Ethical Challenges and Bias in Machine Learning-Based Data Cleaning for Healthcare: A Case Study on Diabetic Patient Records

Chanda Kumari , Research Scholar,
Department of Computer Science & Engg,
Sarala Birala University, Ranchi,India
Email: chanda.research@gmail.com

Dr. Dipti Kumari(co-author) Assistant Professor
Department of Computer Science & Engg.
Sarala Birla University, Ranchi
Email: dipti.kumari@sbu.ac.in

Abstract

Objective: Machine learning (ML) is increasingly used to automate data cleaning in healthcare, improving data quality for predictive analytics and clinical decision-making. However, the adoption of ML-based data cleaning introduces ethical challenges and potential biases that can affect the fairness, reliability, and integrity of patient records. This study explores the ethical implications of using ML-driven techniques for handling missing values, inconsistencies, and outliers in diabetic patient data. We analyze the risks of algorithmic bias, data representativeness, and privacy concerns in automated preprocessing. Through a case study on diabetic patient records, we evaluate the fairness of ML-based imputation techniques and their impact on predictive modeling. We discuss strategies to mitigate bias, ensure transparency, and align ML-driven data cleaning with healthcare ethics and regulatory standards such as GDPR and HIPAA. The study highlights the need for explainable and fair ML models to maintain trust and accuracy in healthcare data management.Outliers in healthcare datasets, particularly in diabetic patient records, can significantly impact predictive modeling and clinical decision-making. This study explores the challenges posed by outliers in diabetic datasets, reviews current trends in detection and mitigation, and evaluates effective solutions to enhance data quality and model accuracy.

Methods: This study analyzes ethical concerns in ML-based data cleaning using a publicly available diabetic patient dataset. Various data cleaning techniques were assessed, including missing value imputation, anomaly detection, and bias analysis. We applied supervised and unsupervised ML techniques, including K-Nearest Neighbors (KNN) imputation, Multiple Imputation by Chained Equations (MICE), Isolation Forests, and Density-Based Clustering (DBSCAN). Ethical concerns were evaluated using fairness-aware ML frameworks, disparate impact analysis, and explainability methods such as SHAP (Shapley Additive Explanations) to assess bias in model decisions.A publicly available Kaggle diabetic dataset was used to analyze outlier issues. Various detection techniques, including Z-score, Interquartile Range (IQR), and machine learning-based anomaly detection, were applied. The impact of different outlier handling strategies on predictive model performance was assessed.

Results: The findings highlight that ML-based data cleaning improves predictive model accuracy but introduces bias if improperly applied. Imputation techniques led to varying degrees of data distortion, disproportionately affecting minority patient groups. The bias analysis revealed that traditional rule-based cleaning methods resulted in more consistent imputations across demographic groups, whereas ML-based approaches amplified disparities in missing data handling. Using fairness-aware ML techniques reduced disparate impact, and explainability methods such as SHAP provided transparency into bias propagation. Overall, integrating fairness constraints into data cleaning frameworks significantly improved ethical compliance and model reliability.Experimental results show that addressing outliers through appropriate detection and mitigation techniques improves model accuracy, with Support Vector Machine (SVM) accuracy increasing from 65% (raw data) to 74% (cleaned data). The study highlights recent advancements, including deep learning-based anomaly detection, as promising solutions.

Conclusion: Machine learning-based data cleaning presents significant opportunities for improving data quality in healthcare, but it also raises ethical concerns, particularly regarding bias and fairness. This study highlights the impact of ML-driven imputation techniques on different demographic groups, revealing the risks of amplifying disparities if not handled properly. By integrating fairness-aware ML approaches, explainable AI techniques, and privacy-preserving data processing, we can create more equitable and transparent healthcare analytics systems. Future work should focus on developing standardized frameworks to assess bias in ML-based data cleaning and refining algorithms to minimize unintended consequences in clinical decision-making.Outlier detection and handling are crucial for improving predictive accuracy in diabetic datasets. This paper provides insights into contemporary techniques and proposes hybrid approaches for robust anomaly detection in medical datasets.

Keywords: Machine Learning, Data Cleaning, Algorithmic Bias, Fairness in AI, Healthcare Data Ethics, Explainable AI, Privacy-Preserving Machine LearningOutliers, diabetic dataset, anomaly detection, data cleaning, machine learning, predictive modeling

1. Introduction

1.1 Background and Motivation

High-quality healthcare data is crucial for effective disease management, yet real-world datasets often contain missing values, inconsistencies, and outliers. Machine learning (ML) has emerged as a powerful tool to automate data cleaning processes. However, automated methods may inadvertently introduce bias, disproportionately impacting underrepresented patient groups. Addressing these ethical challenges is critical to ensuring fairness and accuracy in healthcare analytics.

1.2 Research Objectives

This paper aims to:

- Identify ethical concerns in ML-based data cleaning for diabetic patient records.
- Evaluate potential biases in data imputation and anomaly detection.
- Assess the impact of automated cleaning on predictive models.
- Propose strategies for mitigating bias and ensuring ethical data practices.

1.3 Hypotheses

Based on the literature review and research objectives, the following hypotheses are proposed:

H1: ML-based data cleaning improves data quality and fairness in predictive models.

H2: Bias in training data leads to systematic errors in imputation techniques.

H3: Explainable AI methods enhance transparency and trust in ML-based data cleaning.

H4: Privacy-preserving data cleaning methods mitigate ethical risks while maintaining accuracy.

2. Literature Review

2.1 ML-Based Data Cleaning in Healthcare

Mehrabi et al. (2021) provide a comprehensive survey on bias and fairness in ML applications, indicating that imbalanced datasets can lead to biased predictions, particularly in healthcare settings. Similarly, Rajkomar et al. (2018) emphasize the importance of fairness-aware ML techniques to enhance health equity.

Chouldechova (2017) discusses the implications of disparate impact in predictive models, demonstrating how incorrect imputations in underrepresented groups can exacerbate existing healthcare disparities. Kleinberg et al. (2018) further highlight algorithmic discrimination in automated decision-making, calling for more transparent and interpretable models.

2.2 Ethical Concerns in ML Data Cleaning

Obermeyer et al. (2019) examine racial bias in healthcare algorithms, revealing that ML models trained on biased datasets can perpetuate disparities in clinical outcomes. Their study underscores the need for fairness-aware data cleaning methods.

Binns (2018) explores the intersection of ML and political philosophy, discussing ethical frameworks for bias mitigation in AI-driven decision-making. These insights inform strategies to ensure ethical ML-based data cleaning in healthcare.

Danks & London (2017) analyze algorithmic bias in autonomous systems, reinforcing the necessity of accountability and transparency in ML applications. Their findings align with Hao (2019), who warns against biases introduced during AI training phases.

3. Research Methodology

3.1 Dataset and Methodology

We evaluate the fairness and impact of ML-based data cleaning using a publicly available diabetic patient dataset. The following techniques are applied:

- **Imputation Methods:** KNN, Multiple Imputation by Chained Equations (MICE), and Autoencoders.
- **Anomaly Detection:** Isolation Forests and Density-Based Clustering (DBSCAN).
- **Bias Analysis:** Comparing imputation accuracy across different patient demographics.

3.2 Ethical Evaluation Metrics

To assess fairness and ethical impact, we use:

- **Disparate Impact Analysis:** Measuring bias in imputations based on demographic attributes.
- **Explainability Metrics:** Using SHAP values to evaluate model transparency.
- **Privacy Risk Assessment:** Evaluating data anonymization effectiveness.

4. Results and Discussion

Metric	Raw Data	Cleaned Data
Predictive Model Accuracy	65%	More than74%
Bias in Imputation (Disparate Impact)	High	Reduced
Privacy Risk Score	Moderate	Low
Model Explainability (SHAP)	Limited	Improved

4.1 Key Findings

- **Bias in Missing Data Handling:** Imputed values for minority groups deviated significantly from original patterns, introducing predictive disparities.
- **Impact on Model Performance:** Predictive models trained on cleaned data showed improved accuracy overall but reduced performance for certain subgroups.
- **Privacy Concerns:** Ensuring compliance with data anonymization standards was challenging when handling outlier detection.

5. Mitigation Strategies for Ethical Data Cleaning

5.1 Fairness-Aware Machine Learning

Incorporating fairness constraints in ML models can mitigate bias. Techniques like adversarial debiasing and re-weighting can help balance imputations across patient groups.

5.2 Transparent and Explainable AI (XAI)

Using explainable AI methods such as SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-Agnostic Explanations) can improve interpretability and trust in ML-based data cleaning.

5.3 Privacy-Preserving Data Cleaning

Applying differential privacy and federated learning can ensure secure and compliant data preprocessing without exposing sensitive patient information.

5.4 Inclusive and Representative Training Data

Healthcare institutions should ensure diverse and representative training datasets to prevent bias amplification in ML-based data cleaning systems.

6. Conclusion and Future Work

ML-driven data cleaning holds great potential for improving healthcare data quality, but it also presents significant ethical challenges. Addressing biases, ensuring transparency, and adhering to privacy regulations are crucial for ethical adoption. Future work should focus on fairness-aware ML techniques, real-time monitoring of bias, and explainable AI solutions for responsible data preprocessing in healthcare.

References

1. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys*, 54(6), 1-35.
2. Rajkomar, A., Hardt, M., Howell, M. D., Corrado, G., & Chin, M. H. (2018). Ensuring fairness in machine learning to advance health equity. *Annals of Internal Medicine*, 169(12), 866-872.
3. Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big Data*, 5(2), 153-163.
4. Kleinberg, J., Ludwig, J., Mullainathan, S., & Sunstein, C. R. (2018). Discrimination in the age of algorithms. *Journal of Legal Analysis*, 10, 113-174.
5. Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447-453.
6. Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. *Proceedings of the 2018 Conference on Fairness, Accountability, and Transparency*.
7. Danks, D., & London, A. J. (2017). Algorithmic bias in autonomous systems. *IJCAI Proceedings*, 4691-4697.
8. Hao, K. (2019). Training data for AI systems can introduce bias. *MIT Technology Review*.
9. Barocas, S., Hardt, M., & Narayanan, A. (2019). *Fairness and Machine Learning: Limitations and Opportunities.* MIT Press.
10. Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. *New England Journal of Medicine*, 380(14), 1347-1358.