Open Access Research Journal of
Engineering and Technology
Journals home page: https://oarjpublication/journals/oarjet/
ISSN: 2783-0128 (Online)

OARJ OPEN ACCESS RESEARCH JOURNALS

(RESEARCH ARTICLE)

Algorithmic bias reduction strategies in machine learning

Ransford Antwi, Moses Garuba * and Ikemefuna Uba

Department of Electrical and Computer Engineering, Hampton University, Hampton, Virginia, United States.

Open Access Research Journal of Engineering and Technology, 2026, 10(02), 170-177

Publication history: Received on 24 April 2026; revised on 01 June 2026; accepted on 03 June 2026

Article DOI: https://doi.org/10.53022/oarjet.2026.10.2.0046

Abstract

As Artificial Intelligence (AI) and Machine Learning (ML) systems are increasingly integrated into critical societal domains such as finance, healthcare, and autonomous infrastructure, the prevalence of algorithmic bias has emerged as a paramount global concern. This research addresses the lack of a standardized, comprehensive understanding of how systematic deviations enter the AI lifecycle, leading to social harm, eroded public trust, and significant legal risks. Algorithmic bias is defined as systematic, non-random errors in machine learning algorithms that generate outcomes disproportionately favoring or disfavoring only one side. At the societal level, Fink's Seven-Phase Issue-Development Process is utilized to monitor the evolution of algorithmic failures—such as gender disparities in search results and racial inaccuracies in computer vision—before they escalate into public crises. At the technical level, the research evaluates the efficacy of selecting best-suited estimators and optimizers. Empirical comparisons conducted via Python-based benchmarking demonstrate that architectures such as Convolutional Long Short-Term Memory (CLSTM) and Extra Trees (ET) significantly reduce systematic deviation in imbalanced datasets compared to sub-optimal models like Multinomial Naive Bayes (MNB) or Gradient Boosted Regression Trees (GBRT). The findings underscore that effective bias reduction requires a dual-track strategy which includes aligning technical model selection with the best data system and implementing ethical auditing throughout the development pipeline. This research provides a robust framework for training data best practices and the need to use the best estimators, optimizers, and regressors in machine learning.

Keywords: Algorithm Bias; Machine Learning; Taxonomy Development; Framework Integration; Technical Benchmarking; Systematic Mitigation; Optimizers; Estimators; Decision Tree Regressor; Linear Regression

1. Introduction

Bias is fundamentally defined as a consistent deviation from what is considered normal or standard. Algorithmic bias occurs when this deviation is encoded within a computer program, particularly in autonomous systems powered by Machine Learning (ML). As stated by Jonker and Rogers [1], it often reflects or reinforces existing gender biases, particularly in occupational and financial domains. The central problem addressed by this paper is the lack of a standardized, comprehensive understanding of where, when, and how these systematic errors enter the AI lifecycle. As AI systems utilize complex algorithms to discover patterns and predict outputs, their pervasive integration into critical domains such as hiring, criminal justice, finance, and healthcare has made the integrity of their underlying logic a paramount global concern [2]. The central problem is the lack of a standardized, comprehensive understanding of how systematic errors enter the AI lifecycle across various demographic and operational factors. Biased algorithms impact system outputs in ways that lead to harmful decisions, promote discrimination, and erode trust in the institutions that deploy them.

Consequently, this research aims to answer the question: 'What are the distinct forms through which algorithmic bias is introduced and perpetuated by machine learning systems, and what specific strategies can be implemented to identify and mitigate these biases?' [3] This study is significant because the consequences of algorithmic bias create tangible social and economic harm, such as disproportionately denying financial opportunities to marginalized groups or

* Corresponding author: Moses Garuba

misclassifying critical medical data. By identifying ways to reduce these biases, this research fills the gap in existing mitigation frameworks and helps organizations navigate the substantial legal and financial risks associated with biased automated decision-making.

The structure of this paper proceeds as follows. Section 2 presents a literature review covering the origins, manifestations, and societal consequences of algorithmic bias as documented in prior research. Section 3 outlines the mitigation and prevention strategies, with a focus on Fink's Seven-Phase Issue-Development Process as a societal monitoring framework, and on principled estimator and optimizer selection as technical countermeasures. Section 4 details the methodology, including the datasets, models, and experimental procedure employed. Section 5 reports the empirical results, and Section 6 provides a comprehensive discussion of findings. Section 7 concludes with broader implications for AI development and deployment.

2. Literature Review

2.1. Origins and Manifestations of Bias

Existing research highlights that algorithmic bias often originates from biased training data and model design [4]. Technical manifestations are frequently seen in computer vision, where software struggles with diversity. For instance, the 2009 Hewlett-Packard (HP) webcam controversy [5] revealed that facial-tracking algorithms, trained on datasets dominated by lighter-skinned faces, were unable to recognize Black users. Similarly, the 2010 Nikon Blink Warning incident [6] demonstrated how a Western-centric morphological standard caused software to inaccurately identify Asian individuals as blinking when their eyes were open. These cases show that when data is not curated for diversity, the resulting technology is naturally discriminatory.

Beyond technical failure, algorithmic bias reinforces societal stereotypes. A 2015 University of Washington study found that Google's CEO image search results were significantly skewed: while 27% of US CEOs were women at the time, only 11% of the top search results featured them [7]. This creates a feedback loop that shapes public perception of leadership. Kate Crawford's 2016 work, Artificial Intelligence's White Guy Problem [8], synthesizes these issues as systemic, arguing that AI reflects the biases of its creators—often an overwhelmingly white and male demographic—meaning bad data is simply a reflection of existing social inequalities. Her opinion piece in the New York Times in 2016 states: 'Sexism, racism, and other forms of discrimination are being built into the machine learning algorithms that underlie technology behind many intelligent systems that shape how we are categorized and advertised to. We need to be vigilant about how we design and train these machine-learning systems, or we will see ingrained forms of bias built into the artificial intelligence of the future.'

Like all technologies before it, artificial intelligence will reflect the values of its creators and builders. Inclusivity matters from the company or person who designs it, to who sits on the company boards, and which ethical perspectives are included. Otherwise, we risk constructing machine intelligence that mirrors a narrow and privileged vision of society, with its old, familiar biases and stereotypes.

While previous work has identified these failures, there remains a gap in establishing a standardized trend or pattern for mitigation. Current literature and journals establish four primary sources of bias: training data, algorithm design, proxy data, and evaluation [9]. This study contributes to the field by proposing a combination of methods—integrating ethical auditing, multi-metric fairness evaluation, and diverse data curation—to carefully choose the best model for data, addressing these interconnected failure points simultaneously rather than in isolation.

2.2. Bias in Criminal Justice and Healthcare

The pervasive impact of algorithmic bias extends well beyond technology products into life-altering institutional decisions. In the domain of criminal justice, risk assessment tools such as the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) have come under sustained scrutiny. Investigative reporting demonstrated that the system assigned significantly higher recidivism risk scores to Black defendants than to their white counterparts when controlling for actual subsequent criminal behavior. This case illustrates that when historical incarceration data—itself shaped by racially unequal policing practices—is used as ground truth for training, the resulting model inherits and amplifies those disparities rather than correcting for them.

Similarly, the healthcare domain is not immune. A widely cited study published in Science revealed that an algorithm used by major US health systems to allocate care management resources exhibited significant racial bias. The system used health costs as a proxy for health needs; however, because Black patients historically incur lower costs for

comparable levels of illness due to systemic barriers in healthcare access, the model systematically identified healthier Black patients as needing the same level of care as sicker white patients. The authors estimated that the algorithm reduced the number of Black patients identified as needing extra care by more than half. These cases underscore that algorithmic bias is not a purely technical artifact but a societal problem that requires both technical and institutional remediation.

2.3. Legal and Regulatory Landscape

The growing recognition of algorithmic harm has prompted regulatory scrutiny worldwide. In the European Union, the General Data Protection Regulation (GDPR) grants individuals the right not to be subject to decisions based solely on automated processing that produce significant effects, and mandates transparency in algorithmic decision-making. The EU's Artificial Intelligence Act, proposed in 2021 and advancing toward implementation, establishes a risk-based framework that classifies high-stakes AI applications in employment, education, and law enforcement as high-risk systems subject to strict conformity requirements.

In the United States, the Equal Credit Opportunity Act and the Fair Housing Act impose anti-discrimination obligations that extend to algorithmic credit and housing decisions. The Federal Trade Commission has increasingly signaled scrutiny of AI systems that produce discriminatory outcomes, while several states and municipalities have enacted more targeted legislation, such as New York City's Local Law 144, which requires bias audits for automated employment decision tools. This regulatory evolution underscores that addressing algorithmic bias is not merely an ethical aspiration but a concrete legal obligation, reinforcing the urgency of the technical and organizational frameworks proposed in this research.

2.4. Mitigation and Prevention Strategies

2.4.1. Fink's Seven-Phase Issue-Development Process

The primary methodology employed in this research is Fink's Seven-Phase Issue-Development Process framework [10]. This model is utilized to present the evolution of AI-related issues as they manifest at a societal level within the United States, particularly concerning machine learning algorithms. By mapping the progression of an issue from its latent stage to a crisis, this study aims to provide a structured timeline for anticipating and preventing future harm to the public. This proactive mapping is critical because, as seen in the Google search disparities and the HP webcam controversy, algorithmic issues often move rapidly from technical oversights to widespread social crises.

2.5. Estimator Selection

The term estimator [11] refers to a rule, formula, or algorithm used in the prediction of an unknown population parameter based on observed data. While standard estimators are designed solely to maximize accuracy, this research proposes the use of fairness-aware estimators to address biases in algorithm design. By selecting models that prioritize algorithmic transparency and interpretability, developers can identify the systematic deviations that lead to the misclassification of classes such as facial recognition, genders, and races. For instance, employing constrained classification models as estimators ensures that the decision boundaries do not rely on proxy variables that correlate with protected characteristics, such as race or gender.

2.6. Optimizer Selection

The mitigation strategy further integrates specialized optimizers that move beyond simple loss minimization. Standard optimizers often prioritize the majority demographic in a dataset, which Crawford identified as a primary driver of the White Guy Problem [12]. To resolve this, the proposed method utilizes optimizers that incorporate fairness constraints (such as demographic parity or equalized odds) directly into the objective function. This ensures that the model does not learn and reinforce existing socioeconomic inequalities, such as the historical gender imbalances found in Google's CEO image search results. By penalizing the model for high error rates on marginalized subgroups during the training phase, these optimizers prevent the feedback loop that typically results in discriminatory outcomes.

2.7. Integration into the AI Lifecycle

Integrating these technical selections ensures that fairness is not an afterthought but a fundamental part of the AI development lifecycle. When combined with the auditing stages of Fink's Seven-Phase Process, the use of best-suited estimators and optimizers provides a robust defense against the legal and financial risks associated with biased systems. This dual approach ensures that technical accuracy (choosing the best estimator) and social equity (mitigating bias) are pursued simultaneously.

2.8. Diverse Data Curation and Pre-Processing

Before estimator or optimizer selection can be effective, the quality and representativeness of training data must be assessed. This research advocates for a structured data curation process that includes demographic auditing of datasets to identify underrepresented groups, application of resampling techniques such as Synthetic Minority Over-sampling Technique (SMOTE) to address class imbalance, and the use of fairness-aware feature engineering to remove or transform proxy variables that encode protected characteristics.

Pre-processing interventions are particularly valuable in regulated domains. In hiring algorithms, for instance, removing variables such as zip code or educational institution from training data can reduce proxy discrimination without sacrificing significant predictive power. By combining these data-level interventions with the algorithmic-level fairness constraints described above, organizations can construct a more resilient, multi-layered defense against the introduction of bias at any stage of the AI pipeline.

2.9. Post-Processing and Fairness Auditing

Even after deployment, continuous monitoring is essential. Post-processing techniques such as threshold calibration can be applied to model outputs to ensure that error rates are equalized across demographic subgroups. Fairness auditing—using metrics such as demographic parity difference, equalized odds, and disparate impact ratio—provides operational definitions against which deployed systems can be assessed on a regular basis.

3. Methodology

To ensure the integrity of the findings, a variety of analytical tools and algorithms are employed to benchmark performance against fairness:

Estimators: The study compares simple models like Multinomial Naive Bayes (MNB) and Linear Regression against more sophisticated architectures such as Convolutional Long Short-Term Memory (CLSTM) Neural Networks, Extra Trees (ET), and Decision Tree Regressors.

Optimizers: Specialized optimizers are utilized to incorporate fairness constraints directly into the objective function, preventing the feedback loop that reinforces societal stereotypes.

Frameworks: Fink's Seven-Phase Issue-Development Process is used to monitor the evolution of these technical errors as they move into the public domain.

3.1. Procedure

The experiment follows a systematic four-step process to transition from data gathering to bias mitigation:

- Utilizing Fink's framework, latent algorithmic issues in current systems are identified (e.g., gender bias in search results or skin-tone recognition failures).
- In confluence with Fink's framework, models are trained on both linear and complex datasets using Python libraries such as NumPy and Matplotlib.
- A linear model of the form $y = 2x + 3 + \varepsilon$ was created and a complex model of the form $y = \sin(x) + \varepsilon$ was constructed to obtain a sample set of data for training.
- Predicted vs. True data values are compared to calculate the Systematic Deviation. For instance, testing whether a Decision Tree Regressor overfits the complex data compared to a Linear Regression model.
- Evaluation of the data fitted on the graph is conducted at this stage. A linear regression model was used to train a complex dataset and vice versa.

The study also evaluates the behavior of models when applied to mismatched datasets—i.e., training a model designed for one distributional shape on data generated from a different underlying process. This deliberate mismatching reproduces conditions analogous to real-world deployment scenarios in which a model trained on a majority-group distribution is then applied to minority subgroups with systematically different data distributions. The degree of systematic deviation observed in these cross-trained configurations provides a quantitative measure of inductive bias and its potential for producing inequitable outcomes.

4. Results

Table 1 shows a complete list of comparative studies of ML algorithms (estimators).

Table 1 Comparative Studies of ML Algorithms (Estimators)

ID	Year	Ref.	Domain	Data Category	Estimators	Most Suitable Estimator
CS1	2022	[13]	Agriculture	Spectra Data	RF, SVM, Cubist	Ensemble Method
CS2	2016	[14]	Biology	Numeric	SVM, KNN, DT	SVM
CS3	2020	[15]	Business	Text (Reviews)	MNB, CLSTM	CLSTM
CS4	2015	[16]	Computer Vision	Remote Sensing	RF, ET, GBRT	ET

Table 2 shows the estimator names and their corresponding descriptions.

Table 2 Estimator Names and Descriptions

Algorithm (Estimator) Name	Abbreviation	Description
Random Forest	RF	Fundamental ML algorithm that makes use of the output of multiple decision trees for classification, regression and other tasks [17]
Support Vector Machines	SVM	Classification algorithm that finds an optimal line or hyperplane to maximize the distance between each class [18]
Cubist	Cubist	Rule-based model that contains a tree whose final leaves use linear regression models [19]
k-Nearest Neighbors	KNN	A fundamental ML algorithm used for classification [20]
Decision Trees	DT	Fundamental ML algorithm that uses a tree-like model of decisions and can be used for both regression and classification tasks [21]
Multinomial Naive Bayes	MNB	Probabilistic classifier based on Bayes' theorem that focuses on calculation of text data's distribution [22]
Convolutional Long Short-Term Memory	CLSTM	Type of LSTM Neural Network that also utilizes a convolutional layer [23]
Extra Trees	ET	An ensemble ML approach that utilizes multiple randomized decision trees [24]
Gradient Boosted Regression Trees	GBRT	Type of additive model that makes predictions by combining decisions from a set of other models [25]

Figure 1 shows true linear data fitted with a linear regressor model. Figure 2 shows true complex data fitted with a linear model. Figure 3 shows true linear data fitted with a well-trained linear regression model. Figure 4 shows true complex data fitted with a well-trained Decision Tree regression model.

The visual representations provided in Figures 1 through 4 collectively illustrate the central quantitative finding of this study: systematic deviation is minimized when model complexity is matched to data complexity. When a linear model is applied to linear data (Figure 1 and Figure 3), the predicted values closely track the true values, resulting in low residuals and minimal systematic bias. Conversely, when the same linear model is applied to sinusoidal (complex) data (Figure 2), a pronounced systematic deviation is observed: the predicted line fails to capture the oscillatory nature of the data, producing large and patterned residuals. This pattern of failure, whereby a model's structural assumptions are inconsistent with the data-generating process, constitutes inductive bias.

Figure 4 demonstrates the corrective power of complexity matching: when the Decision Tree Regressor is applied to complex data, the fitted curve closely follows the sinusoidal fluctuations, substantially reducing systematic deviation.

These results have direct implications for the study of algorithmic fairness. When a model is trained primarily on data from one demographic group and applied to another group whose data follows a different distribution, the result is analogous to Figure 2—systematic deviation that disproportionately affects the underrepresented group.

5. Discussion

In Table 1, the authors in [13] made use of three algorithms to estimate soil organic carbon; the Ensemble method was the most suitable algorithm since the results obtained from the experiment were precise and less biased. From the authors in [14], SVM was chosen over KNN and DT since it yielded superior results with biological data used to predict bacterial heterotrophic fluxomics. Moreover, the necessity of selecting the best estimator is particularly evident in the business domain. For instance, when classifying client reviews, empirical evidence shows that CLSTM Neural Networks are more appropriate than MNB. While MNB often produces biased results by over-learning dominant classes in imbalanced datasets, the CLSTM architecture serves as a superior estimator that prevents the neglect of minority data points, thereby mitigating the risk of exclusionary insights [15]. Lastly, research evaluating computer vision applications in remote sensing compared the effectiveness of Random Forest (RF), Extra Trees (ET), and Gradient Boosted Regression Trees (GBRT). The findings established that Extra Trees (ET) emerged as the most accurate estimator for this domain, while Gradient Boosted Regression Trees (GBRT) produced the least accurate results [16].

5.1. Model Suitability and the Mitigation of Inductive Bias

The provided simulation illustrates a critical finding in bias mitigation: the selection of an estimator must be dictated by the underlying distribution of the data. In the experiment, two distinct models—Linear Regression and a Decision Tree Regressor—were tested against linear and complex (sinusoidal) datasets. The results demonstrate that Linear Regression serves as the best estimator for linear data but fails significantly when applied to complex patterns. This failure is a form of inductive bias, where the model's inherent simplicity prevents it from capturing the true nature of the data, leading to systematic underperformance that could, in real-world applications, result in the exclusion of marginalized groups.

5.2. Linear vs. Complex Architectures

When linear trends are applied to the linear dataset, the Linear Regression model achieves high fidelity. Using a complex model here would be unnecessary and might introduce high variance. On the other hand, when faced with complex data, the Linear Regression model produces a straight line that ignores the periodic fluctuations of the true data. This systematic deviation is exactly what Crawford (2016) identifies as a precursor to broader algorithmic failures. Hence, mitigation through complexity matching using the Decision Tree Regressor acts as a more flexible estimator. It is capable of approximating the non-linear nests or patterns within the complex dataset. This demonstrates that mitigating bias often requires transitioning from rigid, one-size-fits-all algorithms to architectures that can adapt to the nuances of diverse datasets.

6. Conclusion

These experiments demonstrate that mitigating algorithmic bias is not merely a matter of better data, but of technical alignment. By utilizing Python-based benchmarking to compare estimators—similar to the comparison of CLSTM vs. MNB in business reviews or ET vs. GBRT in remote sensing—researchers can identify which models are prone to neglecting minority classes or complex patterns. Integrating these best estimators with fairness-aware optimizers constitutes a robust technical defense against algorithmic bias, ensuring that AI systems are both accurate and equitable across all segments of the public.

The simulation further reinforces that estimator selection is inseparable from bias mitigation. Linear Regression performs optimally on linear datasets but fails on complex, non-linear patterns, illustrating a classic inductive bias in which overly simplistic models systematically misrepresent reality. Such misalignment mirrors broader algorithmic failures identified in the literature, where rigid models overlook structural nuances and inadvertently propagate exclusionary outcomes. Conversely, the Decision Tree Regressor adapts to non-linear fluctuations, demonstrating that flexible architectures can better capture heterogeneous patterns and reduce systematic error.

Collectively, these findings affirm that mitigating algorithmic bias requires deliberate alignment between model complexity and data complexity. Estimators must be chosen not for convenience or convention, but for their capacity to faithfully represent diverse data distributions, thereby ensuring fairness, accuracy, and inclusivity in real-world decision-making systems.

The societal monitoring component of this research, grounded in Fink's Seven-Phase Issue-Development Process, provides a complementary framework that organizations can use to track the evolution of bias-related incidents before they become public crises. Historical cases such as the HP webcam controversy, the Nikon blink detection failure, and the gender disparities in Google's executive image search results all followed recognizable trajectories: an initial technical oversight in data curation or model design that, without active monitoring and intervention, escalated into reputational and legal consequences. By applying Fink's framework proactively, organizations can intervene at the latent or emerging stages of an issue, before media attention and regulatory action force reactive remediation.

Looking forward, the field of algorithmic fairness must grapple with the fundamental tension between model accuracy and distributional equity. In many real-world settings, these objectives are not fully aligned, and optimizing for one may come at a cost to the other. Future research should focus on developing optimization frameworks that characterize and navigate this trade-off transparently, enabling policymakers and practitioners to make informed decisions about acceptable levels of systematic deviation in high-stakes applications.

In conclusion, this research demonstrates that algorithmic bias is neither inevitable nor intractable. It is the product of specific, identifiable choices made at each stage of the AI lifecycle—in data collection, model selection, optimizer design, and deployment monitoring. By making these choices deliberately, with fairness as an explicit objective alongside accuracy, developers and organizations can build AI systems that serve all segments of society equitably. The frameworks and findings presented in this paper provide a concrete starting point for that endeavor.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] A. Jonker and J. Rogers, "Algorithmic bias." IBM. [Online]. Available: https://www.ibm.com/think/topics/algorithmic-bias

[2] A. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, "A survey on bias and fairness in machine learning," IEEE Trans. Knowl. Data Eng., vol. 34, no. 6, pp. 218-246, 2022.

[3] S. Barocas, M. Hardt, and A. Narayanan, Fairness and Machine Learning, NIPS Tutorial, 2017.

[4] S. Hajian, F. Bonchi, and C. Castillo, "Algorithmic bias: From discrimination discovery to fairness-aware data mining," IEEE Data Eng. Bull., vol. 36, no. 4, pp. 55-64, 2013.

[5] S. Ferguson, "HP investigating alleged racist facial-tracking software," eWEEK, Dec. 22, 2009. [Online]. Available: https://www.eweek.com/pc-hardware/hp-investigating-alleged-racist-facial-tracking-software/

[6] A. Rose, "Are face-detection cameras racist?" Time, Jan. 22, 2010. [Online]. Available: https://time.com/archive/6906847/are-face-detection-cameras-racist/

[7] E. Handley, "Google's image search gender bias shows underrepresentation of women," Open Access Government, Feb. 25, 2022. [Online]. Available: https://www.openaccessgovernment.org/image-search-gender-bias/130353/

[8] K. Crawford, "Artificial intelligence's white guy problem," The New York Times, Jun. 26, 2016. [Online]. Available: https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html

[9] M. Feldman, S. A. Friedler, J. Moeller, C. Scheidegger, and S. Venkatasubramanian, "Certifying and removing disparate impact," in Proc. IEEE Int. Conf. Big Data, pp. 243-252, 2015.

[10] A. Yapo and J. Weiss, "Ethical implications of bias in machine learning," in Proc. 51st Hawaii Int. Conf. Syst. Sci., 2018, doi: 10.24251/hicss.2018.668.

[11] B. Efron and T. Hastie, Computer Age Statistical Inference: Algorithms, Evidence, and Data Science. Cambridge, U.K.: Cambridge Univ. Press, 2016.

[12] IEEE Standard for Algorithmic Bias Considerations, IEEE Std 7003-2024, 2024.

[13] J. K. M. Biney et al., "Prediction of topsoil organic carbon content with Sentinel-2 imagery," Soil Tillage Res., vol. 220, p. 105379, 2022, doi: 10.1016/j.still.2022.105379.

[14] S. G. Wu et al., "Rapid prediction of bacterial heterotrophic fluxomics," PLoS Comput. Biol., vol. 12, no. 8, p. e1004838, 2016, doi: 10.1371/journal.pcbi.1004838.

[15] A. Rafay, M. Suleman, and A. Alim, "Robust review rating prediction model," in Proc. 2020 Int. Conf. Emerging Trends Smart Technol. (ICETST), pp. 8138-8143, 2020, doi: 10.1109/ICETST49965.2020.9080724.

[16] A. Merentitis and C. Debes, "Many hands make light work—On ensemble learning," IEEE Geosci. Remote Sens. Mag., vol. 3, no. 3, pp. 86-99, 2015, doi: 10.1109/MGRS.2015.2434354.

[17] S. J. Rigatti, "Random forest," J. Insurance Med., vol. 47, no. 1, pp. 31-39, 2017, doi: 10.17849/insm-47-01-31-39.1.

[18] H. Wang and D. Hu, "Comparison of SVM and LS-SVM for regression," in Proc. 2005 Int. Conf. Neural Netw. Brain, vol. 1, pp. 279-283, 2005, doi: 10.1109/ICNNB.2005.1614615.

[19] M. Kuhn, S. Weston, C. Keefer, and N. Coulter, Cubist models for regression (Version 0.0), R package vignette, 2012.

[20] T. Cover and P. Hart, "Nearest neighbor pattern classification," IEEE Trans. Inf. Theory, vol. 13, no. 1, pp. 21-27, 1967.

[21] Y. Y. Song and Y. Lu, "Decision tree methods: Applications for classification and prediction," Shanghai Arch. Psychiatry, vol. 27, no. 2, pp. 130-135, 2015, doi: 10.11919/j.issn.1002-0829.215044.

[22] M. Abbas, K. A. Memon, A. A. Jamali, S. Memon, and A. Ahmed, "Multinomial Naive Bayes classification model for sentiment analysis," IJCSNS Int. J. Comput. Sci. Netw. Secur., vol. 19, no. 3, pp. 62-71, 2019.

[23] Mustaqeem and S. Kwon, "CLSTM: Deep feature-based speech emotion recognition," Mathematics, vol. 8, no. 12, p. 2133, 2020, doi: 10.3390/math8122133.

[24] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," Mach. Learn., vol. 63, no. 1, pp. 3-42, 2006, doi: 10.1007/s10994-006-6226-1.

[25] P. Prettenhofer and G. Louppe, "Gradient boosted regression trees in scikit-learn," in Proc. PyData London 2014, London, U.K., Feb. 2014.