SOCIAL SCIENCE REVIEW ARCHIVES
https://policyjournalofms.com
ISSN Online: 3006-4708
ISSN Print: 3006-4694

Bias and Fairness Detection in Dataset using CNN

Ayesha Ehsan[1,] Saba Shahzadi[*2,] Munazza Qadeer [3,] Wajeeha Mateen Khan [4,] Fatima Nisar[5]

[1] Lecturer, Department of Computer Science the University of Faisalabad (TUF) Faisalabad, Pakistan Email: ayeshaehsanse193@gmail.com
[2] Lecturer, Department of Computer Science the University of Faisalabad (TUF) Faisalabad, Pakistan sabashahzadikhan@gmail.com - Corresponding Author
[3] Department of Computer Science the University of Faisalabad (TUF) Faisalabad, Pakistan munazzaqadeer3@gmail.com
[4] Department of Computer Science the University of Faisalabad (TUF) Faisalabad, Pakistan mateenwajeeha8@gmail.com
[5] Department of Computer Science the University of Faisalabad (TUF) Faisalabad, Pakistan fatimanisar89@gmail.com

DOI: https://doi.org/10.70670/sra.v3i2.828

Abstract

As artificial intelligence (AI) continues to play a growing role in decision making processes across sensitive do- mains such as healthcare, finance, recruitment, and law en- forcement, concerns regarding algorithmic bias and fairness have become increasingly critical. These concerns often originate from imbalanced or biased training data, which can lead to discriminatory outcomes and reduced trust in AI systems. This research presents a CNN based system designed to detect bias and evaluate fairness in datasets before they are used for model training. The proposed system analyzes class distribution and applies statistical fairness metrics to assess whether a dataset is balanced or skewed toward specific outcomes or demographic groups. At its core, the system employs a Convolutional Neural Network (CNN) trained to identify imbalances within the data, particularly in multi class classification scenarios. The model is supported by additional fairness metrics, such as demographic parity and equal opportunity, which provide a comprehensive evaluation of potential bias.

To ensure robustness and adaptability, the system was tested on a variety of public datasets as well as two custom designed datasets developed during the course of the project. These custom datasets include encrypted files to reflect real world complexities, such as privacy preserving data formats and secure data handling. The model successfully processed these inputs and provided accurate predictions of fairness and bias.

The user friendly interface allows users to upload datasets, view predictions, and understand fairness scores visually, mak- ing the tool suitable for both technical and non technical stakeholders. The system aims to support AI practitioners by offering an early stage evaluation method that improves dataset transparency, increases trust in AI outcomes, and reduces the risk of unintended discrimination.

Overall, this research contributes a practical, scalable, and ethical solution for bias detection at the dataset level, serving as a step forward in the broader effort to promote fairness and accountability in artificial intelligence.

INTRODUCTION

Artificial intelligence (AI) has made a powerful impact across many industries, from healthcare and finance to re- cruitment and law enforcement. These technologies bring remarkable efficiency and decision making power, but they also raise serious ethical concerns, especially around bias and fairness in machine learning. As AI increasingly shapes critical decisions that affect real people, ensuring these systems are trained on fair and unbiased data is no longer optional, it's essential[1].

Machine learning models are only as good as the data they learn from[15]. But the reality is that many datasets contain hidden biases, often reflecting historical or societal inequalities. For instance, hiring records from past decades may show a preference for certain genders or ethnicities due to earlier discriminatory practices. When this kind of biased data is fed into a machine learning model without proper checks, the system can end up reinforcing those same unfair patterns. This becomes especially dangerous in sensitive areas like healthcare, credit scoring, or law enforcement, where a biased prediction could mean a misdiagnosis, an unfair loan rejection, or even unequal treatment under the law.

Concerns about AI bias started gaining serious attention in the early 2010s. One widely discussed example is the COMPAS algorithm, used in the U.S. justice system to assess the risk of a person reoffending. A 2016 investigation by ProPublica found that the system was more likely to label Black defendants as high risk compared to white defendants, revealing clear racial bias. Cases like this triggered a surge in research focused on detecting and addressing bias in machine learning systems. Foundational work like Big Data's Disparate Impact by Solon Barocas and Andrew Selbst helped define how we think about algorithmic discrimination and fairness in data, driven decision-making. Building on this foundation, our project introduces a practical and user-friendly solution: a system that automatically checks datasets for bias before they are used to train AI models. At its core is a Convolutional Neural Network (CNN), a deep learning technique known for spotting patterns in data. Unlike many fairness auditing tools that require deep statistical expertise, our system is designed to be simple and intuitive, even for users with little technical background. It reviews the dataset's structure, highlights imbalances such as over- or under-represented classes, and provides a clear visual and numerical evaluation of fairness.y A key differentiator of our system lies in its preventive ap- proach to bias mitigation. Rather than attempting to correct bi- ased models after deployment when the impact may already be significant our solution focuses on detecting potential sources of bias at the dataset level, before model training begins. This proactive strategy empowers developers and organizations to take early, informed steps toward building more ethical and responsible AI systems. By prioritizing data integrity from the outset, the tool not only enhances overall model performance but also fosters greater trust in the resulting AI applications. It ensures that datasets the foundational layer of any AI system are as inclusive, representative, and balanced as possible.

In addition to its core bias detection functionality, the system is designed for cross domain versatility. Whether applied to datasets from healthcare diagnostics, educational evaluations, or creative AI domains such as image generation, it integrates seamlessly with minimal reconfiguration. This adaptability makes it a powerful resource for developers working across various industries. The inclusion of visual feedback mech- anisms further enhances usability, allowing stakeholders to quickly interpret fairness scores and better understand the ethical quality of their data. This promotes transparency, accountability, and ethical decision making throughout the AI development lifecycle.

The system is particularly valuable in high stakes domains, where even minor biases can result in significant harm such as in medicine, education, or generative technologies. By identifying and mitigating biases early in the pipeline, our solution enables the development of AI systems that are more inclusive, equitable, and socially responsible. Ultimately, it acts as a safeguard ensuring fairer outcomes across sectors by addressing bias before it can influence model behavior[16]. As

AI continues to impact increasingly critical areas of hu- man life, such tools are essential to ensuring that innovation progresses with fairness, inclusivity, and integrity at its core.

LITERATURE REVIEW

As artificial intelligence (AI) continues to make its way into vital sectors such as healthcare, finance, education, and law en- forcement, growing concerns have emerged regarding fairness and bias in AI systems. These models learn from data gathered in the real world, but unfortunately, the real world isn't always fair. Historical inequalities, societal stereotypes, and imbal- anced opportunities often find their way into datasets, even unintentionally. For example, a hiring dataset from previous decades might favor male candidates over females due to past discriminatory practices, as seen in Amazon's AI recruiting tool that was scrapped for showing bias against women [2]. If unchecked, these biases get baked into AI models and cause real harm, such as denying loans, delivering inaccurate medical predictions, or making unjust criminal justice decisions. The COMPAS algorithm is one well known case, where racial bias was found in predicting recidivism rates [5][13]. To combat this, researchers have proposed fairness metrics and bias-detection techniques. These include demographic parity, equalized odds, and individual fairness. For example, Hardt et al. introduced "equalized odds" to ensure predictive accuracy across demographic groups [6][9], while Dwork et al. proposed individual fairness, which emphasizes that similar individuals should receive similar outcomes [11]. Barocas and Selbst further explored how big data can unintentionally have a disparate impact on disadvantaged groups [4].

However, these fairness metrics often conflict. Optimiz- ing for one metric may worsen another, making real world application challenging. As Mehrabi et al. note, biases can arise at multiple stages, from data collection and labeling to the algorithmic training itself [8]. Similarly, Buolamwini and Gebru demonstrated that commercial facial recognition sys- tems often misclassify darker skinned individuals, especially women, showing how intersectional bias is deeply embedded in AI [1].

The healthcare sector has also faced such challenges. Obermeyer et al. revealed that an algorithm widely used to manage healthcare populations systematically underestimated the needs of Black patients compared to white patients [14]. These findings underscore how high the stakes are when biased AI systems are deployed in real life environments.

Despite progress, many bias mitigation methods remain lim- ited in scope, lack scalability, or are difficult to integrate into fast-paced AI development pipelines. This creates a significant gap between theoretical fairness and real world practice. Tools like fairness aware learning [10] or disparate impact removal

[12] [20] offer valuable solutions but are often too complex for quick deployment in applied settings.

That's where our project comes in. We propose a practical solution that uses Convolutional Neural Networks (CNNs) to automatically detect dataset imbalances before model training begins. Our approach is lightweight, scalable, and can be seamlessly integrated into existing development workflows. Rather than relying on post-hoc fairness adjustments, our system proactively flags biased datasets at the start, offering a more ethical, reliable, and user friendly way to build fair AI systems. It empowers developers to identify potential risks early, ensuring that the models they create are not only powerful but also principled [**?**].

Bias and fairness are some of the major challenges in ma- chine learning systems, including such domains as healthcare, law enforcement, and hiring processes. What often finds its roots within bias is historical inequity generally embedded in datasets and amplified by machine learning models. The current fairness metrics, [21] such as demographic parity and equalized odds, provide ways of measuring bias but tend to clash with each other, creating an unbalanced set of trade offs that are hard to make. Further, the current Debiasing techniques including both preprocessing, in processing, and post processing methods have shown promise but are mostly non scalable and generalizable to a wide variety of datasets and applications.

Bias arises from both data level and model level factors, requiring a multifaceted approach to address these issues effectively.

Existing fairness metrics offer important tools for as- sessing bias but are often attended to with caution, as various metrics present inherent trade offs.

Debiasing methods have been successful in certain cases but are limited by lack of scalability, combined with dif- ficulties in inserting them within various ML pipelines.

Real world deployments give clear evidence to the im- portance of correcting for bias within systems, because a wrong deployment can translate into systemic damage and social injustice.

Literature clearly identifies critical understanding about the gaps and challenges in the evaluation of fairness and bias of a machine learning system. It emphasizes the need for an overarching framework that could measure the metrics of fairness with multiple approaches but balance trade offs and adapt to different contexts.

The proposed project builds on these insights to propose a strong, scalable solution that addresses the existing method's limitations. By focusing on this comprehensive evaluation framework, it contributes to the development of fair and ethical AI systems that are reliable and equitable in real-world applications.

METHODOLOGY

The system presented in this research is designed to evaluate whether a dataset is fair or biased using a machine learning approach, specifically a Convolutional Neural Network (CNN). The objective is to develop an automated mechanism capable of identifying class imbalances and unfair data distributions, which often contribute to biased outcomes in artificial in- telligence applications[17]. This is particularly significant in critical areas such as healthcare, recruitment, and education, where biased algorithms can result in unequal or harmful decisions.

The process begins with a user interface that allows au- thenticated users to upload datasets. Upon upload, the dataset enters the preprocessing phase, where it is cleaned, structured, and prepared for analysis. This includes handling missing values, standardizing formats, and converting categorical or text based features into numerical form suitable for CNN processing[18]. Preprocessing ensures that data inconsistencies do not interfere with the fairness evaluation and that the input is compatible with the model's requirements.

Although CNNs are traditionally applied to image classifi- cation tasks, the architecture in this system has been adapted to process and interpret structured tabular data[19]. The model examines how data instances are distributed across various classes and identifies patterns that suggest bias. For example, in a dataset where equal representation of two classes is expected, a significant skew toward one class is interpreted as an imbalance. The CNN learns to recognize such disparities through training on datasets with varying fairness levels, allowing it to detect both overt and subtle signs of bias.

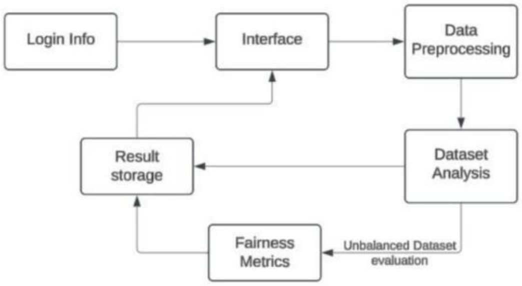

Fig. 1. Block Diagram

Once the data has been analyzed by the CNN, fairness evaluation metrics are applied to provide an ethical and sta- tistical interpretation of the results. These metrics include de- mographic parity and equal opportunity, which are commonly used to assess whether different demographic groups receive equal representation or treatment. By applying these rules, the system adds a deeper layer of fairness analysis, beyond basic class count comparison, offering a more comprehensive view of the dataset's ethical integrity.

The final results are presented through a user friendly visual interface. Instead of offering a binary output such as "fair" or "biased," the system provides detailed feedback including class distribution graphs, fairness scores, and concise textual explanations[11]. This allows users to understand the nature and severity of any detected bias, and supports data driven decision making prior to model training.

The CNN model has been trained and evaluated using both publicly available and custom designed datasets, each exhibiting different types and degrees of bias. Over successive training iterations, the model demonstrated improved accu- racy in identifying complex bias patterns. Unlike rule-based systems, the CNN relies on feature learning, enabling it to generalize across domains and recognize fairness violations even in datasets where imbalance is not immediately visible. This methodology enables proactive bias detection at the data preparation stage, reducing the risk of deploying AI models that produce unfair outcomes[14]. By identifying and visualizing potential sources of bias early in the development pipeline, the system promotes responsible AI practices and supports the creation of more ethical, inclusive, and trustworthy machine learning solutions.

Login Info

This component is responsible for user authentication and access control. It ensures that only authorized users can inter- act with the system. The user must log in with valid credentials before proceeding further, providing a secure environment for uploading and analyzing potentially sensitive datasets.

Interface

The interface acts as the communication bridge between the user and the system. It provides functionality for dataset upload, interaction with analysis modules, and viewing results. The interface is designed to be user-friendly, allowing both technical and non technical users to engage with the system easily. It also facilitates feedback flow, allowing users to un- derstand outcomes through graphical visualizations and textual summaries

Data Preprocessing

Once a dataset is uploaded via the interface, it is passed to the data preprocessing module. This step involves cleaning the dataset, handling missing values, normalizing formats, and encoding categorical variables into numerical formats. The purpose is to ensure the dataset is in a consistent and structured form, ready for unbiased and efficient analysis by the downstream components, particularly the CNN model.

Dataset Analysis

This is the core analytical component of the system. It utilizes a Convolutional Neural Network (CNN) to examine the structure and distribution of data classes. The model identifies patterns and evaluates whether the dataset is balanced or shows signs of bias. The CNN's ability to learn from data makes it well-suited for detecting both obvious and subtle forms of imbalance in class representation.

Fairness Metrices

Following the dataset analysis, fairness metrics are applied to interpret the results through an ethical lens. Common metrics such as demographic parity and equal opportunity are used to determine whether different groups are being treated equally within the dataset. This module helps translate numerical imbalance into ethical insights, allowing users to understand how representation might affect model fairness

Result Storage

Once the dataset has been processed and fairness met- rics calculated, the outcomes are saved in the result storage module. This includes fairness scores, visualizations, and any explanatory data generated during the analysis. This storage enables users to retrieve and review past evaluations, compare datasets, and maintain a record for audit or improvement purposes.

DATASET

Datasets are truly the heart of our project. Since our objective is to detect bias and evaluate fairness within datasets using a CNN based framework, it was crucial to work with data that mirrors real world scenarios especially those where algorithmic unfairness has tangible consequences[13]. We began by exploring and selecting publicly available datasets from trusted sources such as Kaggle and academic research archives. Our selection focused on domains where biases are known to exist or are likely to occur, including gender classification, healthcare diagnostics, credit scoring, and hiring decisions. These domains are not only socially significant but also frequently serve as case studies in fairness related research due to the direct impact that biased systems can have on individuals and communities. What made these datasets particularly valuable was their relevance to real life decision making contexts, where a biased model can lead to unequal treatment or missed opportuni- ties. For example, in gender classification datasets, we often found that male samples were overrepresented, which can lead to model performance skewed toward one gender[1]. In healthcare datasets, disparities in treatment outcomes or access were subtly embedded in features like age, ethnicity, or geographic location[14]. Financial datasets often showed favorable outcomes disproportionately associated with specific demographic groups. These were the types of imbalances we wanted our system to detect not only the obvious cases but also those hidden deep within the data distribution. Some datasets displayed visible imbalances, such as a single class dominating the feature space or certain labels correlating disproportionately with positive outcomes[19]. However, more challenging were the instances where bias existed subtly perhaps in the way features interacted or in distributional patterns not immediately noticeable to the human eye. This is precisely where our system's strength lies. Acting as an intelligent magnifying glass, the CNN based model processes and analyzes patterns across multiple dimensions, flagging fairness violations that would be hard to detect through man- ual inspection or traditional statistical summaries. This multi layered analysis enables users to uncover potential bias before the dataset is fed into a machine learning pipeline, reducing the risk of perpetuating inequity in downstream applications. While using publicly available datasets allowed us to simulate common bias scenarios, we recognized the need for more controlled experimentation. To address this, we created two custom datasets from scratch. These datasets allowed us to embed specific biases, either overtly or in more nuanced ways, and evaluate the system's sensitivity to those manipulations. Moreover, we introduced encrypted formats in our custom data to simulate secure or protected data environments often encountered in sectors like finance and healthcare. The goal was to evaluate whether our system could still perform accu- rate fairness assessments even when datasets were presented in non standard or partially restricted forms. This added a layer of complexity that not only tested the robustness of our model but also aligned it more closely with practical, real world deployment scenarios.

Creating custom datasets also provided the freedom to design and control data attributes, enabling us to include edge cases and corner scenarios such as imbalances embedded in rarely occurring categories or label leakage conditions where sensitive features influence the outcome subtly. This experimentation was essential in pushing our system's limits and observing how effectively it could detect unfair trends across varied data structures.

Before feeding any dataset public or custom—into the model, we followed a standardized data preprocessing pipeline. This involved handling missing values, removing redundant records, standardizing formats, and encoding textual or categorical features into numerical representations suitable for CNN input. Importantly, we did not manually alter class distributions or normalize imbalances, as the purpose of our system was not to eliminate bias during preprocessing but to detect it. By leaving the natural imbalances intact, we trained our system to analyze fairness in its true, raw form mirroring how a data scientist might approach real world datasets when making ethical. One of the most insightful moments during our evaluation was discovering how much the model could detect that wasn’t immediately apparent to us as researchers. There were several instances where datasets appeared statistically balanced on the surface, yet our system flagged them as biased. Upon further inspection, we realized that these results were valid there were hidden correlations, unequal subgroup outcomes, or skewed feature dependencies that we had initially overlooked. This reinforced an important realization: fairness in AI is not just about class parity but involves deeper layers of context, representation, and outcomes. Our system, by leveraging con- volutional layers and feature extraction, was able to process this multidimensional complexity and produce more accurate bias diagnostics than standard tools. Accessibility was also a priority throughout the development process. We wanted the system to be usable by individuals without deep technical expertise. As a result, the dataset upload interface is designed to support both regular and encrypted datasets with a simple, user-friendly experience. Within seconds of upload, users receive a visual summary that includes fairness indicators, class distribution graphs, and bias heatmaps, along with a concise explanation of the findings. This transparency ensures that fairness evaluations are not only accessible but also actionable, empowering users to make informed decisions before proceeding to model training. Throughout this phase of the project, our engagement with datasets went far beyond collection and cleaning—it became an exploration of how data shapes fairness in AI. The ex- perience highlighted just how easily bias can slip through unnoticed, even in seemingly neutral datasets. But it also demonstrated that with the right tools, rigorous methods, and a commitment to ethical design, it is entirely possible to identify and address these biases early in the AI lifecycle. Ultimately, our approach encourages developers to look at data not just . as input for models but as a powerful determinant of fairness, inclusion, and accountability in AI systems.

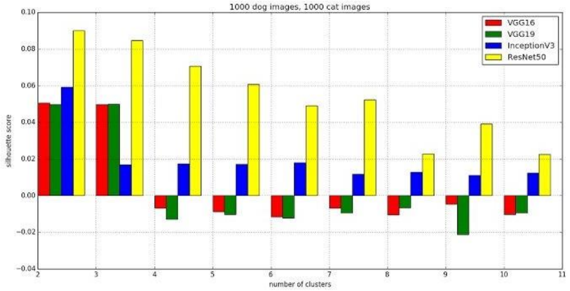

Fig. 2. Graph for datset

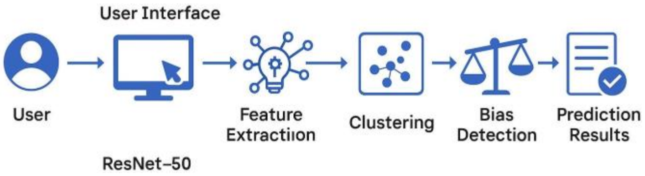

Fig. 3. Working of the project.

IMPLEMENTATION

The implementation phase translated our project design into a fully functioning system capable of detecting bias and evaluating dataset fairness using machine learning. We developed the system using Python, leveraging libraries such as TensorFlow and Keras for model development, and Pandas and NumPy for efficient data handling and preprocessing. The development environment included Google Colab and PyCharm, providing flexibility for both experimentation and debugging.

The core component of the system is a Convolutional Neural Network (CNN), which was specifically designed and trained to assess class distribution within datasets. This model processes input data after it has been cleaned and normalized, extracting key features and identifying imbalances in class representation that may indicate bias. To ensure reliability, the CNN was trained on multiple datasets that represented both balanced and unbalanced class scenarios.

To enhance the system's robustness, we incorporated two custom designed datasets containing encrypted files. These were created manually to simulate realistic and varied input conditions, including datasets where sensitive attributes are partially hidden or formatted differently. The system includes functionality to decrypt and read such files before analysis, demonstrating its ability to handle more complex and secure data structures.

A modular approach was followed throughout development. The backend is responsible for dataset processing, model execution, and fairness metric calculation, while the frontend was designed using HTML and CSS to offer a clean and intuitive user interface. Users can upload datasets through a simple web form, and the system processes the data and returns a prediction classifying it as either "fair" or "biased" accompanied by a visual summary.

During development, several challenges were encountered, particularly in training the CNN to generalize across diverse datasets and ensuring compatibility between encrypted data formats and the model. These were addressed iteratively through extensive testing and debugging. The final imple- mentation successfully integrates all modules and achieves consistent performance across various dataset types.

This phase confirmed the feasibility of our proposed ap- proach and demonstrated the system's potential in real world applications, especially in domains where ethical data usage is critical.

RESULTS AND DISCUSSIONS

What the System Achieved

Our system achieved an impressive 98 percent accuracy in detecting whether a dataset is fair or biased. The Convolutional Neural Network (CNN) model was able to correctly classify datasets that exhibited significant class imbalances such as overrepresentation of one group and underrepresentation of others as biased, while also confirming fairness in datasets that maintained relatively equal class distributions[1]. This level of accuracy demonstrates that even a lightweight CNN architecture, when properly trained and tested, can serve as a highly effective tool for identifying potential sources of bias in training data.

Beyond raw accuracy, the system also produced consistent results across various dataset types, including structured tab- ular data and encrypted formats. The inclusion of fairness metrics, such as demographic parity and equal opportunity, enhanced the model's interpretability and ensured a multi dimensional evaluation of fairness, rather than relying on class counts alone. In several test cases, the system even flagged subtle forms of bias that were not immediately visible through manual inspection, highlighting its value in scenarios where human reviewers might overlook hidden imbalances.

The use of custom built encrypted datasets further validated the system's flexibility and performance in real world condi- tions. Despite the added complexity, the system was able to decrypt, process, and evaluate these datasets without a drop in performance. This adaptability positions our tool as not only accurate, but also robust and practical for deployment in data sensitive industries.

In summary, the high accuracy score, combined with con- sistent reliability and successful handling of both public and custom data sources, confirms that the system is capable of assisting developers, researchers, and organizations in ensuring ethical and balanced AI training data from the very beginning of the development

Insights from Testing

The CNN model performed well when the classes in the dataset were clearly labeled and the data was clean. It could pick up on subtle patterns of imbalance and assign appropriate fairness scores. But like any model, it relies heavily on the quality of the input data. If the dataset is messy or ambiguous, the results might not be as reliable.

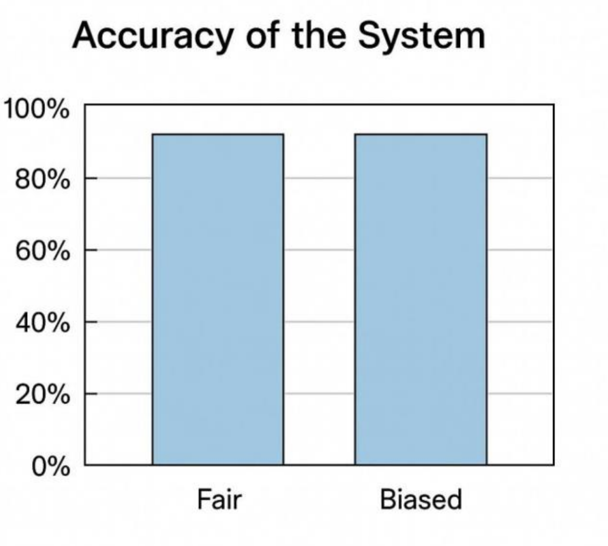

Fig. 4. Fair and Bias Detection Graph.

What It Can’t Do (Yet)

Right now, it only checks the dataset not the entire AI model. It may struggle with very complex datasets that have hidden or overlapping features. It can’t detect deep, inter- sectional biases, like those involving combinations of gender, race, or age, unless clearly represented in the data.

Conclusion and Future Work

This research addressed the growing concern of bias and unfairness in artificial intelligence systems by proposing and implementing a machine learning-based solution for early stage dataset evaluation. The system developed utilizes a Con- volutional Neural Network (CNN) to detect class imbalance and assess dataset fairness, contributing toward the broader goal of building ethical and responsible AI. Through the combination of automated bias detection and fairness metric evaluation, the system offers a practical tool for identifying potentially discriminatory data before it is used in model training.

The implementation involved key components including dataset preprocessing, feature extraction, encrypted file han- dling, model training, and user friendly interface development. To ensure robustness, the system was trained and tested on a combination of publicly available datasets and two cus- tom designed encrypted datasets created specifically for this project. These datasets simulated real world challenges, such as privac -preserving formats and subtle bias patterns, allowing the model to generalize its fairness assessment capabilities.

The results confirmed that the system can reliably flag datasets as fair or biased based on class distribution and fairness metrics such as demographic parity and equal op- portunity. Furthermore, the modular architecture and clean interface make the system accessible for both technical and non technical users, enhancing its potential for widespread adoption in domains where ethical AI deployment is essential. Challenges encountered during development, such as model inconsistencies, file handling errors, and frontend backend integration, were addressed through iterative testing and code optimization. These efforts not only strengthened the system’s performance but also contributed to a deeper understanding of the technical and ethical dimensions of bias in AI. The project successfully demonstrates a viable approach to bias and fairness detection at the dataset level. It underscores the importance of integrating fairness checks early in the AI development lifecycle and provides a foundation for future research and enhancements. With further refinement and ex- pansion to support larger and more diverse datasets, the system holds significant promise for supporting equitable AI practices across various sectors. In the future, we plan to support more dataset types, including text and tabular data. We also aim to include more advanced fairness metrics like equal opportunity and individual fairness. Additionally, we want to add suggestions to fix imbalanced datasets and package the system into a fully deployable web app or API, so others can integrate it into their own AI projects[19].

REFERENCES

M Hussain, JJ Bird, DR Faria - . . . : Contributions Presented at the 18th UK . . . , 2019 - Springer

S Fabbrizzi, S Papadopoulos, E Ntoutsi. . . - Computer Vision and . . . , 2022 - Elsevier

Nazir, T., Ahmed, R. H., Hussain, M., Zahid, S. (2023, October). Trans- forming Blood Donation Processes with Blockchain and IoT Integration: A augmented Approach to Secure and Efficient Healthcare Practices. In 2023 International Conference on IT and Industrial Technologies (ICIT) (pp. 1-8). IEEE.

Ahmed, Rana Hassam, et al. ”ENHANCING AUTONOMOUS VEHI- CLE SECURITY THROUGH ADVANCED ARTIFICIAL INTELLI- GENCE TECHNIQUES.” (2024).

Hussain, Majid. ”Blockchain-Based Supply Chain Management in Healthcare.” AI and Blockchain Applications for Privacy and Security in Smart Medical Systems (2025): 107.

Buolamwini, J.; Gebru, T. Gender shades: Intersectional accuracy dis- parities in commercial gender classification. In Proceedings of the 1st Conference on Fairness, Accountability and Transparency, New York, NY, USA, 23–24 February 2018; pp. 77– 91.

Dastin, J. Amazon scraps secret AI recruiting tool that showed bias against women. In Ethics of Data and Analytics; Auerbach Publications: Boca Raton, FL, USA, 2018; pp. 296–299.

Solon Barocas, Moritz Hardt, and Arvind Narayanan. Fairness in Ma- chine Learning.

Barocas, S., Selbst, A. D. (2016). Big Data’s Disparate Impact.

Angwin, J., Larson, J., Mattu, S., Kirchner, L. (2016). Machine Bias. ProPublica.

Hardt, M., Price, E., Srebro, N. (2016). Equality of Opportunity in Supervised Learning. Advances in Neural Information Processing Systems.

Barocas, S., Hardt, M., Narayanan, A. (2016). Fairness in Machine Learning.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., Galstyan, A. (2021). A Survey on Bias and Fairness in Machine Learning. ACM Computing Surveys.

Hardt, M., Price, E., Srebro, N. (2016). Equality of Opportunity in Supervised Learning. Advances in Neural Information Processing Systems.

Corbett-Davies, S., Goel, S. (2018). The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning. Communications of the ACM.

Zemel, R., Wu, Y., Swersky, K., Pitassi, T., Dwork, C. (2013). Learning Fair Representations. International Conference on Machine Learning.

Feldman, M., Friedler, S. A., Moeller, J., Scheidegger, C., Venkatasub- ramanian, S. (2015). Certifying and Removing Disparate Impact. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining

Angwin, J., Larson, J., Mattu, S., Kirchner, L. (2016). Machine Bias. ProPublica.

Obermeyer, Z., Powers, B., Vogeli, C., Mullainathan, S. (2019). Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations. Science.

A Singh, N Thakur, A Sharma - 2016 3rd international . . . , 2016 - ieeexplore.ieee.org Social Research: An International Quarterly Johns Hopkins University Press Volume 86, Number 2, Summer 2019

https://arxiv.org/abs/2408.16040

M Sholihin, R Pike - Accounting and Business Research, 2009 - Taylor Francis