openly available, which enables developers to extend, reuse, and evaluate existing code (Norori et al., 2021). If a system's predictions have the possibility of causing individual disparity, internal management should be in charge of keeping an eye on the results and ensuring their validity. The data scientist must also take into account the distortion that any model-based intervention may have on the data-generating or sampling distribution while creating the classification system's feedback loop (d'Alessandro et al., 2017).

3.4 Performance Measure

Another crucial question is how to measure the performance of AI systems given a particular scenario or dataset. The numerous bias quantification metrics, often known as fairness metrics or performance matrices, are reviewed in this section. Accuracy is a good measure for the performance of the classification algorithms on a balanced dataset. At the same time, it is not appropriate to use accuracy as a performance measure for an imbalanced dataset (Weng & Poon, 2008). Instead of accuracy, metrics such as the F1 score, receiver operating characteristic (ROC) curve, or

precision-recall (PR) curve may better evaluate model performance on imbalanced data (Tamimi & Juweid, 2017). Each of these metrics offers distinct advantages and considerations that are crucial for assessing model fairness and bias. For instance, while the F1 score balances precision and recall, providing a comprehensive view of classification performance, it may not capture the nuances of true positive and false positive rates across varying decision thresholds. The ROC curve, on the other hand, visualises the trade-off between sensitivity and specificity, offering insights into how well the model discriminates between classes. However, it assumes equal importance for false positive and false negative errors, which may not align with fairness considerations in certain applications that could potentially disadvantage certain groups if they are more prone to one type of error. Similarly, the PR curve focuses on precision and recall, highlighting the model's ability to correctly classify positive instances while managing false positives. Nevertheless, it too has limitations, particularly in contexts where class distributions are highly imbalanced, as it may exaggerate the performance of minority classes (Saito & Rehmsmeier, 2015). Understanding the implications of each metric can help mitigate biases and ensure equitable outcomes in AI systems.

For regression tasks, common performance metrics include Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-squared (R^2) score. MSE and MAE quantify the average magnitude of errors in predictions, while R^2 measures the proportion of the variance in the dependent variable that is predictable from the independent variables (Plevris et al., 2022). Clustering tasks, on the other hand, rely on metrics such as silhouette score and Davies-Bouldin index to evaluate the quality of cluster assignments. The silhouette score measures how similar each data point is to its own cluster compared to other clusters, indicating the cohesion and separation between clusters. The Davies-Bouldin index evaluates the average similarity between each cluster and its most similar cluster, providing a measure of clustering quality (Ros et al., 2023).

Model evaluation can be generally classified into internal and external testing. Internal testing occurs within the training process, employing techniques like train-validation split and cross-validation. External testing, on the other hand, is conducted on an unseen external dataset, often with different populations. Thus, external testing can help to identify real-world issues (Faghani et al., 2022). A confusion matrix is a performance matrix that can be used to evaluate bias in different classes for balanced and imbalanced datasets. The confusion matrix comprises four fundamental components: true positives, false positives, true negatives, and false negatives, which represent the intersections between predicted and actual classifications. These metrics are essential in assessing a model's performance. For instance, they help identify whether the model tends to underpredict (false negatives) or overpredict (false positives) certain classes. A high false negative rate may indicate the model's conservative nature, frequently missing instances of the positive class. Conversely, a high false positive rate might suggest that the model is overly inclusive, often misclassifying negative instances as positive. Various matrices, such as accuracy, intersection over union (IoU), Dice similarity coefficient (DSC), and the Hausdorff distance (HD), exist to assess the performance of segmentation models. The choice of metric

becomes pivotal for a fair evaluation, especially when dealing with outliers or imbalanced datasets (Taha & Hanbury, 2015).

3.5 Deployment

Deployment bias may be avoided by highlighting the technical and social implications of the ML model by evaluating the relevant social environment and moral standards (Martin, 2019). In addition, it's important to express the limitations of the systems in various usage situations and create monitoring plans that account for changes in the algorithm when the context evolves (Buolamwini & Gebru, 2018). Moreover, to enhance deployment accuracy, researchers should gather external testing data from sites resembling their intended deployment environments (Faghani et al., 2022). Regarding the model deployment process, it typically involves three main stages: integration, monitoring, and updating. The integration step specifically encompasses two primary activities: establishing the infrastructure to support the model and implementing the model itself in a usable and sustainable format. While this may imply a clear division of responsibilities, with researchers developing the model and developers creating its operational infrastructure, the domains of model performance and infrastructure often intersect. Thus, it's recommended to involve researchers throughout the entire development process to ensure a cohesive approach (Paleyes et al., 2022). Moreover, monitoring and updating are essential for ensuring the sustained accuracy and reliability of deployed models. After the model is initially deployed, it is important to update it to align with changes in data and the environment. Monitoring involves using tools to track system health, resource usage, and response times, which helps prevent operational issues from affecting model performance.

3.6 Maintenance

An AI system needs to adapt to changes in the environment over time (Budiman et al., 2016). Several factors have an impact on the model's decision-making process. For instance, model drift is a significant concern, representing the diminishing effectiveness of the model over time. This degradation arises due to various factors, notably changes in the data and input-output distribution, referred to as data drift (Quinonero-Candela et al., 2008) and shifts in the decision criteria used by the model, known as concept drift (Jameel et al., 2020). It is necessary to be able to update the model after the first deployment is complete, in order to guarantee that it perfectly reflects the most recent trends in data and the environment (Paleyes et al., 2022). There are several methods for adjusting models to updated data, such as scheduled regular retraining and continually learning from data (Diethe et al., 2019). It is possible to create a feedback loop in which input to the model is adjusted to influence the behaviour of the model while ensuring the model remains up-to-date (Paleyes et al., 2022). In addition, transfer learning offers a strategy, where a model developed for one task is adapted to a new but related task. This can be particularly effective when the new task has limited labelled data, leveraging the knowledge gained from the initial task to improve performance on the new one (Weiss et al., 2016).

4 Applications in Agriculture

Technology has transformed farming, and it has had a wide range of effects on the agriculture sector (see Fig. 4) (Dara et al., 2022; Shamshiri et al., 2018). The incorporation of AI into agriculture is an evolution built on years of research and technological progress. In its initial stages, AI in agriculture centered around rule-based and expert systems aimed at simulating human decision-making (Sriram & Philip, 2016). Yet, contemporary strides in machine learning, deep learning, and extensive data analysis have ushered AI into an advanced era. This empowers AI to unveil concealed patterns and insights from vast datasets (Bhagat et al., 2022; Mourtzinis et al., 2021). Figure 5 depicts the bibliometric visualisation of the annual distribution of publications resulting from a search conducted in the Web of Science (WoS) all databases and the core collection database on research related to AI applications in agriculture and food industry from 2010 to 2023. It offers a significant insight into the rising trend observed in scholarly publications related to AI in the agri-food sector over the examined period. Although the focus is on data from the WoS collection, this trend may broadly mirror the overall growth in research interest and advancements within the field of AI applied to agriculture and food, showcasing the burgeoning attention and progress in this domain.

In many countries around the world, agriculture is the important industry, with an increasing global population and lots of data points on temperature, soil, water, weather, etc. are generated daily by farms. These data are utilised in real-time by AI and ML models to derive insightful knowledge, including determining optimal seed planting times, selecting appropriate crops, and choosing hybrid seeds for higher yields.

One significant challenge in the agricultural sector is the labour shortage, driven by factors such as the aging population of farmers and the migration of younger workers to urban areas. This shortage impacts the efficiency and timing of critical agricultural tasks. AI and robotics offer promising solutions to mitigate this labour

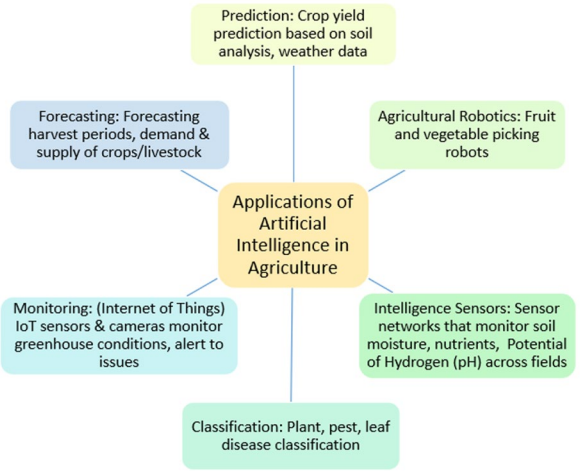

Fig. 4 AI applications in agriculture. Adapted and modified from (Dara et al., 2022)

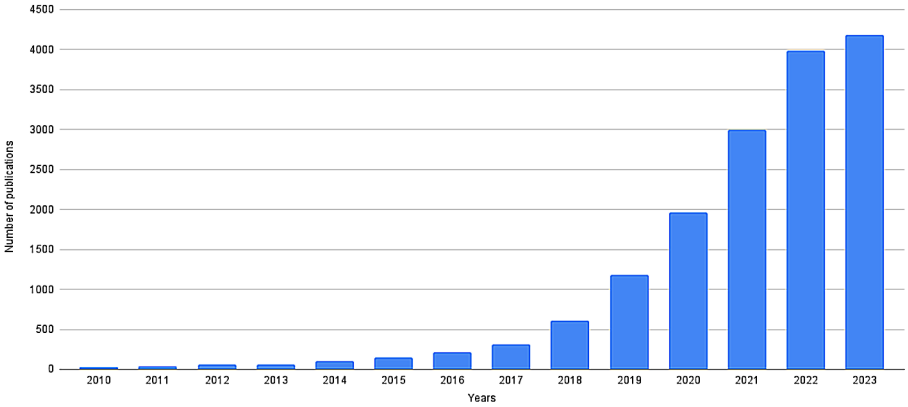

Fig. 5 Annual trends in publications on research related to AI applications in agriculture

crisis. AI-powered grippers and harvesting machines can automate the processes of picking and packing crops, as well as transferring fruits from the field to packing stations. By leveraging AI-driven robots, the agricultural sector can significantly improve productivity, sustainability, and resilience against workforce challenges. Therefore, the need for AI in agriculture is essential. Digital agriculture may suffer from harm and unexpected consequences of poor design and setup of intelligent systems. Addressing biases in AI in agriculture is crucial to ensure fair and equitable access to technology, improve agricultural practices for all farmers, and promote sustainable and inclusive food production systems. Additionally, forecasting agricultural commodity prices poses a formidable challenge due to its reliance on multifaceted factors such as weather patterns, seasonal variations, global economic conditions, and country-specific situations (Gu et al., 2022).

In some instances, governmental decisions significantly impact agricultural practices. For example, recent actions taken by the Sri Lankan government aimed at transitioning toward organic and environmentally friendly farming practices through the prohibition of certain agrochemical imports sparked widespread debates and protests. However, by November 30, 2021, in response to the backlash, the government reversed the ban, allowing private companies to resume importing these products (WorldBank, 2022). Despite the lifting of the agrochemical ban, subsequent challenges, including foreign exchange shortages impacting fertilizer availability, persist in affecting agricultural outputs across various sectors in 2022. Moreover, inadequate linkages between private sector needs for innovation and research and development institutions contribute to limited distribution and adoption of modern farming technologies. These challenges are particularly felt in regions with high poverty rates, such as the Northern and Eastern Provinces, which are centers of estate sector (plantation-based agriculture) concentration.

Figure 6 presents six primary categories into which the biases associated with AI systems applied in agriculture can be classified and Table 2 represents the biases in agricultural domain with mitigation strategies.

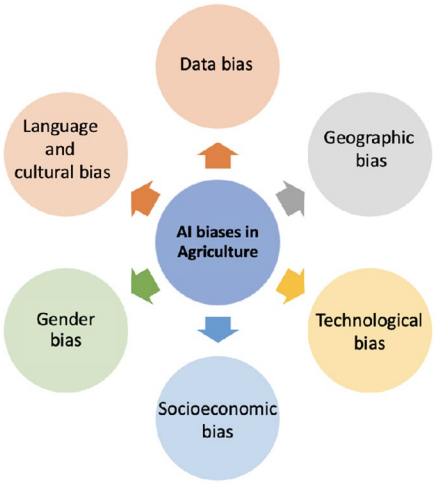

Fig. 6 Type of biases in agriculture AI

4.1 Biases in Agriculture AI

4.1.1 Data Bias

There are a variety of plant species cultivated on the farm. AI systems may favour the majority type if the developer of the AI system uses imbalanced data among the species which will lead to a biased AI system. For example, if a fruit-picking AI robot (Zhou et al., 2022) is trained with a particular kind of fruit in huge quantities, it may favour that variety. It will outperform the large dataset class when used in multi-variety gardens. Also, the fruits may differ in variety and the robot/ system may pluck the unripe fruit.

Most of the time, we may not always receive a properly balanced real-world dataset for training. Thus, Sambasivam and Opiyo (2021) used different techniques including class weight, focal loss, SMOTE and different image dimensions on an imbalanced dataset for Cassava disease detection and classification and achieved promising results. There are AI technologies with computer vision that are used for identifying pests or diseases in the fields (Johannes et al., 2017). In such instances, the model may perform well in the training dataset, but when applied to the field, performance sometimes degrades noticeably. There are various reasons for this low performance such as the actual situation could be different from the training data, various pests might naturally invade the field instead of what was anticipated, and diseases might vary due to climatic changes. Some diseases provide symptoms with a variety of features, and many diseases might cause symptoms that are extremely similar and present at the same time (Barbedo, 2016), etc. The key challenges in classifying fruits/vegetables and detecting fruit diseases from images discussed by Dubey and Jalal (2015) include variations in produce appearance depending on ripeness, complex backgrounds, limited training data, similarity between healthy and diseased tissues, and dependence on accurate segmentation algorithms to handle illumination changes, extract meaningful features, and precisely detect defects. Another issue is that one leaf may have more than one disease at the same location. Since we

Table 2 Agricultural biases and mitigation strategies

Author(s)	Types of biases	Mitigation	Application
Gardezi et al. (2023)	Data bias	Use large, diverse and representative datasets that cover all relevant groups/classes to reduce unintended biases from creeping in during training	Trustworthy AI system
	Application bias	Consider inclusion of diverse contexts and groups during design and development to ensure models work for all relevant stakeholders and situations	
McVey et al. (2023)	Aggregation bias	Using unsupervised machine learning approaches like hierarchical clustering that do not rely on predefined assumptions and can identify patterns at the individual level	Livestock farming data streams
Zhou et al. (2022)	Data bias	Delete rows with missing values	AI robots
Lee et al. (2022)	Data bias	Data was collected from both sides of each tree and the results were averaged and taking measurements from multiple angles helps reduce occlusion bias, Data collection was done at a consistent time of day and amount of sunlight to ensure proper detection under uniform lighting conditions to avoid natural lighting variation	Fruit flower cluster detection
Pádua et al. (2022)	Seasonal/ temporal bias	The study acquires remote sensing data and field measurements at three different time periods (August, September, October) to help account for changes in leaf area index over the growing season	Estimating leaf area index (LAI) in a chestnut grove
Rauf et al. (2022)	Human error and bias	Using remote sensing techniques that are more objective and standardized to mitigate the error in manually collecting ground truth data	Rice variety identification
Restrepo-Arias et al. (2022)	Class imbalance	Use data augmentation techniques like flipping and rotating to under-represented classes	Plant disease detection
Thomas et al. (2022)	Selection bias	Characterizing the dataset for limitations and making sure it represents the target population. Appropriately recruiting populations and up/downsampling data to address imbalances	Neutrinos
	Data imbalance	Balancing classes/groups in the data through upsampling underrepresented populations or downsampling overrepresented ones. Appropriate balancing methods can address biases that arise from imbalanced data	
	Measurement error	Taking multiple measurements of the same variable, collecting data with precision under controlled conditions, and being transparent about levels of measurement error, XAI models can help understand how measurement error propagates	
Sambasivam and Opiyo (2021)	Data bias	Class weight, focal loss, SMOTE and different image dimensions on an imbalanced dataset	Disease detection and classification
Camaréna (2021)	Socioeconomic bias	Take a collaborative and participatory approach to AI development through code design and involve diverse stakeholders from the beginning, Framing AI development within an ethics framework like the Australian Government's AI Ethics Principles, which includes fairness, non-discrimination, and accountability	Support farmers

Table 2 (continued)

Author(s)	Types of biases	Mitigation	Application
Zossou et al. (2020)	Gender bias	Implement gender-sensitive rural extension and learning approaches that have been tested and proven effective, Leverage modern technology such as farmer-to-farmer videos, mobile phones, radios, televisions, and social networks to disseminate information and facilitate learning and encouraging community members to access information directly through these tools, reducing their dependence on traditional group leaders	Rice farming practices
Xiong et al. (2020)	Data bias	Use blockchain technology, decentralizing data management among network participants rather than central administrators enhances data security and reduces the risk of manipulation or loss, Ensure transparency by allowing all participants to independently verify recorded transactions and data, eliminating the risk of biased data collection or use by a single central authority	Food supply chain
Moore and Rutherfurd (2020)	Social desirability bias	Choose data collection methods that reduce recall period, like daily journals instead of long-term recall questions	Environmental behaviours
Shikuku (2019)	Gender bias	Technology dissemination can be enhanced when the disseminating farmer is female, regardless of whether the contact farmers are male or female	Agricultural technologies
	Time-variant selection bias	Use propensity score matching, where treated (linked) farmers are matched with control (non-linked) farmers with similar baseline characteristics, helping to mitigate time-variant selection bias	

can only assign individual labels to the diseases if they are physically separated in the same leaves, it is quite difficult to distinguish between the symptoms if they overlap in the same region (Mondello et al., 2018).

Soil texture analysis involves classifying soil based on particle types like clay, sand, and silt, which significantly influence soil quality and suitability for agriculture. The challenge lies in collecting a diverse and representative dataset for training machine learning algorithms to accurately predict soil texture using techniques such as near-infrared spectroscopy. However, due to the variability of soil composition across regions and even within local environments, achieving a comprehensive training dataset becomes impractical. This limitation introduces bias into the data used for machine learning, potentially leading to inaccuracies in predicting soil texture under different conditions (Bronson et al., 2021).

Resource bias in AI models for crop recommendations arises when certain crops or agricultural practices receive disproportionate attention due to the abundance and availability of data. This bias can lead to underrepresented or overlooked recommendations for other equally significant crops. Often, AI models are trained on data sourced from large-scale, resource-rich farms, which may not provide suitable or feasible recommendations for small-scale or resource-constrained farmers. Additionally, disparities in access to technology, data availability, infrastructure, financial resources, educational opportunities, technical support, information dissemination, and regulatory frameworks all significantly influence the agricultural sector. These

factors collectively contribute to the prevalence of resource bias in agriculture, affecting the relevance and applicability of AI-driven recommendations across different farming contexts.

An ***aggregation bias*** (McVey et al., 2023) can emerge due to the obscuring effects resulting from data aggregation or averaging techniques applied during model development. Specifically, when indicators representing the outcome of interest are observed at an individual level, but data collection and subsequent modeling occur at an aggregated level, significant variations among individual units may become masked. This discrepancy between the actual phenomenon's level and the level at which the data is analysed can introduce systematic errors into model assumptions and predictions if left unaddressed. For instance, daily weather metrics such as temperature and rainfall were averaged to create a singular value for each zone per day, rather than retaining raw hourly or sub-daily observations. This method disregards potential variations within a day, such as differing accumulation of heat units among crops in various parts of a zone due to variations in rainfall patterns. Consequently, these nuanced intra-zone differences are overlooked. As a result, predictive models fail to fully capture genuine relationships that exist on a more localised scale than the predefined zones, leading to aggregation bias. This bias obscures the analysis by smoothing out within-zone heterogeneity that significantly influences crop yields.

Time-variant selection bias (Shikuku, 2019) arises when subjects in a study lack comparability over time due to inherent characteristics that influenced their assignment to different groups. For example, farmers who established information exchange links with others may differ in terms of motivation and learning ability, potentially introducing bias into the estimates if not considered.

4.1.2 Geographic Bias

Agricultural data is often collected from specific regions, which may not represent the diversity of conditions and challenges faced by farmers in different locations. Models trained on geographically biased data, which often reflect the interests and conditions of major producing regions rather than broader needs, might restrict the applicability of findings and agricultural practices in other areas with different needs and conditions (Byerlee et al., 2014). For instance, a model developed using data from a specific region of paddy fields might heavily rely on factors such as soil type, localised weather patterns (e.g., humidity levels), specific crop varieties, and the use of particular fertilizers. Kamilaris and Prenafeta-Boldú (2018) highlight the difficulties in applying deep learning to agriculture, primarily due to the fact that the majority of data collected for developing agricultural deep learning models originate from particular regions, specific crop varieties, and limited environmental conditions. As a result, these biased datasets can lead to models that are optimised and tuned only for those specific conditions, potentially limiting their applicability and accuracy when deployed in different agricultural contexts.

4.1.3 Technological Bias

AI applications in agriculture often rely on specific types of sensors or technologies. If these technologies are biased towards certain types of data or conditions, the resulting AI models may not be robust enough for diverse situations. Automated technology, such as drones or robots, might do harm to plants while farming. For example, while plucking fruit, the robot arms might inadvertently damage nearby fruit, affecting other plants and resulting in wastage. To overcome this, research is now being done to genetically alter plant structures to accommodate robots (Sparrow & Howard, 2021). Sensor nodes face a significant risk of damage and failure when exposed to harsh outdoor conditions such as extreme temperatures, humidity, or physical damage from animals or farm equipment (Okengwu et al., 2023). The energy consumption associated with data collection, processing, and transmission further contributes to their limited lifespan. As a consequence, these situations introduce bias in data collection and processes (Rehman et al., 2014). For instance, sensor failures or malfunctions in certain conditions may lead to incomplete or inaccurate data collection, affecting the overall quality and reliability of the data obtained.

The potential impact on each country will hinge on a variety of factors, particularly their readiness for the emergence of AI. This readiness encompasses technological prerequisites like internet accessibility and robust, cloud-connected computing capabilities. Nonetheless, similar to other technological advancements, the adoption of AI could exacerbate inequalities and marginalize disadvantaged populations. For instance, the widespread automation of farms may disrupt established ways of life and ecosystems, with profits accruing to tech corporations. The unconsidered advancement of AI on the continent may amplify disparities, leading to increased inequality, economic disruption, and social tensions. This, in turn, could have serious consequences, particularly for the underrepresented and technologically disadvantaged segments of the population. Furthermore, it is important to maintain the trust among farmers while considering data gathering. The ability of AI to analyse both individual and group behaviour is remarkable. Therefore, maintaining their privacy is essential. Recently, the term “data trusts” reached common usage as a governance framework for the collection and exchange of personal data (Brewer et al., 2021).

4.1.4 Socioeconomic Bias

AI models trained on data may reflect socioeconomic disparities in agriculture. For instance, certain technologies or practices might be recommended based on economic viability, potentially neglecting the needs of subsistence farmers. Economic barriers often determine which farmers can adopt new technologies (Okengwu et al., 2023). Precision farming technologies, such as sensor-based irrigation systems and satellite-guided crop monitoring, are often recommended for their potential to optimize resource use and maximize yields. However, these advanced technologies require significant upfront investment and entail ongoing maintenance costs. Farmers’ access to agricultural information is a crucial factor that can be influenced by various socio-economic characteristics. The study (Rehman et al., 2013) aimed to identify various sources of agricultural information accessible to farmers and exam-

ine how socioeconomic characteristics influence this access. The researchers found a significant association between farmers' educational levels and the size of their land holdings with their ability to access agricultural information. Specifically, higher educational attainment and larger land holdings were linked to better access to crucial agricultural data and resources.

Innovations such as hybrid seeds and genetically modified organisms (GMOs) have exacerbated power inequities between small and large farmers, as well as between farmers and powerful agribusinesses. This refers to how economic power is concentrated among larger entities, disadvantaging smaller farmers who may struggle to afford expensive technologies or comply with legal requirements (Bronson et al., 2021). The exclusive use of AI on large, financially well-off, and homogenous farms creates a significant barrier to AI adoption by smaller farms, impoverished farmers, and rural communities. The absence of well-defined norms and limitations for domain-specific human information further complicates the situation. Therefore, in such scenarios, it is crucial to ensure that AI is trained on diverse sources to ensure its representativeness and relevance (Lin et al., 2017).

4.1.5 Gender Bias

In some regions, agricultural practices may be influenced by traditional gender roles, leading to biased data collection and AI models that may not be inclusive or equitable for all genders involved in farming. For instance, men are often more involved in growing cash crops for sale, while women typically focus on cultivating crops for personal consumption and household use. This division of labour can lead to unequal access to resources, support, and technology for women farmers. Gender discrimination significantly affects agricultural practices, as lenders, suppliers, and other stakeholders in the agricultural sector often exhibit biases based on gender. One of the difficulties farmers confront is credit. There is gender bias in lending because fewer women apply for loans and the loan application procedure has insufficient documentation (Escalante et al., 2009).

4.1.6 Language and Cultural Bias

If AI models are developed using data and resources primarily in one language or culture, they may not be as effective or accessible for farmers who speak other languages or come from different cultural backgrounds. The adoption of digital technologies often requires a certain level of digital literacy and technical knowledge, which is not evenly distributed among farmers. Farmers with higher education levels are more likely to adopt and benefit from AI technologies, while those with lower educational attainment struggle to integrate these innovations into their farming practices.

Since inaccurate predictions and the use of unsustainable techniques may have negative effects on the farmer and the health of the stock, transparency could be advantageous for end users in order to avoid the negative effects of unwarranted or excessive data exposure. Errors and unintentional data leaks might be avoided by making the disclosure process more clear and emphasising the data that is made

available to others. However, increased transparency can potentially overwhelm the operators with data (Linsner et al., 2022).

4.2 Recommendation and Suggestion for Bias Mitigation

Addressing AI-based biases in agriculture requires a multifaceted approach that involves various stakeholders, including researchers, developers, policymakers, and farmers. Data collected in digital agriculture can be heterogeneous, noisy, incomplete and imbalanced. Preprocessing techniques like data cleansing, normalisation and transformation are used to handle such issues and prepare the data for effective analysis (Chergui & Kechadi, 2022). For example, ideal position of the image during capturing is a crucial task for the ML processing. If a device/robot captures the image for a ML model by recording the middle position of the plant leaves then the feature extraction, segmentation, edge detection, and removal of background are expected to be more accurate. The presence of dust and raindrops on the plants will also have an impact on the performance. Therefore, preprocessing methods for noise removal have to be applied to images that will help in resolving this issue. Another suggestion is to snap a number of pictures and filter them based on visual clarity. While addressing the technical aspects of data collection devices, it's equally important to recognise that the complexity of farming fields extends far beyond these tools. By systematically capturing data across diverse fields, times of day, sunlight levels, and different days, a more robust and representative dataset can be collected that helps reduce biases compared to only sampling from a small subset of the highly variable farm conditions. This comprehensive data collection approach can help to minimize biases in the data (Lenain et al., 2021). Furthermore, in leaf disease diagnosis, algorithms should be designed to recognise that a single leaf may be affected by multiple diseases simultaneously. This consideration is crucial for accurate disease detection and management, ensuring that all potential diseases are identified and addressed effectively.

The impact of closing this gender gap varies by region, influenced by the extent of women's engagement in agriculture, their control over production and land, and the existing gender disparities. To mitigate gender bias, targeted policy interventions are crucial. These should focus on eliminating discrimination against women in accessing agricultural resources, education, extension services, financial services, and labour markets. Additionally, fostering women's participation in flexible, efficient, and fair rural labour markets is essential (Quisumbing et al., 2014). By implementing these reforms, we can make significant strides towards closing the gender gap and achieving equitable growth in the agricultural sector.

Geographic Information Systems (GIS) play a crucial role in mitigating geographical bias in agriculture by capturing, analyzing, and managing spatial data. GIS enables farmers to map and assess soil properties, crop conditions, and environmental factors, optimizing resource use like fertilizers, water, and pesticides according to specific field needs. It aids in disaster management by predicting and mitigating the impact of natural events and streamlines supply chain logistics by optimizing routes and storage (Sergieieva, 2022). By leveraging GIS, farmers can adapt to local conditions and utilize insights from broader data, leading to more sustainable and

productive agricultural practices. Additionally, to ensure optimal growth, plants must receive precise nutrients tailored to their specific needs. Given that the required nutrients vary based on factors such as plant type, soil, water availability, and weather conditions, it is more effective to treat each plant individually within the same region rather than applying a generalised approach.

In agriculture, maintaining reliable sensor networks demands the use of redundant sensors and robust designs (Rehman et al., 2014). Equipment downtime can significantly impact data collection and result in financial losses. This issue is further compounded by restrictions from manufacturers, which often limit the ability to repair machines promptly and effectively. To mitigate this, predictive maintenance offers a solution. By using data from sensors and cameras, advanced analytics can monitor equipment health in real-time, detect any deviations from normal performance. This technology predicts potential failures before they occur, allowing timely interventions. Additionally, it helps manage aspects like fuel usage, machine availability, and service schedules, optimizing equipment operation and ensuring long-term reliability. Implementing predictive maintenance thus reduces technical biases and enhances overall efficiency in farming operations (Kyslyi & Kovalenko, 2024). Moreover, for effective implementation, agricultural field has to be configured such that robots and AI sensor systems can be used effectively to prevent damage to the plants (Séverac et al., 2021). Crops should be seeded with adequate spacing to facilitate the unobstructed movement of robots within the field (Li et al., 2022). Establishing clear, designated pathways ensures that robots can move efficiently across the field and reach different areas without obstructing the work of human pickers. Additionally, implementing robust connectivity solutions is crucial for real-time data transmission, enabling AI systems to monitor conditions and make timely decisions to support the picking process and prevent plant damage.

The economics of conservation agriculture are complex and depend on the unique situation of individual farmers. Site-specific analysis is needed to properly assess the potential for adoption in a given region (Pannell et al., 2014). Decision-makers and data developers should shift from market-driven to community-centered governance to address biases in agriculture. This approach fosters inclusive alternatives to traditional intensification and commercialisation, supporting marginalised groups such as small-scale, migrant, and indigenous farmers (Bronson et al., 2021). It is crucial to diversify training datasets and consider the unique needs and constraints of diverse farming communities to mitigate socioeconomic bias. By focusing on community-led strategies and equitable access to technology, we can promote sustainable agricultural practices and ensure that all farmers benefit from technological advancement.

The primary markets for AI tools in agriculture are expected to be the United States, Europe, India and potentially China (Chandra, 2023; Sparrow et al., 2021). However, a significant concern arises regarding the applicability of AI systems trained on data from these regions for local contexts elsewhere. In cases where AI systems are retrained using local data, a bias towards certain crops might still persist, impacting the suitability of the technology for various agricultural practices (Sparrow et al., 2021). Since AI applications demand a vast quantity of data, farmers should express interest in contributing to the data collection for training the AI models in order to consider different soil and other factors that vary geographically. Likewise,

developers must prioritize privacy and confidentiality during data collection. Promote the use of XAI models in agriculture, allowing farmers to understand how AI recommendations or decisions are made. Transparent AI systems can help build trust and enable farmers to identify and question any biased outcomes. Koenderink et al. (2010) introduced the notion of "bounded transparency" which balances transparency and opacity based on practical considerations in agriculture.

While current XAI methods offer explanations based on feature activation and importance, these explanations may not be sufficient to fully comprehend the decision-making process in the context of agricultural technology, particularly in disease recognition applications. As a result, they have limited practical use in real-world agricultural scenarios. To address this limitation, it is essential to transform these explanations into human-understandable forms. One potential approach is to integrate the concepts of expert systems with the activated convolution features of XAI, thereby providing more interpretable and meaningful explanations for agricultural experts and stakeholders (Ranasinghe et al., 2022). Additionally, plant nutrient deficiencies vary in presentation across early, mild, and severe stages, with overlapping symptoms between different deficiencies. For instance, calcium and boron deficiencies can initially show similar symptoms but diverge as they progress. CNN models used for diagnosis may yield varied and uncertain predictions due to evolving symptoms. Applying XAI models to interpret CNN outputs may enhance understanding but could be limited by model calibration issues. Given the complexity of symptom evolution, traditional XAI methods may lack reliability. A stage-based analysis of features is recommended to better correlate feature activations with probable deficiency classes (Ranasinghe et al., 2022).

Encouraging collaboration between agricultural experts, data scientists, and social scientists can better understand the socio-economic and cultural aspects of farming. An interdisciplinary approach can help identify and address biases in AI systems for agriculture more effectively. Educating farmers, agricultural extension workers, and other stakeholders about AI-based technologies, their potential sources of biases, and how to interpret and responsibly use recommendations generated by AI systems is crucial for responsible deployment and adoption of these technologies in agriculture. The European Commission is actively supporting AI initiatives in agriculture through programs like the Common European Agricultural Data Space, AI testing facilities under Horizon Europe, and Digital Innovation Hubs (DIHs) (Baerdemaeker et al., 2023). Asian countries are increasingly supporting AI advancements in agriculture through various initiatives. For instance, India's AgriStack is designed to create a comprehensive digital ecosystem for farmers, integrating technology to improve agricultural efficiency and productivity (Beriya, 2022). Efforts are also being made to enhance farmers' digital skills and validate innovative technologies through demonstration cases.

AI models should be regularly monitored and evaluated for biases after deployment, and continuously updated and retrained to ensure they remain fair and relevant by considering changes in agricultural practices and external factors that may introduce new biases over time. Moreover, Ryan (2022) discussed one of the key social and ethical challenges is sustainability. Several articles emphasize sustainability as a major concern for farmers and the agricultural sector. Deploying AI must not

only meet current needs but also help ensure long-term environmental, economic and social sustainability of agriculture (Ryan, 2022; Shankar et al., 2020). However, sustainability receives much less attention in general AI ethics guidelines compared to its importance in discussions of agricultural AI, indicating a disconnect that needs to be addressed. Ensuring AI technologies are developed and implemented with sustainability top of mind will be important for responsible development in this sector.

5 Conclusion

In this manuscript, we explore the prevalence of bias in AI systems, and how it affects the way the systems perform, as well as existing strategies for circumventing the bias. This study further considers the agriculture sector, where AI-based decision-making systems are often used. Initially, we introduce different types of biases that exist within the AI pipeline. Having identified six types of main biases through our literature search, namely, data-centric bias, social bias, algorithmic bias, performance measure bias, feedback bias, and deployment bias, we describe them in detail. Some of the biases are interleaved with each other. While AI has tremendous potential to improve our lives, the biases introduced during data collection, model development, and system deployment can negatively affect user experiences and outcomes if not properly addressed.

We discuss an overview of strategies to mitigate biases across various phases of the AI pipeline, from problem understanding to maintenance. We identified that mitigation strategies to address bias can generally be addressed using technical methodologies with the exception of cases where social bias needs to be addressed with human involvement, diversification of participants, technical experts and open science. Effective bias mitigation requires multidisciplinary collaboration and an iterative approach that continuously evaluates models as contexts change over time. Open communication of limitations and outcomes is also important to build public trust. However, it's important to note that while some mitigation strategies effectively address existing biases, they might inadvertently introduce new biases. Additionally, we provide a list of available tools that aid in mitigating biases.

Biases can significantly impact the effectiveness and ethics of AI applications in agriculture if not properly addressed. A variety of biases, such as data, geographic, technological, socioeconomic and cultural biases, may arise at different stages of developing and deploying agricultural AI systems. It is crucial to adopt measures to identify, understand and mitigate these biases to ensure fair, transparent and beneficial use of AI. An interdisciplinary approach involving farmers, domain experts, data scientists and social scientists can facilitate comprehensive bias analysis and mitigation strategies. Collaborative data collection representing diverse conditions can make models more robust and inclusive. In addition to employing techniques such as data preprocessing, model regularization, and post-deployment monitoring to enhance fairness, regulatory frameworks ensuring data ownership and privacy are indispensable. The acceptance and utilisation of AI systems by farmers hinge on robust regulation and security protocols. Meanwhile, educational initiatives and transparent model explanations can build trust among stakeholders. Overall, a con-

certed effort is needed across technical, social and policy dimensions to develop agricultural AI responsibly and maximize its potential to support sustainable and equitable food systems worldwide.

To conclude, this manuscript provides an in-depth review of the world of biases in applications of AI, along with how to mitigate those biases. The agricultural application areas discussed are intended to give an overall idea of how to identify and mitigate issues related to bias in the field. Although this is not a systematic review, it is presented with the goal of providing value to stakeholders who engage with AI tools and pipelines in the agricultural sector. In closing, addressing bias in agricultural AI will require a holistic and multi-disciplinary research approach. Future work must prioritize developing standardized methodologies to detect and mitigate technical biases by employing quantitative analysis, while also exploring the social dimensions of bias through studies of rural communities. A deeper understanding of local dynamics is vital for equitable design and deployment of technologies. A holistic approach, incorporating diverse perspectives and long-term evaluations, can guide the development of agricultural AI that empowers stakeholders equitably, fosters trust, and promotes fair and sustainable growth through the power of AI.

Acknowledgements The authors thank Professor Saman Halgamuge (University of Melbourne, Australia) for initiating this work and sharing his insights in FAIR AI.

Author Contributions Conceptualization: Damith, Rajith, Ramanan and Mayuravaani; Literature Search: Mayuravaani and Ramanan; Writing: Original draft preparation: Mayuravaani and Ramanan; Review and Editing: Maneesha.

Funding No external funding.

Data Availability No datasets were generated or analysed during the current study.

Declarations

Ethics Approval and Consent to Participate Not applicable.

Conflict of Interest On behalf of all authors, the corresponding author states that there is no conflict of (or competing) interest.

References

Agarwal, A., Dudík, M., & Wu, Z. S. (2019). Fair regression: Quantitative definitions and reduction-based algorithms. *International conference on machine learning* (pp. 120–129).

Arazo, E., Ortego, D., Albert, P., O'Connor, N. E., & McGuinness, K. (2020). Pseudo-labeling and confirmation bias in deep semi-supervised learning. *2020 international joint conference on neural networks (IJCNN)* (pp. 1–8).

Atemezing, G. A., & Troncy, R. (2013). Towards interoperable visualization applications over linked data. Talk given at the 2nd European data forum (EDF), Dublin, Ireland (april 2013). http://goo.Gl/jhvrax

Baerdemaeker, J. D., et al. (2023). *Artificial intelligence in the agri-food sector applications, risks and impacts.* STUDY - Panel for the Future of Science and Technology. Retrieved Aug 20, 2023 from https://www.cema-agri.org/publications/21-articles/1013-european-parliament-think-tank-publishes-study-on-artificial-intelligence-in-the-agri-food-sector

Barbedo, J. G. A. (2016). A review on the main challenges in automatic plant disease identification based on visible range images. *Biosystems Engineering*, *144*, 52–60. https://doi.org/10.1016/j.biosystemseng.2016.01.017

Barber, D. (2012). *Bayesian reasoning and machine learning*. Cambridge University Press.

Basolo, A. L. (1995). Phylogenetic evidence for the role of a pre-existing bias in sexual selection. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *259*(1356), 307–311. https://doi.org/10.1098/rspb.1995.0045

Bastedo, M. N., Umbricht, M., Bausch, E., Byun, B.-K., & Bai, Y. (2023). Contextualized high school performance: Evidence to inform equitable holistic, test-optional, and test-free admissions policies. *AERA Open*, *9*, 23328584231197413.

Bellamy, R. K., Dey, K., Hind, M., Hoffman, S. C., Houde, S., Kannan, K., Lohia, P., Martino, J., Mehta, S., Mojsilovic, A., Nagar, S., Ramamurthy, K. N., Richards, J., Saha, D., Sattigeri, P., Singh, M., Varshney, K. R., & Zhang, Y. (2019). AI fairness 360: An extensible toolkit for detecting and mitigating algorithmic bias. *IBM Journal of Research and Development*, *63*(4/5), 4–1. https://doi.org/10.48550/arXiv.1810.01943

Belleflamme, P., & Peitz, M. (2018). Inside the engine room of digital platforms: Reviews, ratings, and recommendations. https://doi.org/10.2139/ssrn.3128141

Beriya, A. (2022). India digital ecosystem of agriculture and agristack: An initial assessment (Tech. Rep.). ICT India Working Paper.

Bhagat, P. R., Naz, F., & Magda, R. (2022). Artificial intelligence solutions enabling sustainable agriculture: A bibliometric analysis. *PloS One*, *17*(6), e0268989.

Bolam, M. R., Corbett, L. E., Ellero, N. P., Stein Kenfield, A., Mitchell, E. T., Opasik, S. A., & Ryszka, D. (2018). Current work in diversity, inclusion and accessibility by metadata communities: A working report from the ala/alcts metadata standards committee. *Technical Services Quarterly*, *35*(4), 367–376. https://doi.org/10.1080/07317131.2018.1509439

Bolukbasi, T., Chang, K.-W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. *Advances in Neural Information Processing Systems*, *29*.

Brewer, S., Pearson, S., Maull, R., Godsiff, P., Frey, J. G., Zisman, A., Parr, G., McMillan, A., Cameron, S., Blackmore, H., Manning, L., & Bidaut, L. (2021). A trust framework for digital food systems. *Nature Food*, *2*(8), 543–545. https://doi.org/10.1038/s43016-021-00346-1

Bronson, K., Rotz, S., & D'Alessandro, A. (2021). The human impact of data bias and the digital agricultural revolution. In *Handbook on the human impact of agriculture* (pp. 119–137). Edward Elgar Publishing.

Budiman, A., Fanany, M. I., & Basaruddin, C. (2016). *Adaptive online sequential ELM for concept drift tackling*. Computational intelligence and neuroscience, 2016. https://doi.org/10.1155/2016%2F8091267.

Buet-Golfouse, F., & Utyagulov, I. (2022). Towards fair unsupervised learning. *Proceedings of the 2022 ACM conference on fairness, accountability, and transparency* (pp. 1399–1409).

Bulò, S. R., Porzi, L., & Kontschieder, P. (2016). Dropout distillation. *International conference on machine learning* (pp. 99–107).

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Conference on fairness, accountability and transparency* (pp. 77–91).

Byerlee, D., Stevenson, J., & Villoria, N. (2014). Does intensification slow crop land expansion or encourage deforestation? *Global Food Security*, *3*(2), 92–98.

Camaréna, S. (2021). Engaging with Artificial intelligence (AI) with a bottom-up approach for the purpose of sustainability: Victorian farmers market association, Melbourne Australia. *Sustainability*, *13*(16), 9314.

Chandra, V. S. (2023). Role of artificial intelligence in indian agriculture: A review. *Agricultural Reviews*, *44*(4), 558–562.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). Smote: Synthetic minority oversampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/jair.953

Chen, I., Johansson, F. D., & Sontag, D. (2018). Why is my classifier discriminatory? *Advances in Neural Information Processing Systems*, *31*. https://doi.org/10.48550/arXiv.1805.12002

Chergui, N., & Kechadi, M. T. (2022). Data analytics for crop management: A big data view. *Journal of Big Data*, *9*(1), 1–37.

Chierichetti, F., Kumar, R., Lattanzi, S., & Vassilvitskii, S. (2017). Fair clustering through fairlets. *Advances in Neural Information Processing systems*, *30*.

Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big Data*, *5*(2), 153–163. https://doi.org/10.48550/arXiv.1703.00056

Claudy, M. C., Aquino, K., & Graso, M. (2022). Artificial intelligence can't be charmed: The effects of impartiality on laypeople's algorithmic preferences. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.898027

Crabbé, J., & van der Schaar, M. (2022). Label-free explainability for unsupervised models. arXiv preprint arXiv:2203.01928

Cunningham, P., & Delany, S. J. (2021). Underestimation bias and underfitting in machine learning. *Trustworthy AI-integrating learning, optimization and reasoning: First international workshop, tailor 2020, virtual event, september 4–5, 2020, revised selected papers 1* (pp. 20–31).

d'Alessandro, B., O'Neil, C., & LaGatta, T. (2017). Conscientious classification: A data scientist's guide to discrimination-aware classification. *Big Data*, *5*(2), 120–134. https://doi.org/10.48550/arXiv.1907.09013

Dara, R., Hazrati Fard, S. M., & Kaur, J. (2022). Recommendations for ethical and responsible use of artificial intelligence in digital agriculture. *Frontiers in Artificial Intelligence*, *5*, 884192.

Dastin, J. (2018). Amazon scraps secret ai recruiting tool that showed bias against women. 296–299.

Datta, A., Fredrikson, M., Ko, G., Mardziel, P., & Sen, S. (2017). Proxy non-discrimination in data-driven systems. arXiv preprint arXiv:1707.08120. https://doi.org/10.48550/arXiv.1707.08120

Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world. *Harvard Business Review*, *96*(1), 108–116.

Diethe, T., Borchert, T., Thereska, E., Balle, B., & Lawrence, N. (2019). Continual learning in practice. arXiv preprint arXiv:1903.05202. https://doi.org/10.48550/arXiv.1903.05202

Du, W., & Wu, X. (2021). Fair and robust classification under sample selection bias. *Proceedings of the 30th acm international conference on information & knowledge management* (pp. 2999–3003).

Dubey, S. R., & Jalal, A. S. (2015). Application of image processing in fruit and vegetable analysis: A review. *Journal of Intelligent Systems*, *24*(4), 405–424.

Emmanuel, T., Maupong, T., Mpoeleng, D., Semong, T., Mphago, B., & Tabona, O. (2021). A survey on missing data in machine learning. *Journal of Big Data*, *8*(1), 1–37. https://doi.org/10.1186/s40537-021-00516-9

Escalante, C. L., Epperson, J. E., & Raghunathan, U. (2009). Gender bias claims in farm service agency's lending decisions. *Journal of Agricultural and Resource Economics*, 332–349. https://doi.org/10.22004/AG.ECON.54550

Faghani, S., Khosravi, B., Zhang, K., Moassefi, M., Jagtap, J. M., Nugen, F., Vahdati, S., Kuanar, S. P., Rassoulinejad-Mousavi, S. M., Singh, Y., Vera Garcia, D. V., Rouzrokh, P., & Erickson, B. J. (2022). Mitigating bias in radiology machine learning: 3. Performance metrics. *Radiology: Artificial Intelligence*, *4*(5), e220061. https://doi.org/10.1148/ryai.220061

Feurer, M., Klein, A., Eggensperger, K., Springenberg, J., Blum, M., & Hutter, F. (2015). Efficient and robust automated machine learning. *Advances in Neural Information Processing Systems*, *28*. https://doi.org/10.1007/978-3-030-05318-56

Ganaie, M. A., Hu, M., Malik, A., Tanveer, M., & Suganthan, P. (2022). Ensemble deep learning: A review. *Engineering Applicationsntroduction to Bayesian Networ of Artificial Intelligence*, *115*, 105151. https://doi.org/10.48550/arXiv.2104.02395

Gao, Y., Gao, L., Li, X., & Yan, X. (2020). A semi-supervised convolutional neural network-based method for steel surface defect recognition. *Robotics and Computer-Integrated Manufacturing*, *61*, 101825. https://doi.org/10.1016/j.rcim.2019.101825

Gardezi, M., Joshi, B., Rizzo, D. M., Ryan, M., Prutzer, E., Brugler, S., & Dadkhah, A. (2023). Artificial intelligence in farming: Challenges and opportunities for building trust. *Agronomy Journal*.

Gijsbers, P., & Vanschoren, J. (2019). Gama: Genetic automated machine learning assistant. *Journal of Open Source Software*, *4*(33), 1132. https://doi.org/10.21105/joss.01132

Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., & Kagal, L. (2018). Explaining explanations: An overview of interpretability of machine learning. *2018 IEEE 5th international conference on data science and advanced analytics (DSAA)* (pp. 80–89).

Gjoka, M., Kurant, M., Butts, C. T., & Markopoulou, A. (2010). Walking in facebook: A case study of unbiased sampling of OSNS. *2010 proceedings IEEE infocom*, 1–9.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, *63*(11), 139–144. https://doi.org/10.48550/arXiv.1406.2661

Gorski, L., Ramakrishna, S., & Nowosielski, J. M. (2020). Towards Grad-CAM based explainability in a legal text processing pipeline. arXiv preprint arXiv:2012.09603. https://doi.org/10.48550/arXiv.2012.09603

Greshake Tzovaras, B., & Tzovara, A. (2019). The personal data is political. *The Ethics of Medical Data Donation*, 133–140. https://doi.org/10.1007/978-3-0302370188116894-04363-6

Gu, Y. H., Jin, D., Yin, H., Zheng, R., Piao, X., & Yoo, S. J. (2022). Forecasting agricultural commodity prices using dual input attention LSTM. *Agriculture*, *12*(2), 256. https://doi.org/10.3390/agriculture12020256

Halgamuge, S. (2021). FAIR AI: A conceptual framework for democratisation of 21st century AI. *2021 international conference on instrumentation, control, and automation (ICA)* (pp. 1–3).

Haque, M., Saxena, D., Weathington, K., Chudzik, J., & Guha, S. (2024). Are we asking the right questions?: Designing for community stakeholders' interactions with ai in policing. arXiv preprint arXiv:2402.05348

Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (Vol. 2). Springer.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the Econometric Society*, 153–161. https://doi.org/10.2307/1912352

Hellström, T., Dignum, V., & Bensch, S. (2020). Bias in machine learning–what is it good for? arXiv preprint arXiv:2004.00686. https://doi.org/10.48550/arXiv.2004.00686

Holmes, D. E., & Jain, L. C. (2008). *Introduction to bayesian networks*. Springer.

Howard, A., & Borenstein, J. (2018). The ugly truth about ourselves and our robot creations: The problem of bias and social inequity. *Science and Engineering Ethics*, *24*, 1521–1536. https://doi.org/10.1007/s11948-017-9975-2

Jabbar, H., & Khan, R. Z. (2015). Methods to avoid over-fitting and under-fitting in supervised machine learning (comparative study). *Computer Science, Communication and Instrumentation Devices*, *70*, 163–172. https://doi.org/10.3850/978-981-09-5247-1017

Jabbari, S., Joseph, M., Kearns, M., Morgenstern, J., & Roth, A. (2017). Fairness in reinforcement learning. *International conference on machine learning* (pp. 1617–1626).

Jameel, S. M., Hashmani, M. A., Alhussain, H., Rehman, M., & Budiman, A. (2020). A critical review on adverse effects of concept drift over machine learning classification models. *International Journal of Advanced Computer Science and Applications*, *11*(1). https://doi.org/10.14569/ijacsa.2020.0110127

James, W. (2017). *Facets: An open source visualization tool for machine learning training data*. Google AI Blog. Retrieved 2022-11-10, from https://ai.googleblog.com/2017/07/facets-open-source-visualization-tool.html

Jing, Y., Yang, Y., Feng, Z., Ye, J., Yu, Y., & Song, M. (2019). Neural style transfer: A review. *IEEE Transactions on Visualization and Computer Graphics*. *26*(11), 3365–3385. https://doi.org/10.1109/TVCG.2019.2921336

Johannes, A., Picon, A., Alvarez-Gila, A., Echazarra, J., Rodriguez-Vaamonde, S., Navajas, A. D., & Ortiz-Barredo, A. (2017). Automatic plant disease diagnosis using mobile capture devices, applied on a wheat use case. *Computers and Electronics in Agriculture*, *138*, 200–209. https://doi.org/10.1016/j.compag.2017.04.013

Johnson, B., & Brun, Y. (2022). Fairkit-learn: A fairness evaluation and comparison toolkit. *Proceedings of the ACM/IEEE 44th international conference on software engineering: Companion proceedings* (pp. 70–74).

Joo, H.-T., & Kim, K.-J. (2019). Visualization of deep reinforcement learning using Grad-CAM: How AI plays atari games? *2019 IEEE conference on games (COG)* (pp. 1–2).

Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, *147*, 70–90.

Kamishima, T., Akaho, S., Asoh, H., & Sakuma, J. (2012). Fairness-aware classifier with prejudice remover regularizer. *Machine learning and knowledge discovery in databases: European conference, ECML PKDD 2012, Bristol, Uk, september 24–28, 2012. proceedings, Part II 23* (pp. 35–50).

Koenderink, N. J., Broekstra, J., & Top, J. L. (2010). Bounded transparency for automated inspection in agriculture. *Computers and Electronics in Agriculture*, *72*(1), 27–36.

Kusner, M. J., Loftus, J., Russell, C., & Silva, R. (2017). Counterfactual fairness. *Advances in Neural Information Processing Systems, 30*. https://doi.org/10.48550/arXiv.1703.06856

Kyslyi, A., & Kovalenko, S. (2024). *Key agro challenges solved by advanced data analytics*. Infopulse. Retrieved June 21, 2024 from https://www.infopulse.com/blog/data-analytics-use-cases-agriculture

Lagioia, F., et al. (2020). The impact of the general data protection regulation (GDPR) on artificial intelligence. https://doi.org/10.2861/293

Lee, J., Gadsden, S. A., Biglarbegian, M., & Cline, J. A. (2022). Smart agriculture: A fruit flower cluster detection strategy in apple orchards using machine vision and learning. *Applied Sciences*, *12*(22), 11420.

Lenain, R., Peyrache, J., Savary, A., & Séverac, G. (2021). *Agricultural robotics: Part of the new deal?: Fira 2020 conclusions*. éditions Quae.

Li, X., Lloyd, R., Ward, S., Cox, J., Coutts, S., & Fox, C. (2022). Robotic crop row tracking around weeds using cereal-specific features. *Computers and Electronics in Agriculture*, *197*, 106941.

Lin, Y.-P., Petway, J. R., & Settele, J. (2017). Train artificial intelligence to be fair to farming. *Nature*, *552*(7683), 334–335. https://doi.org/10.1038/d41586-017-08881-3

Linsner, S., Steinbrink, E., Kuntke, F., Franken, J., & Reuter, C. (2022). Supporting users in data disclosure scenarios in agriculture through transparency. *Behaviour & Information Technology*, *41*(10), 2151–2173. https://doi.org/10.1080/0144929X.2022.2068070

Liu, H., & Cocea, M. (2017). Granular computing-based approach for classification towards reduction of bias in ensemble learning. *Granular Computing*, *2*, 131–139. https://doi.org/10.1007/s41066-016-0034-1

Lohia, P. (2021). Priority-based post-processing bias mitigation for individual and group fairness. arXiv preprint arXiv:2102.00417. https://doi.org/10.48550/arXiv.2102.00417

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30*. https://doi.org/10.48550/arXiv.1705.07874

Martin, K. (2019, June). Designing ethical algorithms. *MIS Quarterly Executive*. https://doi.org/10.17705/2msqe.00012

Mayuravaani, M., & Manivannan, S. (2021). A semi-supervised deep learning approach for the classification of steel surface defects. *2021 10th international conference on information and automation for sustainability(iciafs)* (pp. 179–184).

McVey, C., Hsieh, F., Manriquez, D., Pinedo, P., & Horback, K. (2023). Invited review: Applications of unsupervised machine learning in livestock behavior: Case studies in recovering unanticipated behavioral patterns from precision livestock farming data streams. *Applied Animal Science*, *39*(2), 99–116.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys (CSUR)*, *54*(6), 1–35. https://doi.org/10.48550/arXiv.1908.09635

Mirowski, P. (2018). The future (s) of open science. *Social Studies of Science*, *48*(2), 171–203. https://doi.org/10.1177/0306312718772086

Mondello, V., Songy, A., Battiston, E., Pinto, C., Coppin, C., Trotel-Aziz, P., & Fontaine, F. (2018). Grapevine trunk diseases: A review of fifteen years of trials for their control with chemicals and biocontrol agents. *Plant Disease*, *102*(7), 1189–1217. https://doi.org/10.1094/pdis-08-17-1181-fe

Moore, H. E., & Rutherfurd, I. D. (2020). Researching agricultural environmental behaviour: Improving the reliability of self-reporting. *Journal of Rural Studies*, *76*, 296–304.

Mourtzinis, S., Esker, P. D., Specht, J. E., & Conley, S. P. (2021). Advancing agricultural research using machine learning algorithms. *Scientific Reports*, *11*(1), 17879.

Mundhenk, T. N., Chen, B. Y., & Friedland, G. (2019). Efficient saliency maps for explainable AI. arXiv preprint arXiv:1911.11293. https://doi.org/10.48550/arXiv.1911.11293

Nabi, R., Malinsky, D., & Shpitser, I. (2019). Learning optimal fair policies. *International conference on machine learning* (pp. 4674–4682).

Neal, B., Mittal, S., Baratin, A., Tantia, V., Scicluna, M., Lacoste-Julien, S., & Mitliagkas, I. (2018). A modern take on the bias-variance tradeoff in neural networks. arXiv preprint arXiv:1810.08591. https://doi.org/10.48550/arXiv.1810.08591

Norori, N., Hu, Q., Aellen, F. M., Faraci, F. D., & Tzovara, A. (2021). Addressing bias in big data and AI for health care: A call for open science. *Patterns*, *2*(10), 100347. https://doi.org/10.1016/j.patter.2021.100347

Nunan, D., Aronson, J., & Bankhead, C. (2018). Catalogue of bias: Attrition bias. *BMJ Evidence-based Medicine*, *23*(1), 21–22. https://doi.org/10.1136/ebmed-2017-110883

O'donovan, P., Leahy, K., Bruton, K., & O'Sullivan, D. T. (2015). Big data in manufacturing: A systematic mapping study. *Journal of Big Data*, *2*, 1–22. https://doi.org/10.1186/s40537-015-0028-x

Okengwu, U., Onyejegbu, L., Oghenekaro, L., Musa, M., & Ugbari, A. (2023). Environmental and ethical negative implications of ai in agriculture and proposed mitigation measures. *Scientia Africana*, *22*(1), 141–150.

Pádua, L., Chiroque-Solano, P. M., Marques, P., Sousa, J. J., & Peres, E. (2022). Mapping the leaf area index of castanea sativa miller using uav-based multispectral and geometrical data. *Drones*, *6*(12), 422.

Paleyes, A., Urma, R.-G., & Lawrence, N. D. (2022). Challenges in deploying machine learning: A survey of case studies. *ACM Computing Surveys*, *55*(6), 1–29. https://doi.org/10.48550/arXiv.2011.09926

Pannell, D. J., Llewellyn, R. S., & Corbeels, M. (2014). The farm-level economics of conservation agriculture for resource-poor farmers. *Agriculture, Ecosystems & Environment*, *187*, 52–64.

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., & Lerer, A. (2017). Automatic differentiation in pytorch.

Plevris, V., Solorzano, G., Bakas, N. P., & Ben Seghier, M. E. A. (2022). Investigation of performance metrics in regression analysis and machine learning-based prediction models. *8th European congress on computational methods in applied Sciences and engineering (ECCOMAS Congress 2022)*.

Pot, M., Kieusseyan, N., & Prainsack, B. (2021). Not all biases are bad: Equitable and inequitable biases in machine learning and radiology. *Insights into Imaging*, *12*(1), 1–10. https://doi.org/10.1186/s13244-020-00955-7

Prince, A. E., & Schwarcz, D. (2019). Proxy discrimination in the age of artificial intelligence and big data. *Iowa Law Review*, *105*, 1257.

Puiutta, E., & Veith, E. M. (2020). Explainable reinforcement learning: A survey. *International cross-domain conference for machine learning and knowledge extraction* (pp. 77–95).

Quinonero-Candela, J., Sugiyama, M., Schwaighofer, A., & Lawrence, N. D. (2008). *Dataset shift in machine learning*. MIT Press. Retrieved from https://ieeexplore.ieee.org/servlet/opac?bknumber=6267199

Quisumbing, A. R., Meinzen-Dick, R., Raney, T. L., Croppenstedt, A., Behrman, J. A., & Peterman, A. (2014). Closing the knowledge gap on gender in agriculture. *Gender in Agriculture: Closing the Knowledge Gap*, 3–27.

Ranasinghe, N., Ramanan, A., Fernando, S., Hameed, P., Herath, D., Malepathirana, T., & Halgamuge, S. (2022). Interpretability and accessibility of machine learning in selected food processing, agriculture and health applications. *Journal of the National Science Foundation of Sri Lanka*, *50*, 263–276. https://doi.org/10.4038/jnsfsr.v50i0.11249

Rauf, U., Qureshi, W. S., Jabbar, H., Zeb, A., Mirza, A., Alanazi, E., & Rashid, N. (2022). A new method for pixel classification for rice variety identification using spectral and time series data from sentinel-2 satellite imagery. *Computers and Electronics in Agriculture*, *193*, 106731.

Rayana, S., Zhong, W., & Akoglu, L. (2016). Sequential ensemble learning for outlier detection: A bias-variance perspective. *2016 IEEE 16th international conference on data mining (ICDM)* (pp. 1167–1172).

Rehman, A. U., Abbasi, A. Z., Islam, N., & Shaikh, Z. A. (2014). A review of wireless sensors and networks' applications in agriculture. *Computer Standards & Interfaces*, *36*(2), 263–270.

Rehman, F., Muhammad, S., Ashraf, I., Mahmood, C. K., Ruby, T., & Bibi, I. (2013). Effect of farmers' socioeconomic characteristics on access to agricultural information: Empirical evidence from pakistan. *Journal of Animal and Plant Sciences*, *23*, 324–329. Retrieved from https://api.semanticscholar.org/CorpusID:86290768

Restrepo-Arias, J. F., Branch-Bedoya, J. W., & Awad, G. (2022). Plant disease detection strategy based on image texture and bayesian optimization with small neural networks. *Agriculture*, *12*(11), 1964.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135–1144).

Rios, R., Miller, R. J. H., Manral, N., Sharir, T., Einstein, A. J., Fish, M. B., Ruddy, T. D., Kaufmann, P. A., Sinusas, A. J., Miller, E. J., Bateman, T. M., Dorbala, S., Di Carli, M., Van Kriekinge, S. D., Kavanagh, P. B., Parekh, T., Liang, J. X., Dey, D., Berman, D. S., & Slomka, P. J. (2022). Handling missing values in machine learning to predict patient-specific risk of adverse cardiac events: Insights from refine spect registry. *Computers in Biology and Medicine*, *145*, 105449. https://doi.org/10.1016/j.compbiomed.2022.105449

Robinson, S., Narayanan, B., Toh, N., & Pereira, F. (2014). Methods for pre-processing smartcard data to improve data quality. *Transportation Research Part C: Emerging Technologies*, *49*, 43–58. https://doi.org/10.1016/j.trc.2014.10.006

Roelofs, R., Shankar, V., Recht, B., Fridovich-Keil, S., Hardt, M., Miller, J., & Schmidt, L. (2019). A meta-analysis of overfitting in machine learning. *Advances in Neural Information Processing Systems*, *32*.

Roh, Y., Lee, K., Whang, S. E., & Suh, C. (2020). Fairbatch: Batch selection for model fairness. arXiv preprint arXiv:2012.01696. https://doi.org/10.48550/arXiv.2012.01696

Ros, F., Riad, R., & Guillaume, S. (2023). Pdbi: A partitioning davies-bouldin index for clustering evaluation. *Neurocomputing*, *528*, 178–199.

Roscher, R., Bohn, B., Duarte, M. F., & Garcke, J. (2020). Explainable machine learning for scientific insights and discoveries. *IEEE Access*, *8*, 42200–42216. https://doi.org/10.1109/ACCESS.2020.2976199

Ryan, M. (2022). The social and ethical impacts of artificial intelligence in agriculture: Mapping the agricultural AI literature. *AI & Society*, 1–13.

Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. *PloS One*, *10*(3), e0118432.

Salam, M. A., Azar, A. T., Elgendy, M. S., & Fouad, K. M. (2021). The effect of different dimensionality reduction techniques on machine learning overfitting problem. *International Journal of Advanced Computer Science and Applications*, *12*(4), 641–655. https://doi.org/10.14569/IJACSA.2021.0120480

Samadi, S., Tantipongpipat, U., Morgenstern, J. H., Singh, M., & Vempala, S. (2018). The price of fair pca: One extra dimension. *Advances in Neural Information Processing Systems*, *31*.

Sambasivam, G., & Opiyo, G. D. (2021). A predictive machine learning application in agriculture: Cassava disease detection and classification with imbalanced dataset using convolutional neural networks. *Egyptian Informatics Journal*, *22*(1), 27–34. https://doi.org/10.1016/j.eij.2020.02.007

Saranya, A., & Subhashini, R. (2023). A systematic review of explainable artificial intelligence models and applications: Recent developments and future trends. *Decision Analytics Journal*, *100230*. https://doi.org/10.1016/j.dajour.2023.100230

Sengupta, K., & Srivastava, P. R. (2022). Causal effect of racial bias in data and machine learning algorithms on user persuasiveness & discriminatory decision making: An empirical study. arXiv preprint arXiv:2202.00471. https://doi.org/10.48550/arXiv.2202.00471

Sergieieva, K. (2022) Gis in agriculture: Best practices for agritech leaders. *Earth Observing System*. Retrieved 2024-06-21, from https://eos.com/blog/gis-in-agriculture/

Séverac, G., Savary, A., Peyrache, J., & Lenain, R. (2021). Agricultural robotics: Part of the new deal? Fira 2020 conclusions: With 27 agricultural robot information sheets.

Shamshiri, R., Kalantari, F., Ting, K., Thorp, K. R., Hameed, I. A., Weltzien, C., Ahmad, D., & Shad, Z. M. (2018). Advances in greenhouse automation and controlled environment agriculture: A transition to plant factories and urban agriculture. https://doi.org/10.25165/j.ijabe.20181101.3210

Shankar, P., Werner, N., Selinger, S., & Janssen, O. (2020). Artificial intelligence driven crop protection optimization for sustainable agriculture. *2020 IEEE/ITU international conference on artificial intelligence for good (AI4G)* (pp. 1–6).

Shikuku, K. M. (2019). Information exchange links, knowledge exposure, and adoption of agricultural technologies in northern uganda. *World Development*, *115*, 94–106.

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, *6*(1), 1–48. https://doi.org/10.1186/s40537-019-0197-0

Sjoding, M. W., Dickson, R. P., Iwashyna, T. J., Gay, S. E., & Valley, T. S. (2020). Racial bias in pulse oximetry measurement. *New England Journal of Medicine*, *383*(25), 2477–2478. https://doi.org/10.1056/nejmc2029240

Smith, G., Czerwinski, M., Meyers, B., Robbins, D., Robertson, G., & Tan, D. S. (2006). Facetmap: A scalable search and browse visualization. *IEEE Transactions on Visualization and Computer Graphics*, *12*(5), 797–804. https://doi.org/10.1109/TVCG.2006.142

Sparrow, R., Howard, M., & Degeling, C. (2021). Managing the risks of artificial intelligence in agriculture. *NJAS: Impact in Agricultural and Life Sciences*, *93*(1), 172–196.

Sparrow, R., & Howard, M. (2021). Robots in agriculture: Prospects, impacts, ethics, and policy. *Precision Agriculture*, *22*, 818–833. https://doi.org/10.1007/s11119-020-09757-9

Sriram, N., & Philip, H. (2016). Expert system for decision support in agriculture. *TNAU Agritech*.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, *15*(1), 1929–1958.

Stinson, C. (2022). Algorithms are not neutral: Bias in collaborative filtering. *AI and Ethics*, *2*(4), 763–770. https://doi.org/10.1007/s43681-022-00136-w

Sun, K. H., Huh, H., Tama, B. A., Lee, S. Y., Jung, J. H., & Lee, S. (2020). Vision-based fault diagnostics using explainable deep learning with class activation maps. *IEEE Access*, *8*, 129169–129179. https://doi.org/10.1109/ACCESS.2020.3009852

Suresh, H., & Guttag, J. (2021). A framework for understanding sources of harm throughout the machine learning life cycle. *Equity and Access in Algorithms, Mechanisms, and Optimization*, 1–9.

Syerov, Y., Shakhovska, N., & Fedushko, S. (2020). Method of the data adequacy determination of personal medical profiles. *Advances in Artificial Systems for Medicine and Education*, *Ii*(2), 333–343.

Sypherd, T., Nock, R., & Sankar, L. (2021). Being properly improper. arXiv preprint arXiv:2106.09920. https://doi.org/10.48550/arXiv.2106.09920

Taha, A. A., & Hanbury, A. (2015). Metrics for evaluating 3d medical image segmentation: Analysis, selection, and tool. *BMC Medical Imaging*, *15*(1), 1–28. https://doi.org/10.1186/s12880-015-0068-x

Tamimi, A. F., & Juweid, M. (2017). Epidemiology and outcome of glioblastoma. *Exon Publications*, 143–153. https://doi.org/10.15586/codon.glioblastoma.2017.ch8

Tarrant, M., & North, A. C. (2004). Explanations for positive and negative behavior: The intergroup attribution bias in achieved groups. *Current Psychology*, *23*(2). https://doi.org/10.1007/BF02903076

Thomas, D. M., Kleinberg, S., Brown, A. W., Crow, M., Bastian, N. D., Reisweber, N., Lasater, R., Kendall, T., Shafto, P., Blaine, R., Smith, S., Ruiz, D., Morrell, C., & Clark, N. (2022). Machine learning modeling practices to support the principles of ai and ethics in nutrition research. *Nutrition & Diabetes*, *12*(1), 48.

Tzovaras, G., et al. (2019). Open humans: A platform for participant-centered research and personal data exploration. *GigaScience*, *8*(6), giz076. https://doi.org/10.1093/gigascience/giz076

van Giffen, B., Herhausen, D., & Fahse, T. (2022). Overcoming the pitfalls and perils of algorithms: A classification of machine learning biases and mitigation methods. *Journal of Business Research*, *144*, 93–106. https://doi.org/10.1016/j.jbusres.2022.01.076

Vassiliades, A., Bassiliades, N., & Patkos, T. (2021). Argumentation and explainable artificial intelligence: A survey. *The Knowledge Engineering Review*, *36*, e5.

Verma, A., Murali, V., Singh, R., Kohli, P., & Chaudhuri, S. (2018). Programmatically interpretable reinforcement learning. *International conference on machine learning* (pp. 5045–5054).

Vieth, K., & Bronowicka, J. (2017). *Ethics of algorithms*. Center for Internet and Human Rights. Retrieved Aug 12, 2023 from https://cihr.eu/eoa2015web/

Wang, R., Jia, X., Wang, Q., Wu, Y., & Meng, D. (2022). Imbalanced semi-supervised learning with bias adaptive classifier. *The eleventh international conference on learning representations*.

Wang, Z. J., Kale, A., Nori, H., Stella, P., Nunnally, M., Chau, D. H., & Caruana, R. (2021). Gam changer: Editing generalized additive models with interactive visualization. arXiv preprint arXiv:2112.03245. https://doi.org/10.48550/arXiv.2112.03245

Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2016). A survey of transfer learning. *Journal of Big Data*, *3*, 1–40.

Weng, C. G., & Poon, J. (2008). A new evaluation measure for imbalanced datasets. *Proceedings of the 7th Australasian Data Mining Conference-Volume*, *87*, 27–32.

Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining* (Vol. 1, pp. 29–39).

Wong, E. (2018). Self configuration in machine learning. arXiv preprint arXiv:1809.06463. https://doi.org/10.48550/arXiv.1809.06463

WorldBank. (2022). *Sri lanka development update: Protecting the poor and vulnerable in a time of crisis*. World Bank.

Xiong, H., Dalhaus, T., Wang, P., & Huang, J. (2020). Blockchain technology for agriculture: Applications and rationale. *Frontiers in Blockchain*, *3*, 7. https://doi.org/10.3389/fbloc.2020.00007

Yang, J., Soltan, A. A., Eyre, D. W., & Clifton, D. A. (2023). Algorithmic fairness and bias mitigation for clinical machine learning with deep reinforcement learning. *Nature Machine Intelligence*, *5*(8), 884–894.

Yao, Q., Wang, M., Chen, Y., Dai, W., Hu, Y.-Q., Li, Y.-F., Tu, -W.-W., Yang, Q., & Yu, Y. (2018). Taking human out of learning applications: A survey on automated machine learning. arXiv preprint arXiv:1810.13306. https://doi.org/10.48550/arXiv.1810.13306

Zhang, K., Khosravi, B., Vahdati, S., Faghani, S., Nugen, F., Rassoulinejad-Mousavi, S. M., Moassefi, M., Jagtap, J. M. M., Singh, Y., Rouzrokh, P., & Erickson, B. J. (2022). Mitigating bias in radiology machine learning: 2. Model development. *Radiology: Artificial Intelligence*, *4*(5), e220010. https://doi.org/10.1148/ryai.220010

Zhou, H., Wang, X., Au, W., Kang, H., & Chen, C. (2022). Intelligent robots for fruit harvesting: Recent developments and future challenges. *Precision Agriculture*, *23*(5), 1856–1907. https://doi.org/10.1007/s11119-022-09913-3

Zossou, E., Arouna, A., Diagne, A., & Agboh-Noameshie, R. A. (2020). Learning agriculture in rural areas: The drivers of knowledge acquisition and farming practices by rice farmers in West Africa. *The Journal of Agricultural Education and Extension*, *26*(3), 291–306. https://doi.org/10.1080/1389224X.2019.1702066

Authors and Affiliations

Mathuranathan Mayuravaani[1] · Amirthalingam Ramanan[1] · Maneesha Perera[2] · Damith Asanka Senanayake[2] · Rajith Vidanaarachchi[2]

✉ Mathuranathan Mayuravaani
mayu@univ.jfn.ac.lk

Amirthalingam Ramanan
a.ramanan@univ.jfn.ac.lk

Maneesha Perera
maneesha.perera1@unimelb.edu.au

Damith Asanka Senanayake
damith.senanayake@unimelb.edu.au

Rajith Vidanaarachchi
rajith.vidanaarachchi@unimelb.edu.au

[1] Department of Computer Science, Faculty of Science, University of Jaffna, Thirunelvely, Jaffna, Sri Lanka

[2] Department of Mechanical Engineering, Faculty of Engineering and IT, University of Melbourne, Melbourne, Australia