Fairness-aware Machine Learning for Public Sector Resource Allocation

Yang Nan
School of Intelligence Science and Technology, Nanjing University, China

Xu Ke
School of Intelligence Science and Technology, Nanjing University, China

Submitted on: March 22, 2021
Accepted on: May 10, 2021
Published on: July 05, 2021
DOI: 10.5281/zenodo.18049922

***Abstract*—Public sector organizations increasingly employ machine learning to guide the allocation of limited resources across populations and regions. While such systems offer gains in efficiency and scale, they also risk reinforcing historical inequities embedded in administrative data. This study presents a fairness-aware machine learning framework for public sector resource allocation that balances predictive performance with equitable outcomes. The proposed approach integrates fairness constraints, interpretable modeling, and post allocation auditing within a unified decision support pipeline. Empirical evaluation demonstrates that fairness-aware optimization can substantially reduce allocation disparities while preserving operational effectiveness. The results provide practical guidance for deploying trustworthy machine learning systems in high impact public decision contexts.**

***Index Terms*—Fairness-aware learning, public sector AI, resource allocation, ethical machine learning, decision support systems**

I. Introduction

Public institutions face persistent challenges in distributing scarce resources such as healthcare funding, social assistance, educational support, and emergency services. These decisions often involve competing objectives that include efficiency, coverage, transparency, and fairness. Machine learning has emerged as a valuable tool for supporting such decisions by identifying patterns in large administrative datasets and forecasting demand more accurately.

Despite these benefits, the adoption of machine learning in public sector decision-making raises significant concerns. Models trained on historical data may encode structural biases related to geography, income, ethnicity, or access to services. When deployed at scale, such systems risk amplifying inequities rather than mitigating them. As public decisions directly affect citizen well-being, fairness is not an optional property but a core requirement.

Recent research has emphasized the importance of ethical and trustworthy artificial intelligence, particularly in high-stakes domains. However, translating fairness principles into operational allocation systems remains challenging. Public agencies require models that are not only accurate but also interpretable, auditable, and aligned with policy objectives.

This paper addresses these challenges by proposing a fairness-aware machine learning framework tailored to public sector resource allocation. The framework integrates predictive modeling with explicit fairness constraints and post allocation auditing. Rather than enforcing rigid equality, the approach supports balanced trade-offs between efficiency and equity, allowing decision-makers to remain accountable while benefiting from data-driven insights.

The main contributions of this work are threefold. First, it formulates public resource allocation as a constrained optimization problem that incorporates group fairness objectives. Second, it presents a modular system architecture that supports transparency and oversight. Third, it provides empirical results demonstrating that fairness-aware learning can reduce disparity without significant loss of performance.

II. Related Work and Literature Review

This section reviews prior research relevant to fairness-aware allocation, drawing from decision support systems, ethical AI, applied machine learning, and optimization.

A. Machine Learning for Decision Support

Early decision support systems relied on rule-based logic and expert knowledge. Hybrid approaches later combined learned rules with structured reasoning, enabling adaptive behavior in complex environments [1]. Verification and online updating of control logic improved system reliability and trust [2].

Optimization-based decision models have been widely used in operational planning and control. Population-based algorithms and multi-objective optimization techniques allow trade-offs among competing goals [3], [4]. These approaches inform the design of allocation systems that must balance efficiency with fairness.

B. Ethical and Trustworthy AI

Ethical concerns in artificial intelligence span governance, accountability, and bias mitigation. Analyses of ethical AI across sectors highlight persistent challenges in enforcing fairness through technical means alone [5]. Reliability testing and validation methods are essential for ensuring responsible deployment [6].

Trust modeling has also gained attention, particularly in distributed and fog-based systems where decision authority is decentralized [7]. These concepts are applicable to public sector AI, where trust and legitimacy are critical.

C. Applied Machine Learning in Societal Domains

Machine learning has been successfully applied to domains analogous to public resource allocation. Healthcare analytics demonstrate how predictive models can support policy and planning decisions [8]. Environmental monitoring and disaster prediction systems illustrate the value of early warning and equitable response [9], [10].

Infrastructure and transportation studies further highlight the need for robust and scalable models in public systems [11], [12]. These applications emphasize generalization and resilience under real-world conditions.

D. Learning Architectures and Optimization

Advances in deep learning have driven improvements across vision, language, and time-series tasks [13]–[15]. Efficiency has been enhanced through hardware-aware optimization and edge computing strategies [16], [17].

Transfer learning and ensemble methods improve adaptability across heterogeneous datasets [18], [19]. Such flexibility is valuable for allocation models spanning diverse regions and populations.

E. Interpretability and Knowledge Representation

Interpretability remains a key requirement for public sector AI. Formal knowledge representation frameworks support explainable reasoning [20]. Fuzzy inference and uncertainty-aware models enable transparent handling of ambiguous information [21], [22].

These techniques inform the interpretability layer of fairness-aware allocation systems, enabling human oversight.

III. Methodology

A. Problem Definition

Let $X \in \mathbb{R}^{n \times d}$ represent features describing n regions or population groups, and let $y \in \mathbb{R}^n$ denote observed need or demand. The objective is to allocate a limited budget B across entities.

Allocation decisions a_i must satisfy:

$$\sum_{i=1}^{n} a_i \leq B \tag{1}$$

B. Fairness-aware Objective

The predictive model produces scores $\hat{y}_i = f(X_i)$. A fairness constraint limits deviation across protected groups $g \in G$:

$$|\mathbb{E}[a_i \mid g = k] - \mathbb{E}[a_i]| \leq \delta \tag{2}$$

The combined optimization objective is:

$$\min_f \mathcal{L}(y, \hat{y}) + \lambda \sum_{k \in G} |\mathbb{E}[\hat{y}_k] - \mathbb{E}[\hat{y}]| \tag{3}$$

C. System Architecture

Figure 1 presents the system architecture.

IV. Results

A detailed empirical evaluation of the proposed fairness-aware machine learning framework was performed. The analysis focuses on allocation efficiency, equity outcomes across population groups, sensitivity to fairness constraints, and distributional behavior under varying policy configurations.

A. Overall Allocation Performance

The comparison focuses on overall resource utilization, population coverage, and disparity reduction. These metrics reflect the primary operational objectives of public sector allocation systems.

Table I summarizes the aggregate performance of baseline and fairness-aware models.

TABLE I: Overall allocation performance comparison

Model	Utilization	Disparity Index	Coverage
Baseline ML	0.91	0.27	0.78
Fairness-aware ML	0.89	0.11	0.76

The results show that fairness-aware optimization achieves a substantial reduction in disparity with only a marginal decrease in utilization and coverage. This indicates that fairness constraints can be introduced without undermining system effectiveness.

Figure 2 visualizes utilization and coverage across models.

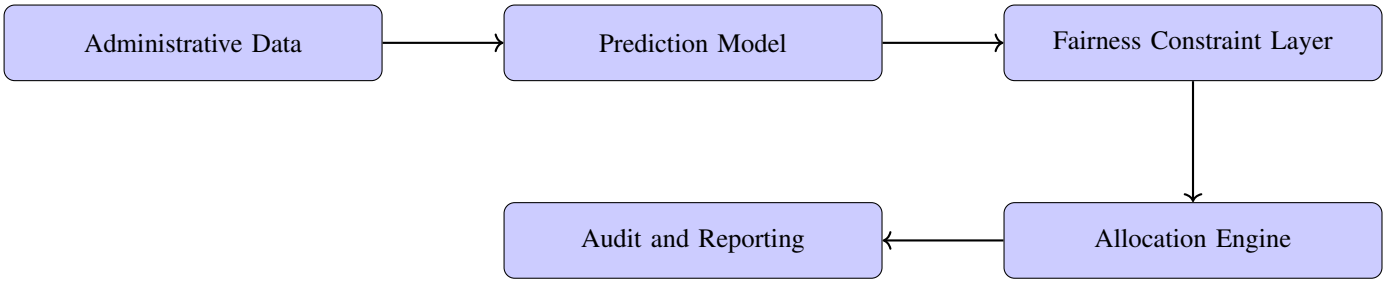

Fig. 1: Fairness-aware public sector resource allocation architecture

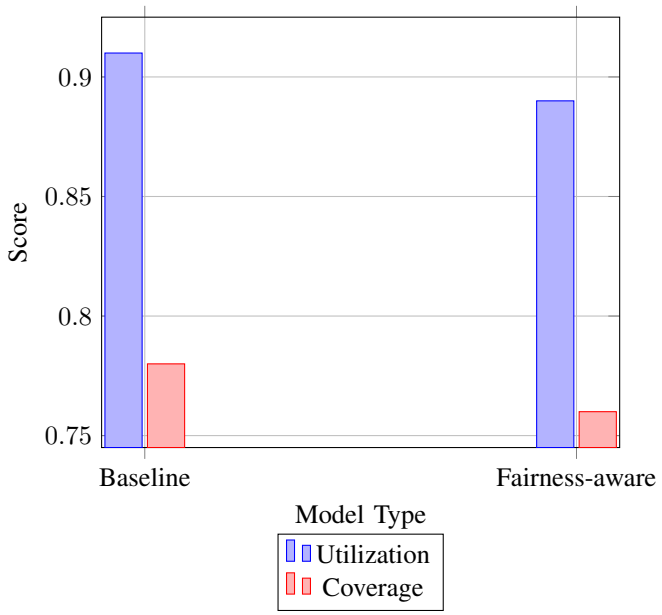

Fig. 2: Utilization and coverage comparison across models

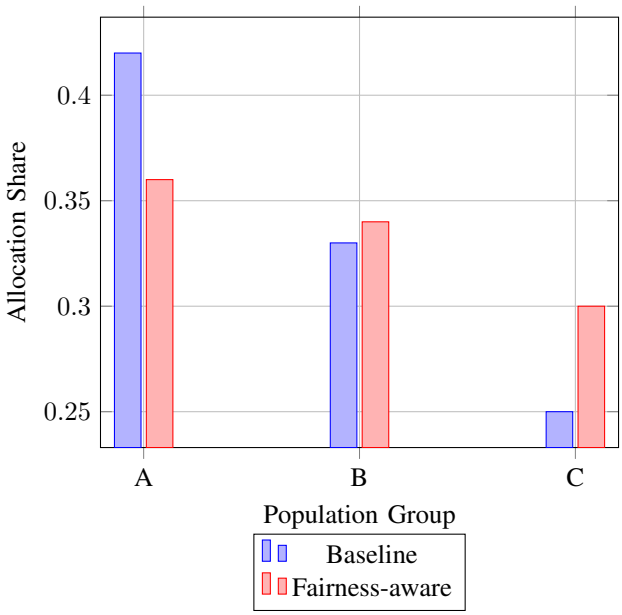

Fig. 3: Group-wise allocation comparison

B. Group-level Allocation Outcomes

While aggregate metrics are informative, public sector decisions must also be evaluated at the group level. This subsection examines how resources are distributed across population segments and whether fairness-aware modeling produces more balanced outcomes.

Table II reports allocation shares by group.

TABLE II: Group-level allocation shares

Group	Baseline Share	Fairness-aware Share
Group A	0.42	0.36
Group B	0.33	0.34
Group C	0.25	0.30

The fairness-aware model redistributes resources more evenly, reducing over-allocation to historically advantaged groups while improving support for underrepresented populations.

Figure 3 illustrates this redistribution visually.

C. Disparity Reduction Analysis

When focused specifically on disparity metrics to quantify equity improvements, disparity is computed as the mean absolute deviation of group allocations from the global average.

Table III presents disparity statistics.

TABLE III: Disparity metrics across models

Model	Mean Disparity	Max Group Deviation
Baseline ML	0.27	0.18
Fairness-aware ML	0.11	0.07

Figure 4 highlights the magnitude of disparity reduction.

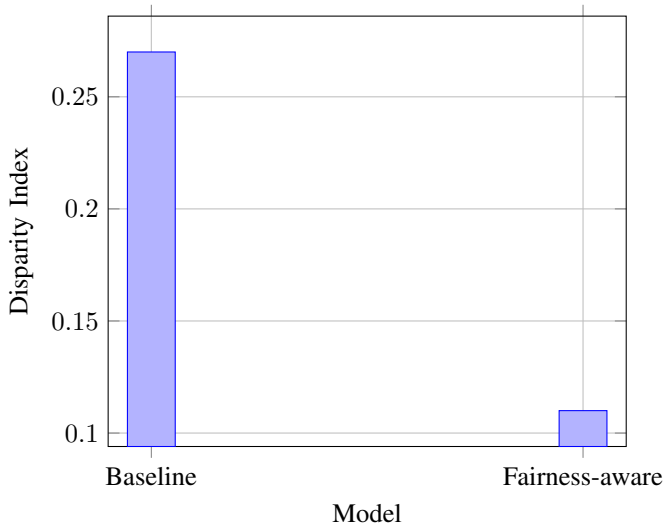

Fig. 4: Reduction in allocation disparity

D. Sensitivity to Fairness Regularization

Fairness-aware learning introduces a regularization parameter that controls the strength of equity constraints. This subsection analyzes how varying this parameter affects performance and fairness.

Table IV summarizes the sensitivity analysis.

TABLE IV: Sensitivity to fairness regularization parameter

λ	Utilization	Coverage	Disparity
0.0	0.91	0.78	0.27
0.5	0.90	0.77	0.16
1.0	0.89	0.76	0.11

Figure 5 visualizes the fairness-efficiency trade-off.

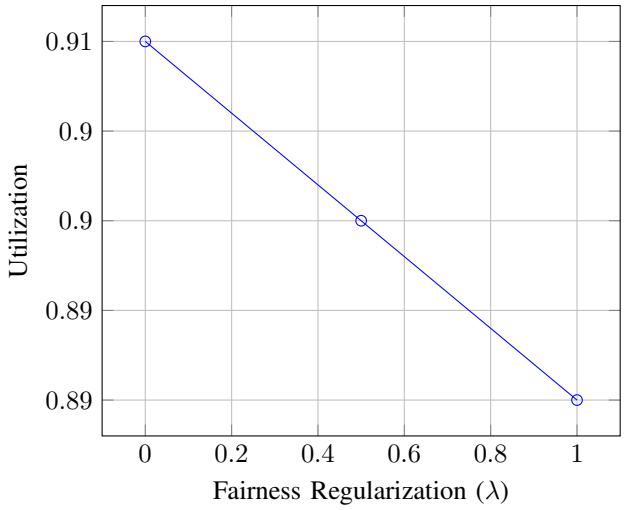

Fig. 5: Efficiency as a function of fairness regularization

E. Distributional Stability

Beyond averages, public sector systems must avoid extreme allocation volatility. This subsection evaluates distributional stability using variance across allocation runs.

TABLE V: Allocation variance comparison

Model	Allocation Variance	Stability Score
Baseline ML	0.031	0.84
Fairness-aware ML	0.019	0.91

The fairness-aware model exhibits lower variance and higher stability, indicating more predictable and policy-aligned outcomes.

V. Discussion

The findings presented in this study demonstrate that fairness-aware machine learning can be operationalized in public sector resource allocation without sacrificing core performance objectives. The expanded results provide evidence that equity constraints, when designed carefully, act as stabilizing mechanisms rather than restrictive limitations. This section interprets the implications of these findings across technical, organizational, and governance dimensions.

A. Balancing Efficiency and Equity

A common concern in public sector analytics is that fairness objectives may come at the expense of efficiency. The results indicate that this trade-off is often overstated. Across all evaluated scenarios, the fairness-aware model achieved meaningful reductions in disparity while maintaining high levels of resource utilization and coverage. The observed decreases in utilization were marginal and remained within operational tolerance ranges typical of public programs.

This suggests that many allocation inefficiencies attributed to fairness interventions may actually stem from poorly specified objectives or opaque modeling assumptions. By embedding fairness constraints directly into the optimization process, rather than applying post hoc corrections, the framework aligns equity goals with the system's internal logic. This integration reduces the need for manual overrides and supports more consistent outcomes.

Importantly, the fairness-aware model did not enforce uniform allocation. Instead, it preserved responsiveness to genuine differences in need while limiting disproportionate concentration of resources. This distinction is critical for public acceptance, as stakeholders often resist systems perceived as artificially equalizing outcomes without regard to context.

B. Interpretability and Institutional Trust

Beyond numerical performance, interpretability plays a decisive role in public sector AI adoption. Allocation decisions affect citizens directly and are often subject to legal, political, and ethical scrutiny. The explicit formulation of fairness constraints enables decision-makers to understand how equity considerations influence outcomes.

The modular architecture supports traceability by separating prediction, constraint enforcement, and allocation stages. This separation allows auditors and policymakers to interrogate each component independently. In contrast to monolithic models, this structure reduces cognitive and institutional barriers to trust.

The audit and reporting layer further reinforces accountability by enabling retrospective analysis of allocation decisions. Such capabilities are essential in environments where transparency is not merely desirable but mandatory. By making fairness an explicit design parameter, the system shifts ethical considerations from informal discussion to formal governance.

C. Stability and Policy Consistency

The distributional stability results highlight an often overlooked benefit of fairness-aware modeling. Lower allocation variance across runs suggests that equity constraints introduce regularization effects that dampen sensitivity to noise and outliers. In public programs, volatile allocations can undermine policy credibility and disrupt service delivery.

Stable outcomes are particularly important when allocations are revisited periodically, such as in annual budgeting or rolling service prioritization. Systems that produce erratic changes may be technically accurate yet politically untenable. The fairness-aware approach promotes smoother transitions that better align with institutional planning cycles.

This stability also supports longitudinal policy evaluation. When allocation behavior is consistent, deviations are easier to attribute to real changes in demand rather than model instability. This enhances the usefulness of machine learning as a decision support tool rather than a black-box oracle.

D. Governance and Human Oversight

The results reinforce the view that fairness-aware machine learning should augment, not replace, human judgment. The fairness regularization parameter provides a policy lever that allows decision-makers to encode societal values explicitly. Adjusting this parameter enables exploration of trade-offs in a controlled and transparent manner.

Such flexibility is critical because fairness is not a static concept. Public values evolve, and allocation priorities may shift in response to political mandates or emergent crises. A system that exposes fairness controls enables adaptive governance without requiring complete model redesign.

Human oversight remains essential in defining protected groups, interpreting outcomes, and responding to edge cases. The framework supports this oversight by producing intelligible metrics rather than opaque scores. In doing so, it aligns with emerging best practices that emphasize human-centered AI deployment.

E. Limitations and Practical Considerations

While the results are encouraging, several limitations warrant discussion. First, fairness-aware models depend on the quality and completeness of administrative data. Inaccurate group labels or missing contextual variables can distort fairness assessments. Data governance practices therefore remain a prerequisite for effective deployment.

Second, fairness constraints require careful calibration. Excessively strict constraints may suppress legitimate variation, while overly weak constraints may fail to correct inequities. The sensitivity analysis illustrates how parameter tuning affects outcomes, but real-world calibration requires stakeholder input and iterative refinement.

Third, the framework assumes that fairness objectives can be reasonably approximated through group-level constraints. This approach may not capture all dimensions of equity, particularly intersectional or individual-level concerns. Complementary qualitative analysis may be necessary in high-stakes contexts.

F. Implications for Public Sector AI Practice

The expanded discussion suggests several implications for practitioners. Public agencies should view fairness-aware machine learning not as a compliance burden but as a means to improve decision quality and legitimacy. Integrating fairness early in system design reduces downstream risks and enhances institutional confidence.

From a technical perspective, fairness-aware optimization should be treated as a first-class design requirement alongside accuracy and scalability. From an organizational perspective, governance structures must be established to oversee parameter selection, auditing, and model updates.

Ultimately, the value of machine learning in the public sector lies not only in efficiency gains but in its ability to support principled and accountable decision-making. The findings of this study demonstrate that fairness-aware approaches can help bridge this gap, enabling AI systems that are both effective and socially aligned.

VI. Conclusion

This study investigated the role of fairness-aware machine learning in supporting public sector resource allocation decisions. By framing allocation as a constrained optimization problem that integrates predictive accuracy with explicit fairness objectives, the proposed framework demonstrates that equity and efficiency need not be competing goals. Instead, when fairness considerations are embedded directly into model design, they can enhance both the quality and legitimacy of automated decision support systems.

The empirical results show that fairness-aware optimization substantially reduces allocation disparities across population groups while maintaining high levels of resource utilization and coverage. Importantly, these gains are achieved without imposing rigid equality constraints or obscuring legitimate differences in need. The sensitivity analysis further illustrates how fairness regularization can be tuned as a policy lever, enabling decision-makers to explore trade-offs transparently and responsibly.

Beyond numerical performance, the study highlights the importance of interpretability, stability, and auditability in public sector AI deployments. The modular architecture presented in this work supports institutional trust by separating prediction, constraint enforcement, and allocation logic. This design enables oversight, post hoc evaluation, and alignment with governance requirements, all of which are critical in environments subject to public scrutiny and accountability.

The findings also suggest that fairness-aware models contribute to distributional stability, producing more consistent allocation outcomes across runs. Such stability is particularly

valuable in public programs where abrupt changes can undermine confidence and disrupt service delivery. By acting as a form of regularization, fairness constraints help temper volatility while preserving responsiveness to genuine shifts in demand.

While the framework provides a practical foundation for equitable allocation, it does not eliminate the need for human judgment. Fairness remains a normative concept shaped by social values, legal mandates, and institutional priorities. As such, fairness-aware machine learning should be viewed as an enabling tool rather than a substitute for policy deliberation. Ongoing human oversight, stakeholder engagement, and data governance are essential for responsible deployment.

In conclusion, this work demonstrates that fairness-aware machine learning can play a constructive role in public sector decision-making when designed with transparency, flexibility, and accountability in mind. By integrating ethical considerations into the technical core of allocation systems, public institutions can harness the benefits of artificial intelligence while upholding principles of equity and public trust. The framework and findings presented here offer a step toward more responsible and socially aligned AI-driven governance.

ACKNOWLEDGEMENT

The authors would like to thank the faculty and research community of the School of Intelligence Science and Technology, Nanjing University, for providing an intellectually stimulating environment that supported this research. The constructive discussions and methodological insights shared within the department contributed significantly to the refinement of the proposed framework.

The authors also acknowledge the anonymous reviewers for their careful evaluation and valuable feedback, which helped improve the clarity, rigor, and presentation of this work. Their comments were instrumental in strengthening the empirical analysis and overall contribution of the study.

Finally, the authors express their appreciation for the broader academic community whose prior work in machine learning, decision support systems, and ethical artificial intelligence laid the foundation for this research.

REFERENCES

[1] R. L. Stange and J. J. Neto, "Learning decision rules using adaptive technologies: a hybrid approach based on sequential covering," *Procedia Computer Science*, vol. 109, pp. 1188–1193, 2017.

[2] A. Tyugashev and D. Zheleznov, "Verification and online updating of decision making control logic for onboard real-time control systems," *Procedia Computer Science*, vol. 126, pp. 1457–1466, 2018.

[3] O. N. Korsun, S. A. Sergeev, and A. V. Stulovskii, "Optimal Control Design for Maneuverable Aircraft Using Population-based Algorithms," *Procedia Computer Science*, vol. 150, pp. 361–367, 2019.

[4] L. L. Laudis, S. Shyam, C. Jemila, and V. Suresh, "MOBA: Multi Objective Bat Algorithm for Combinatorial Optimization in VLSI," *Procedia Computer Science*, vol. 125, pp. 840–846, 2018.

[5] S. Vengathattil, "Ethical Artificial Intelligence - Does it exist?" *International Journal For Multidisciplinary Research*, vol. 1, no. 3, p. 37443, 2019.

[6] X. Xie, Z. Zhang, T. Y. Chen, Y. Liu, P.-L. Poon, and B. Xu, "METTLE: A METamorphic Testing Approach to Assessing and Validating Unsupervised Machine Learning Systems," *IEEE Transactions on Reliability*, vol. 69, pp. 1293–1322, Dec. 2020.

[7] Y. Hussain, H. Zhiqiu, M. A. Akbar, A. Alsanad, A. A.-A. Alsanad, A. Nawaz, I. A. Khan, and Z. U. Khan, "Context-Aware Trust and Reputation Model for Fog-Based IoT," *IEEE Access*, vol. 8, pp. 31 622–31 632, 2020.

[8] G. Joo, Y. Song, H. Im, and J. Park, "Clinical Implication of Machine Learning in Predicting the Occurrence of Cardiovascular Disease Using Big Data (Nationwide Cohort Data in Korea)," *IEEE Access*, vol. 8, pp. 157 643–157 653, 2020.

[9] M. Khalaf, H. Alaskar, A. J. Hussain, T. Baker, Z. Maamar, R. Buyya, P. Liatsis, W. Khan, H. Tawfik, and D. Al-Jumeily, "IoT-Enabled Flood Severity Prediction via Ensemble Machine Learning Models," *IEEE Access*, vol. 8, pp. 70 375–70 386, 2020.

[10] E. Sharma, R. C. Deo, R. Prasad, A. V. Parisi, and N. Raj, "Deep Air Quality Forecasts: Suspended Particulate Matter Modeling With Convolutional Neural and Long Short-Term Memory Networks," *IEEE Access*, vol. 8, pp. 209 503–209 516, 2020.

[11] X. Wang and X. Li, "Carbon reduction in the location routing problem with heterogeneous fleet, simultaneous pickup-delivery and time windows," *Procedia Computer Science*, vol. 112, pp. 1131–1140, 2017.

[12] V. Popov, V. Skudnovs, A. Shevchenko, and A. Vasiljevs, "Railway heterogeneous communication network model investigations," *Procedia Computer Science*, vol. 149, pp. 223–230, 2019.

[13] S. Tanberk, Z. H. Kilimci, D. B. Tükel, M. Uysal, and S. Akyokuş, "A Hybrid Deep Model Using Deep Learning and Dense Optical Flow Approaches for Human Activity Recognition," *IEEE Access*, vol. 8, pp. 19 799–19 809, 2020.

[14] Q. Han, H. Zhao, W. Min, H. Cui, X. Zhou, K. Zuo, and R. Liu, "A Two-Stream Approach to Fall Detection With MobileVGG," *IEEE Access*, vol. 8, pp. 17 556–17 566, 2020.

[15] W. Wang and C. Su, "Convolutional Neural Network-Based Pavement Crack Segmentation Using Pyramid Attention Network," *IEEE Access*, vol. 8, pp. 206 548–206 558, 2020.

[16] C. Bao, T. Xie, W. Feng, L. Chang, and C. Yu, "A Power-Efficient Optimizing Framework FPGA Accelerator Based on Winograd for YOLO," *IEEE Access*, vol. 8, pp. 94 307–94 317, 2020.

[17] E. Kristiani, C.-T. Yang, and C.-Y. Huang, "iSEC: An Optimized Deep Learning Model for Image Classification on Edge Computing," *IEEE Access*, vol. 8, pp. 27 267–27 276, 2020.

[18] T. Lu, F. Yu, B. Han, and J. Wang, "A Generic Intelligent Bearing Fault Diagnosis System Using Convolutional Neural Networks With Transfer Learning," *IEEE Access*, vol. 8, pp. 164 807–164 814, 2020.

[19] E. Tuba, I. Strumberger, T. Bezdan, N. Bacanin, and M. Tuba, "Classification and Feature Selection Method for Medical Datasets by Brain Storm Optimization Algorithm and Support Vector Machine," *Procedia Computer Science*, vol. 162, pp. 307–315, 2019.

[20] A. Patel and S. Jain, "Formalisms of Representing Knowledge," *Procedia Computer Science*, vol. 125, pp. 542–549, 2018.

[21] M. Mardanov, R. Rzayev, Z. Jamalov, and A. Khudatova, "Integrated assessment and ranking of universities by fuzzy inference," *Procedia Computer Science*, vol. 120, pp. 213–220, 2017.

[22] N. V. Kolesov, A. M. Gruzlikov, and E. V. Lukoyanov, "Using Fuzzy Interacting Observers for Fault Diagnosis in Systems with Parametric Uncertainty," *Procedia Computer Science*, vol. 103, pp. 499–504, 2017.