Exploring Fusion Techniques in Multimodal AI-Based Recruitment: Insights from FairCVdb

Swati Swati[1,*], Arjun Roy[1,2] and Eirini Ntoutsi[1]

[1]Research Institute CODE, University of the Bundeswehr Munich, Germany
[2]Institute of Computer Science, Freie University Berlin, Germany

Abstract

The increasing use of decision-making algorithms has led to concerns about transparency and potential discrimination, especially when impacting specific social groups. Investigating how sensitive elements and internal biases influence current multimodal algorithms is crucial. We investigate the fairness and bias implications of multimodal AI-based recruitment systems using the FairCVdb dataset. We evaluate the effectiveness of early and late fusion techniques in a variety of settings, including unbiased ideal conditions and biased real-world scenarios with multimodal biases. Our findings suggest that while late fusion performs well under unbiased conditions, it may exacerbate biases in real-world scenarios. Early fusion, on the other hand, generally produces more equitable results, especially when modality selection is carefully considered. Future research could explore alternative fusion strategies and incorporate additional fairness constraints to improve fairness in algorithmic decision-making. For code and additional insights, visit: https://github.com/Swati17293/Multimodal-AI-Based-Recruitment-FairCVdb

Keywords

Computer vision, Natural language processing, Multimodal, Fairness, Algorithmic Fairness, Multimedia Fusion

1. Introduction

The increasing popularity of decision-making algorithms in society has raised concerns about transparency and the potential for discrimination [1]. Of particular concern is the capability of these algorithms to perpetuate biases or make unfair decisions, particularly when affecting specific social groups [2]. Addressing these concerns requires an investigation into how sensitive elements and internal biases in the data influence current multimodal algorithms, which rely on diverse sources of information [3].

We aim to investigate the fairness and bias implications of multimodal AI-based recruitment systems. Leveraging the FairCVdb [4] dataset as a testbed, we delve into the intricate interplay between modality contribution and fusion techniques in automated recruitment processes. We use FairCVdb for our evaluations as it offers rich and diverse data composed of images, text, and structured data specifically tailored to assess the fairness and bias aspects of AI-driven recruitment algorithms. This dataset contains synthetic profiles intentionally scored with gender

EWAF'24: European Workshop on Algorithmic Fairness, July 01–03, 2024, Mainz, Germany
*Corresponding author.
swati.swati@unibw.de (S. Swati); arjun.roy@unibw.de (A. Roy); eirini.ntoutsi@unibw.de (E. Ntoutsi)
0000-0002-7637-6640 (S. Swati); 0000-0002-4279-9442 (A. Roy); 0000-0001-5729-1003 (E. Ntoutsi)

and ethnicity biases to help investigate the possibility of unfair decision-making.

In multimodal learning, fusion techniques play pivotal roles in integrating information from different modalities [5]. In this study, we focus exclusively on early and late fusion techniques for analysis due to their straightforward interpretability and widespread usage in multimodal AI systems [6, 7]. Early fusion typically involves concatenating features from different modalities at an early stage to create a single feature vector, resulting in a unified representation of the data [7]. This approach simplifies the training process and can capture interactions between modalities effectively [8]. However, early fusion may struggle to handle the heterogeneity of features across modalities [6]. On the other hand, late fusion processes each modality separately before combining their outputs at a later stage. This technique offers flexibility by allowing different processing pathways for individual modalities and accommodating variations in data characteristics and model architectures [9]. Late fusion can capture modality-specific patterns more accurately but may overlook interactions between modalities present at lower levels of representation [10]. By concentrating on these two fusion strategies, our objective is to provide clear insights into their effectiveness and suitability for addressing fairness and bias concerns in automated recruitment processes.

2. Experimental Setup

2.1. Dataset

The FairCVdb dataset [4] comprises of 24,000 synthetic resume profiles, each featuring demographic characteristics (gender and ethnicity), visual data (a facial image), textual data (a short biography), and tabular data (seven common resume attributes such as education and experience). Each profile has been generated based on two gender categories and three ethnic categories. Demographic attributes related to gender and ethnicity determine the facial image, name, and pronouns in the short biography. Profiles in the dataset are scored either blindly, leading to neutral scores, or with a penalty factor applied to specific individuals within a demographic group, resulting in biased scores. See [11] for more details. The design of the scoring system permits certain groups to receive lower scores than others, despite sharing the same competencies. This setup simulates scenarios where cognitive biases, introduced by humans, protocols, or automated systems, influence the decision-making process.

2.2. Evaluation Metrics

For evaluation, we follow the metrics employed in [11]. We use Mean Absolute Error (MAE) to measure accuracy and Kullback-Leibler (KL) divergence between demographic distributions to measure biases. In the gender case, we compare the score distributions for males and females, while in the ethnicity setup, we perform pairwise comparisons and report the average divergence.

3. Results

To predict the scores given the candidate resumes, we train and evaluate the recruitment model as described by the authors in [11]. We modify and expand the testbed to assess early- and late-fusion techniques using all possible combinations of the three modalities.

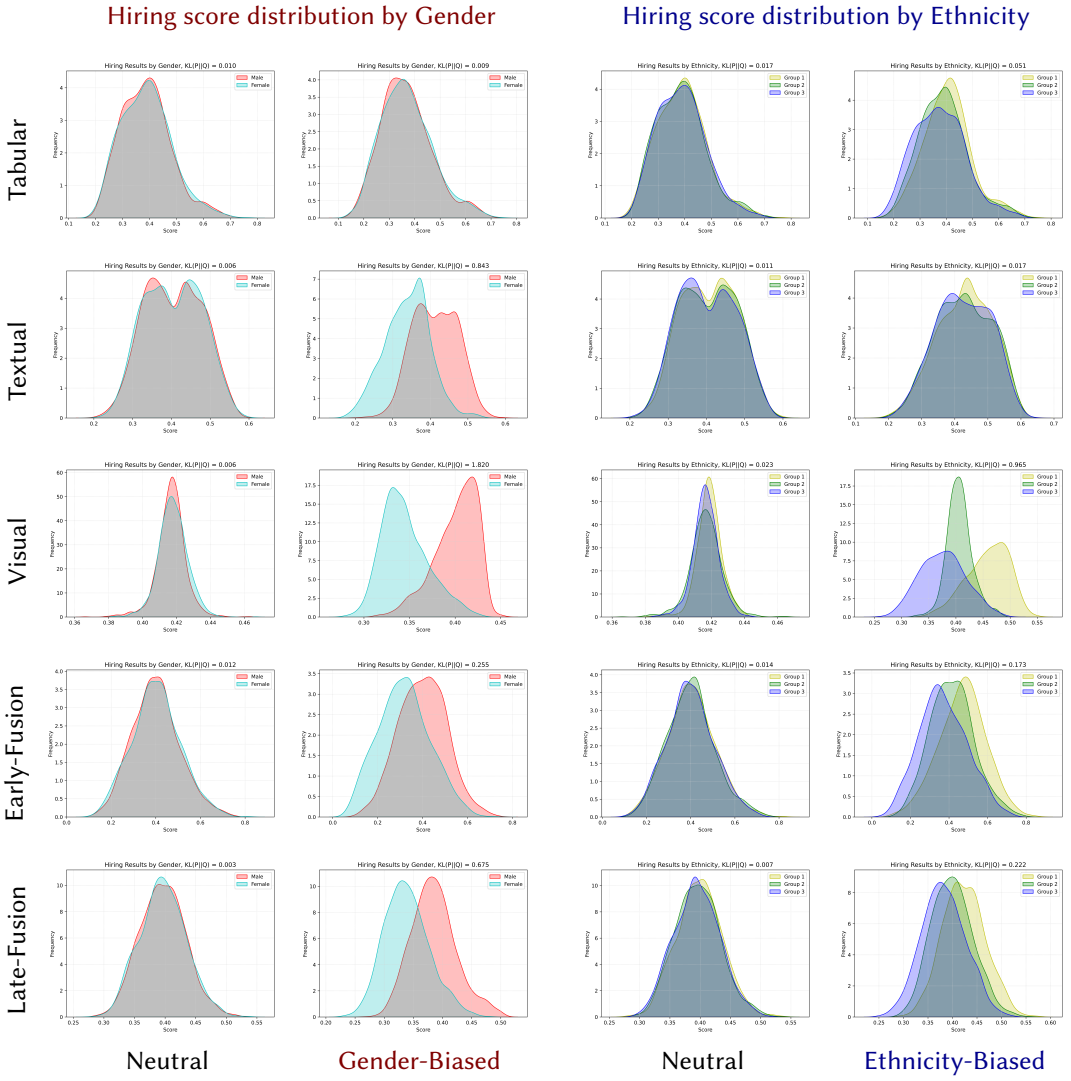

Figure 1: KL-divergence between score distributions across Gender and Ethnicity demographics. "Neutral" refers to models trained on unbiased neutral scores. "Gender/Ethnicity-Biased" models are trained on Gender/Ethnicity biased scores. The terms "Tabular", "Textual", and "Visual" denote models trained exclusively on tabular data, textual data, and visual data, respectively. The terms "Early-Fusion" and "Late-Fusion" denote models that integrate all three modalities at the data representation level and the decision level, respectively. Lower KL-divergence signifies closer alignment between distributions across various demographics, implying more fairness in hiring.

In our experiments, we investigated both unbiased ideal world conditions and real-world scenarios in gender/ethnicity-biased setups as described by authors in [11] (ref. Figure 1). We discovered that in an unbiased ideal world, late fusion emerged as the superior approach for achieving fair outcomes. Late fusion allowed the model to learn from unbiased expert models on individual modalities, facilitating a fair accumulation of decisions. This method enabled the model to aggregate information from diverse modalities without introducing unwanted correlations, leading to more equitable decisions. On the other hand, early fusion, despite combining representations, risked incorporating unwanted correlations, potentially resulting in biased outcomes at the expense of decision accuracy.

Conversely, in biased real-world scenarios characterised by varied biases across modalities, late fusion may exacerbate biases by independently learning biased models for each modality, which can cumulatively impact decision fairness. Early fusion, however, offered greater flexibility in learning optimal combinations of representations. Although early fusion may still capture unwanted correlations, particularly when modalities are not carefully selected, our findings suggest that it generally yields fairer outcomes when integrating multiple modalities such as text, tables, and images. This trend held true across gender- and ethnicity-biased setups, indicating the robustness of early fusion in mitigating biases while enhancing decision accuracy. Notably, minimal differences in Mean Absolute Error (MAE) between early and late fusion techniques were observed across setups, with a disparity of 0.057 in neutral and gender-biased scenarios and a slight reduction to 0.049 in ethnicity-biased conditions.

In general, utilising multimodal data can enhance performance and reduce bias compared to relying on a single modality. However, blindly fusing all modalities may not be effective in every scenario. In our experiments, models trained solely on tabular data outperformed those using late fusion of multiple modalities in terms of both accuracy and fairness. In contrast, fusing textual and visual modalities, regardless of the fusion technique, provided additional context and rich information, resulting in more accurate and fairer outcomes than using these modalities individually. The dataset used can also significantly influence these findings by leveraging the diversity and biases present in the data. This underscores the importance of assessing robustness across multiple datasets and fusion strategies.

Our observations lead us to contemplate the potential of mid-fusion strategies for future exploration. While the success of early fusion with in-processing is evident, we speculate that mid-fusion strategies could offer valuable insights. Mid-fusion provides the opportunity to strategically select, combine, and weight modalities, potentially enhancing both fairness and accuracy in decision-making.

4. Conclusion

In this study, we use the FairCVdb to investigate the bias implications of early- and late-fusion strategies for multimodal AI-based recruitment. We use KL divergence to assess the biases of fusion outcomes in various demographic groups. Our findings suggest that late fusion performs well under unbiased ideal conditions, allowing for the fair accumulation of decisions without introducing unwanted correlations. However, in real-world scenarios with biases, late fusion may exacerbate biases by learning biased models independently for each modality. The results further demonstrate that early fusion generally produces more equitable outcomes, especially

when modalities are carefully selected. Based on the findings, we speculate that mid-fusion strategies may improve both fairness and accuracy through strategic selection and a combination of modalities.

In the future, there exist numerous potential areas for further investigation. The study of different fusion strategies, in addition to traditional early and late fusion techniques, holds the potential to provide new insights into improving fairness in multimodal AI systems. Understanding the risks inherent in using simulated or synthetic data is also crucial for ensuring fairness, transparency, and effectiveness in automated hiring processes. Thus, examining the generalisability of these findings to more datasets and domains than just hiring has the potential to broaden the study's impact and relevance.

Acknowledgments

This research work is funded by the European Union under the Horizon Europe MAMMOth project, Grant Agreement ID: 101070285. UK participant in Horizon Europe Project MAMMOth is supported by UKRI grant number 10041914 (Trilateral Research LTD).

References

[1] C. Schumann, J. Foster, N. Mattei, J. Dickerson, We need fairness and explainability in algorithmic hiring, in: International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS), 2020.

[2] M. Raghavan, S. Barocas, J. Kleinberg, K. Levy, Mitigating bias in algorithmic hiring: Evaluating claims and practices, in: Proceedings of the 2020 conference on fairness, accountability, and transparency, 2020, pp. 469–481.

[3] B. M. Booth, L. Hickman, S. K. Subburaj, L. Tay, S. E. Woo, S. K. D'Mello, Bias and fairness in multimodal machine learning: A case study of automated video interviews, in: Proceedings of the 2021 International Conference on Multimodal Interaction, 2021, pp. 268–277.

[4] A. Pena, I. Serna, A. Morales, J. Fierrez, Bias in multimodal ai: Testbed for fair automatic recruitment, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 28–29.

[5] Z. Xue, R. Marculescu, Dynamic multimodal fusion, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 2574–2583.

[6] K. Gadzicki, R. Khamsehashari, C. Zetzsche, Early vs late fusion in multimodal convolutional neural networks, in: 2020 IEEE 23rd international conference on information fusion (FUSION), IEEE, 2020, pp. 1–6.

[7] L. M. Pereira, A. Salazar, L. Vergara, A comparative analysis of early and late fusion for the multimodal two-class problem, IEEE Access (2023).

[8] G. Barnum, S. Talukder, Y. Yue, On the benefits of early fusion in multimodal representation learning, arXiv preprint arXiv:2011.07191 (2020).

[9] L. M. Pereira, A. Salazar, L. Vergara, On comparing early and late fusion methods, in: International Work-Conference on Artificial Neural Networks, Springer, 2023, pp. 365–378.

[10] K. Bayoudh, R. Knani, F. Hamdaoui, A. Mtibaa, A survey on deep multimodal learning for computer vision: advances, trends, applications, and datasets, The Visual Computer 38 (2022) 2939–2970.

[11] A. Peña, I. Serna, A. Morales, J. Fierrez, A. Ortega, A. Herrarte, M. Alcantara, J. Ortega-Garcia, Human-centric multimodal machine learning: Recent advances and testbed on ai-based recruitment, SN Computer Science 4 (2023) 434.