Fairness in EEG-Based Depression Detection

Annisaa Fitri Nurfirdausi
annisaa.nurfirdausi@studio.unibo.it

May 2025

Abstract

As artificial intelligence becomes increasingly integrated into mental health diagnostics, ensuring fairness and transparency is critical—particularly in sensitive applications such as depression detection. This project explores fairness-aware deep learning using raw EEG signals from the MODMA (Multimodal Open Dataset for Mental-disorder Analysis) dataset [2]. Three deep neural architectures are developed and evaluated: a Convolutional Neural Network (CNN) as a baseline model, a CNN combined with Long Short-Term Memory (CNN-LSTM), and a CNN combined with Gated Recurrent Units and Attention mechanism (CNN-GRU-Attention). To address potential biases, five mitigation techniques are applied across three stages: preprocessing, in-processing, and post-processing. The models are evaluated using both performance metrics (accuracy, precision, recall, F1-score) and fairness metrics (disparate impact, statistical parity, equal opportunity, average odds, and equalized accuracy).

1 Objectives

- Preprocess raw EEG time-series data from the MODMA dataset using filtering and segmentation techniques to enhance signal quality.
- Design and implement three deep learning architectures (CNN, CNN-LSTM, and CNN-GRU-Attention).
- Integrate fairness considerations into model development by applying mitigation techniques across the pre-processing, in-processing, and post-processing stages.
- Identify which model and fairness strategy combination achieves the best balance between predictive performance and demographic fairness. .

2 Introduction

Depression, also referred to as major depressive disorder, is a widespread mental health condition that significantly affects individuals and society. According to the World Health Organization (WHO), approximately 280 million people worldwide experience depression, making it the fourth leading cause of death among individuals aged 15 to 29 years. Experts predict that by 2030, it will become the second leading cause of disease burden [10]. Additionally, the COVID-19 pandemic led to a 27.6% rise in global cases of major depressive disorder [14].

Currently, the depression screening involves physician-administered interviews using questionnaires like the Physical Health Questionnaire Depression Scale (PHQ) [13]. This approach heavily depends on the physician's subjective assessment. Additionally, some patients may hesitate to openly discuss their symptoms due to the stigma associated with depression [15].

EEG has proven to be useful in diagnosing neurological, cognitive, and psychological disorders [11]. Moreover, EEG equipment is relatively easy to maintain and its non-invasive nature allows for monitoring brain activity in real-time and across extended periods [3]. Recent research studies have explored depression detection using machine learning and deep learning techniques. Features like kurtosis, skewness, peak-to-peak, variance, mean, power spectral density, entropy were usually used to be fed into machine learning

algorithms such as K-Nearest Neighbor, Random Forest, Logistic Regression, Support Vector Machine [6] [8] [4]. In recent years, deep learning has become more popular than traditional machine learning in biomedical signal processing. Unlike traditional methods, deep learning uses multiple hidden layers to automatically extract important features from data [17]. It is also better at handling complex, high-dimensional, and nonlinear data, which improves diagnostic accuracy. For example, in 2018, Acharya et al. used a 13-layer CNN to detect depression and achieved high accuracy [1]. In 2022, Wang et al. developed a method using CNN, GRU, and an attention mechanism for depression detection [16].

At the same time, concerns about machine learning (ML) bias are increasingly coming to light. Although algorithmic decision-making can help reduce some of the biases inherent in human judgment, it also carries the risk of producing unfair outcomes when models are trained on biased datasets or suffer from inaccuracies [12]. This can lead to discriminatory treatment of individuals based on sensitive characteristics like race, gender, disability, or political and sexual orientation. The field of machine learning fairness focuses on preventing model decisions from being unjustly influenced by such attributes. For instance, a fair model used to screen job applicants should not favor candidates based on their age. Likewise, a medical diagnosis system should not produce different results for patients solely due to their ethnicity or gender. To the best of our knowledge, the application of bias mitigation techniques in Mental Health, specifically with EEG-based remains relatively underexplored. A notable exception is the work by Kwok et al., who applied bias mitigation strategies to the MODMA dataset using deep learning models such as CNN, LSTM, and GRU [9] which mostly inspired this study. This project aims to contribute to the field by: (1) implementing additional bias mitigation methods, including Equalized Odds Postprocessing, (2) Utilizing the AIF360 library for computing fairness metrics and applying various bias mitigation strategies.

3 Methodology

3.1 Dataset

The MODMA dataset includes 53 subjects with speech data and 52 with EEG recordings [2]. On this project, we are only interested on EEG data. As shown in Table 5.5, the dataset, while relatively small, maintains a balanced distribution between depressed and non-depressed participants. However, both groups contain a higher number of male subjects compared to female participants.

Group	**EEG Subjects**	**Gender (F/M)**
Depressed	23	18 / 29
Non-Depressed	29	18 / 40
Total	52	36 / 69

Table 1: EEG statistics of the MODMA dataset.

3.2 EEG data proprocessing

A finite impulse response (FIR) filter with a frequency range of 0.5–50 Hz was applied, as this range encompasses the majority of depression-related EEG activity. Notch filter is applied to remove power line noise 50 Hz. Following this, the EEG signals were re-referenced to the average electrode, a standard technique in EEG processing used to reduce background noise. Finally, the preprocessed signals were segmented into 8-second epochs.

3.3 Models

3.3.1 CNN

The `CNN` model implemented on this project is a lightweight 1D convolutional neural network designed for sequence classification tasks. It begins with a 1D convolutional layer with 32 filters and a kernel size of 3,

followed by batch normalization and max pooling to reduce dimensionality. This is followed by a second convolutional layer with 64 filters, again followed by batch normalization and max pooling. ReLU activation is applied after each convolutional layer. After the convolutional stages, the output is flattened and passed through a dropout layer to mitigate overfitting. Finally, a fully connected linear layer maps the features to the number of output units.

3.3.2 CNN-LSTM

The `CNN-LSTM` model implemented on this project combines convolutional and recurrent layers for sequence modeling tasks. It starts with a 1D convolutional layer with 32 filters, followed by batch normalization, ReLU activation, and max pooling to extract local features and reduce sequence length. The output is then transposed and fed into an LSTM layer to capture temporal dependencies. A dropout layer is applied to the last hidden state of the LSTM to prevent overfitting, and a final fully connected layer maps the result to the output dimension.

3.3.3 CNN-GRU-Attention

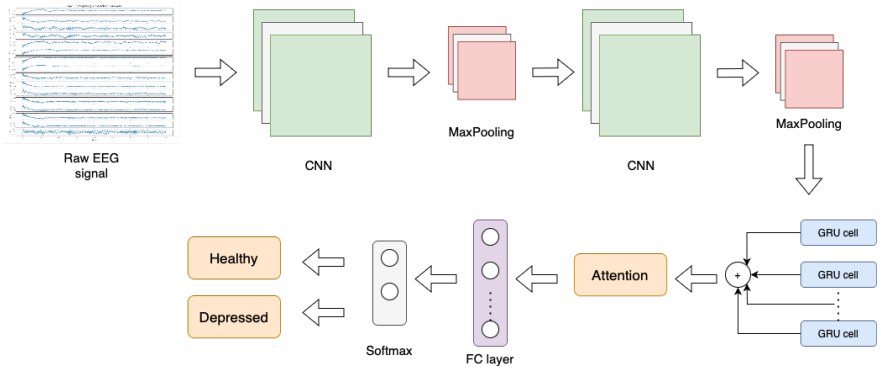


Figure 1: CNN-GRU-Attention architecture

The `CNN-GRU-Attention` model implemented on this project integrates convolutional, recurrent, and attention mechanisms for enhanced sequence modeling as shown on Figure 1 It begins with two 1D convolutional layers (with 64 and 128 filters, respectively), each followed by ReLU activation and max pooling to extract hierarchical features. The resulting representations are passed to a bidirectional GRU to capture temporal dependencies in both directions. An attention mechanism is applied to the GRU outputs to learn a weighted representation of the sequence, focusing on the most informative time steps. Finally, dropout is applied before passing the attended context vector through a fully connected layer for classification.

4 Evaluation Metrics

4.1 Performance Metrics

To evaluate the classification performance of the model, we use the following standard metrics:

- **Accuracy**: Measures the overall proportion of correctly classified instances.
- **Precision**: The ratio of true positives to the total predicted positives, reflecting the quality of positive predictions.
- **Recall (Sensitivity)**: The ratio of true positives to all actual positives, indicating the model's ability to identify positive instances.
- **F1-Score**: The harmonic mean of precision and recall, useful when the class distribution is imbalanced.

4.2 Fairness Metrics

To assess the fairness of the model across different demographic groups, we adopt several fairness metrics implemented via the `aif360` package, specifically using `aif360.metrics.ClassificationMetric`. The metrics evaluated include:

- **Disparate Impact (DI)**: Measures the ratio of favorable outcomes received by the unprivileged group to that received by the privileged group. A value close to 1 indicates fairness.
- **Statistical Parity Difference (SPD)**: The difference in the probability of favorable outcomes between unprivileged and privileged groups. A value close to 0 indicates fairness.
- **Equal Opportunity Difference (EOD)**: The difference in true positive rates between unprivileged and privileged groups. Lower absolute values imply more fairness.
- **Average Odds Difference (AOD)**: The average of the differences in false positive rates and true positive rates between groups. Lower values suggest higher fairness.
- **Equalized Accuracy Difference (EAD)**: Measures the difference in accuracy across groups. Small differences indicate equal treatment.

4.3 Bias Mitigation Strategies

To mitigate bias and promote fairness in model predictions, we employed a variety of strategies categorized into pre-processing, in-processing, and post-processing methods:

- **Data Augmentation (Mixup)**: A technique generating new synthetic samples, increasing representation of minority group samples. In this project, we implemented Mixup, which combines two input samples and their labels using a convex combination. The mixing strength is controlled by the parameter *alpha* (α), which determines the balance between the two input samples. A small α results in a stronger reliance on one sample, while a larger α leads to a more even mix. On this project, we implemented $alpha = 0.2$.
- **Reweighting**: An method where each training sample is assigned a weight based on the joint and marginal probabilities of labels and sensitive attributes, such as gender in this case. The weights are computed using a *beta list* that adjusts the weights for each sample according to the relative frequency of each unique combination of label and sensitive attribute.
- **Massaging**: A label modification strategy that equalizes the distribution of positive and negative labels across sensitive groups. It flips labels in a controlled way, converting 'Depressed' to 'Healthy' for favored groups and vice versa for deprived groups—based on group proportions.
- **Regularization**: A custom regularized loss function that penalizes disparities in True Positive Rate (TPR) and False Positive Rate (FPR) between groups. This method directly incorporates fairness metrics, such as Equal Opportunity and Equalized Odds, into the training process. Specifically, the regularization term focuses on minimizing the difference between TPR and FPR across different groups. These differences are then added to the original loss function to create a fairness-aware objective.
- **Reject Option Classification (ROC)**: A post-processing technique that modifies predictions within a confidence margin defined by lower and upper classification thresholds (set to 0.4 and 0.7 in our implementation). Within this uncertain region near the decision boundary, the method favors the unprivileged group to mitigate bias. The midpoint of these thresholds, $\tau = 0.55$, approximates the threshold suggested by Kamiran et al., and can be adjusted as needed [7].
- **Equalized Odds Postprocessing**: A post-processing method designed to mitigate bias by adjusting the predictions of a trained classifier to satisfy the equalized odds fairness criterion. This criterion requires that the true positive rate (TPR) and false positive rate (FPR) be approximately equal across privileged and unprivileged groups. The method works by solving an optimization problem that minimally alters the classifier's predictions to reduce disparities in error rates while aiming to preserve overall predictive accuracy [5].

5 Experiment

5.1 Experimental Setup

All experiments were conducted on a local machine equipped with an Apple MacBook M1 chip and 8 GB RAM. The dataset was split at the patient level segmented-EEG data using a 70:15:15 ratio for training, validation, and test sets respectively.

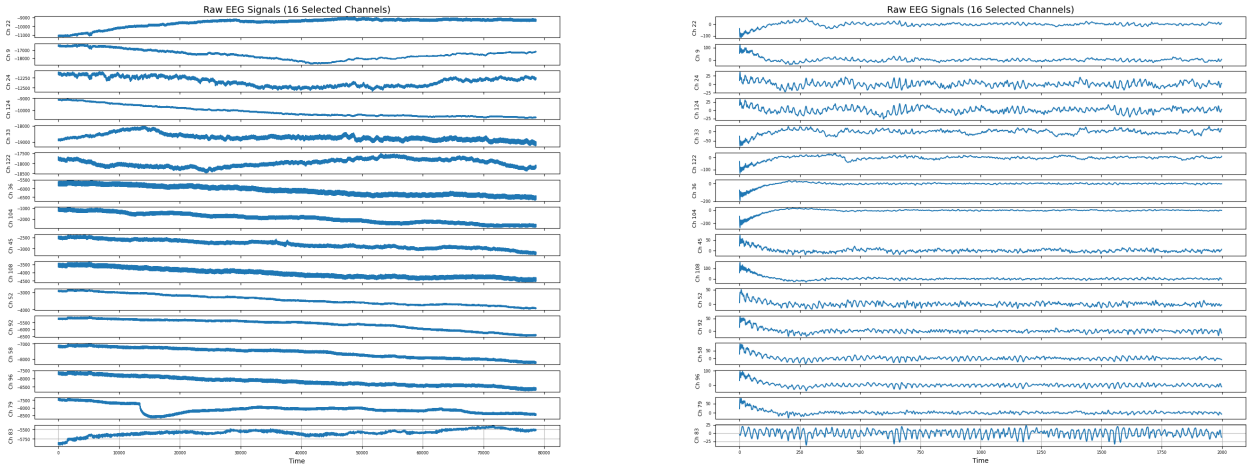
We trained our models using the PyTorch framework. Each model was trained for up to 70 epochs with a batch size of 32. The optimizer used was Adam with a learning rate of 0.0001 and a weight decay of 1×10^{-5} to prevent overfitting. The loss function was categorical cross-entropy ('CrossEntropyLoss'), with optional instance reweighting applied via custom sample weights for fairness experiments.

We incorporated early stopping with a patience of 10 epochs—training terminated if the validation accuracy did not improve for 10 consecutive epochs. Additionally, we employed learning rate scheduling using 'ReduceLROnPlateau', which halves the learning rate if the validation accuracy plateaued for more than 3 epochs. The best model parameters (based on validation accuracy) were retained and restored after training.

5.2 Experimental Result

5.2.1 Data preprocessing

After applying data preprocessing steps—including filtering and segmenting—the resulting processed signals are shown in Figure 2a. These preprocessing steps help standardize the signal quality and prepare the data for further analysis.



(a) Raw EEG Signal

(b) Processed EEG Signal

Figure 2: Comparison of Raw and Processed EEG Signals

The cleaned and segmented data are then used to train three different models: a standalone CNN, a combined CNN-LSTM, and a hybrid CNN-GRU-Attention model. Initially, these models are trained without any bias mitigation techniques which we will call these baseline models. The corresponding performance metrics are presented in Table 2. Among the models, the CNN-GRU-Attention model (`model_cnn_gru`) achieves the highest predictive performance.

Table 2: Performance and Fairness Metrics for Baseline Models

Model	Acc	Prec	Rec	F1	DI	SPD	EOD	FPRD	AOD	EAD
model_cnn	0.791	0.781	0.746	0.763	0.655	-0.226	-0.109	-0.195	-0.152	0.026
model_cnn_gru	0.916	0.898	0.918	0.908	0.795	-0.120	0.043	-0.060	-0.009	0.051
model_cnn_lstm	0.811	0.787	0.799	0.793	0.702	-0.182	-0.016	-0.189	-0.103	0.078

Acc: Accuracy, **Prec**: Precision, **Rec**: Recall, **F1**: F1 Score, **DI**: Disparate Impact, **SPD**: Statistical Parity Difference, **EOD**: Equal Opportunity Difference, **FPRD**: False Positive Rate Difference, **AOD**: Average Odds Difference, **EAD**: Equalized Accuracy Difference.

To evaluate fairness, we define key elements of fairness-aware modeling. The *protected attribute* in this study is gender, as we aim to ensure parity across genders. The *privileged group* is defined as the male group, given their statistical dominance in the dataset. The *favorable label* is associated with the identification of depression, as detecting it leads to treatment, which benefits the individual. In terms of *case type*, this scenario is considered *assistive*, as failing to identify a depressed subject may prevent necessary treatment.

While `model_cnn_gru` excels in predictive performance, it performs poorly on certain fairness metrics—such as Disparate Impact—indicating a substantial imbalance.

5.2.2 Bias Mitigation

To address the mentioned-issues, we explore bias mitigation strategies, beginning with the application of data augmentation, speficially Mixup method. The data number before and after augmentation is shown on Figure 3. In Mixup, the Beta distribution parameter to alpha = 0.2 is set, which encourages sample interpolation coefficients to lie near 0 or 1. This setting generates synthetic examples that remain close to real samples, which preserving the structure of the EEG/audio features while still introducing useful regularization. Empirically, small alpha values like 0.2 have been shown to improve performance on some models as shown in figure Figure 4.It is shown that by augmenting data can help in increasing Disparity Impact to be close to 1 for Model CNN and Model CNN-LSTM. Additionally other metrics, such as Statistial Parity Difference, False Positive Rate Difference, and Average Odds Difference has shown some decreament to 0. However, we noticed that Data Augmentation does not work well on Model CNN-GRU as shown some metrics are not getting better.

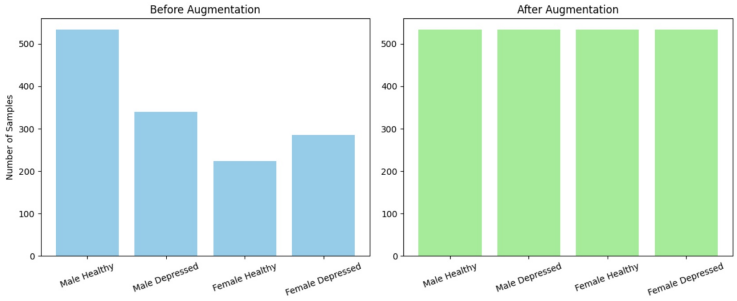


Figure 3: Data Augmentation

Figure 4: Data Augmentation

Apart from Data Augmentation, we implemented Massaging. We applied the massaging technique by first identifying the favored and deprived gender groups based on the observed positive label rates. The favored group is the gender with a higher proportion of positive labels (depressed), while the deprived group has a lower proportion. To close this disparity, we relabel a small number of samples: flipping a subset of Depressed labels to Healthy in the favored group, and vice versa in the deprived group. The number of relabeled samples is determined by the absolute difference in positive label rates between the two groups, scaled by the size of the smaller group. In our case, we relabeled 87 samples. The data after massaging can be shown on Figure 5. The result comparison between baseline models are shown on Figure 6. Overall, massaging labels gave us notable better results on some fairness metrics, such as Disparate Impact and Statistical Parity Difference for all models.

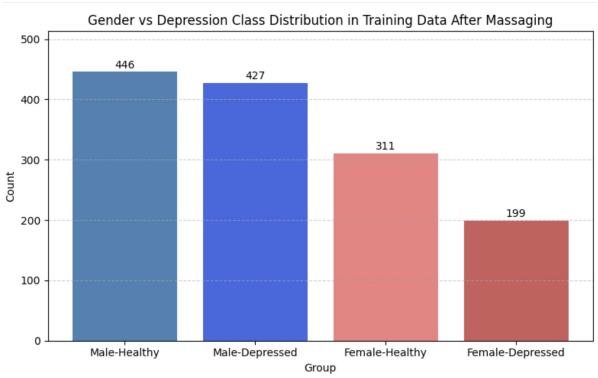


Figure 5: Data After Massaging

The next bias mitigation strategy that we implemented is Reweighing. We apply the Reweighing technique by assigning different importance weights to training samples based on their group membership and label. Specifically, we define male as the privileged group and female as the unprivileged group. Using the AIF360 framework, we construct a BinaryLabelDataset containing the labels and gender attributes, and then apply the Reweighing algorithm. This process computes instance weights (beta_list) such that samples from disadvantaged groups with unfavorable outcomes receive higher weights during training. The comparison result with the baseline models can be shown on Figure 7. It is shown that Reweighing gave better result

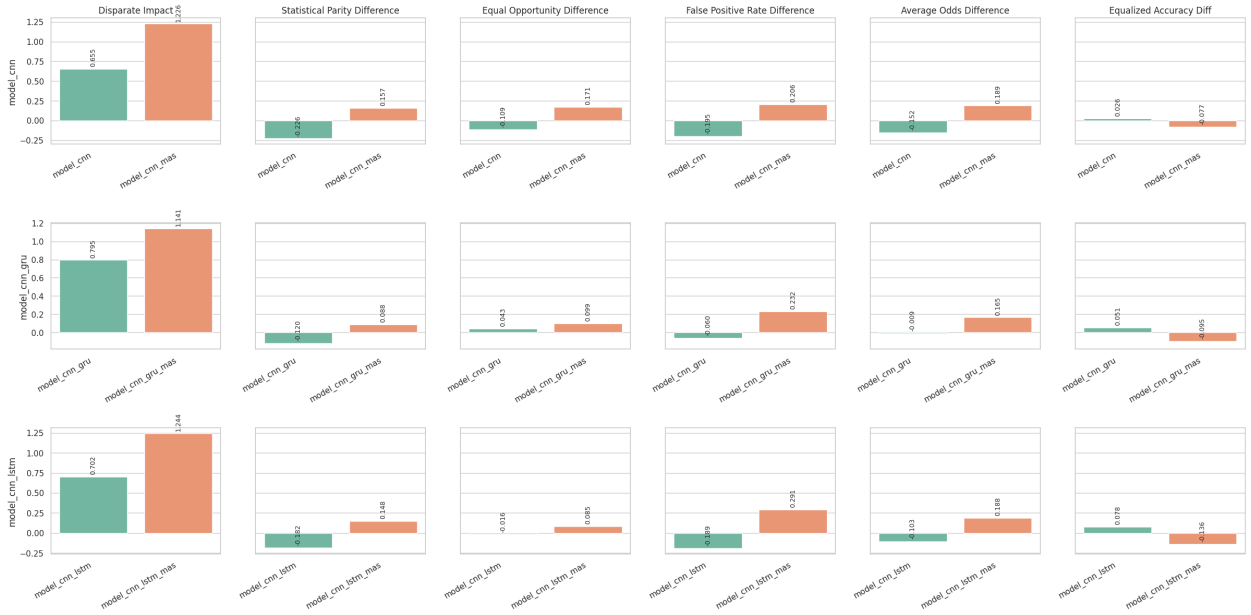


Figure 6: Massaging Result

on Disparate Impact and Statistical Parity Difference for all models. Hoever, the most notable result can be seen on Model CNN-LSTM where all of the metrics are showing better result after applying Reweighing.

The fourth strategy that we implemented was in-processing Regularization. On this method, we penalize the model for unequal true positive rates (TPR) and false positive rates (FPR) across male and female subjects. The regularized loss function combines the standard cross-entropy loss with two fairness penalty terms: the absolute difference in TPR (ΔTPR) and in FPR (ΔFPR) between genders. These are weighted by hyperparameters λeopp and λeodd, respectively, controlling the trade-off between accuracy and fairness. In our experiments, we set both λ values to 0.5 to moderately balance these objectives. The result comparison with the baseline models can be shown on Figure 8. As what we observed, regularization did not give us notable results on all metrics.

The next strategies that we implemented was post-processing: Reject Optional Classifier (ROC) and Equalized Odds Postprocessing(EOP). We define a low threshold (low_class_thresh=0.4) and a high threshold (high_class_thresh=0.7) to delimit this "reject option" region. Predictions with confidence scores between these thresholds are candidates for label flipping to reduce bias. The comparison results with the baseline results are shown on Figure 9. As we can observe, the EOP gave us better result with respect to ROC. For disparate impact, EOP gave value closer to 1 for CNN model and CNN-LSTM model. For Equal Opportinity Difference and False Positive Rate Difference metric, EOP works well for all the models giving the value really closer to 0.

5.2.3 Fairness and Accuracy

Some of the bias mitigation strategies implemented above demonstrated noticeable improvements in fairness metrics. However, since there is often a trade-off between fairness and accuracy, it is also important to evaluate how these strategies affect overall model accuracy. The summarized results, highlighting both fairness improvements and accuracy impacts, are presented in the following Table 3. Previously, we noticed that data augmentation gave noticeable result in terms of fairness metrics. However, if we looked into the prediction metrics, it can be seen there are some slight decrement in accuracy, for CNN-LSTM model previously it was 0.818 to 0.791. The same with reweighting method, where the accuracy drop to 0.684. However there are some methods, like Regularization we noticed give good impact both to prediction metrics and fairness metrics. With CNN-GRU model, previous accuracy was 0.916 increasing to 0.949 with fairness metric like Disparate Impact previously was 0.795 increasing to 0.834.

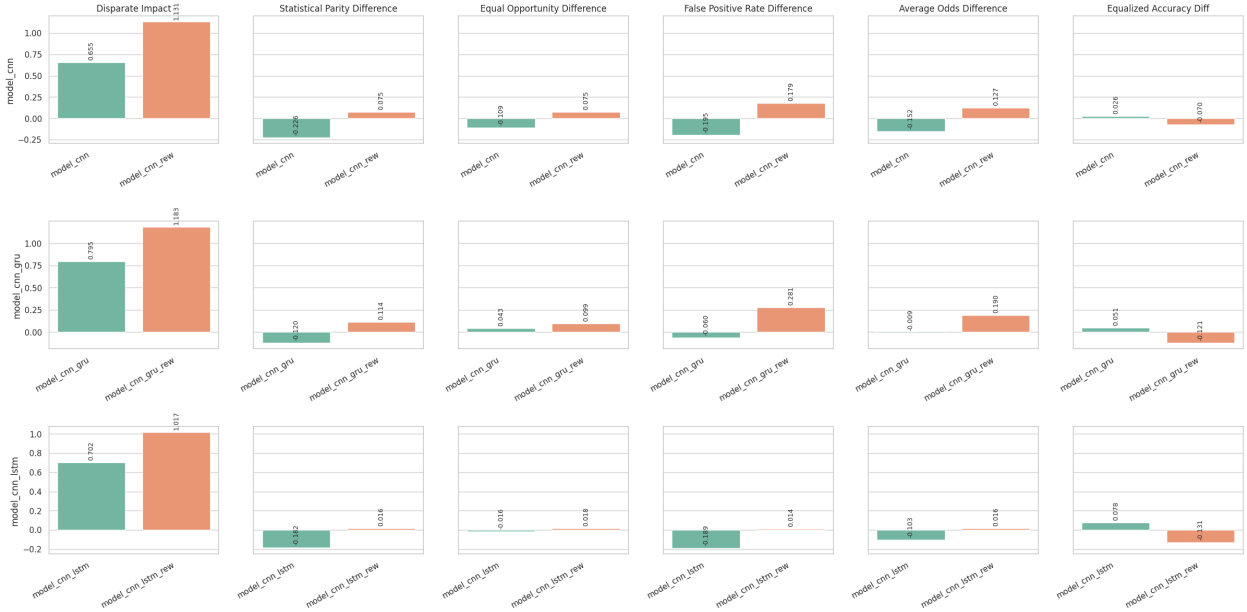


Figure 7: Massaging Result

Figure 8: Regularization Result

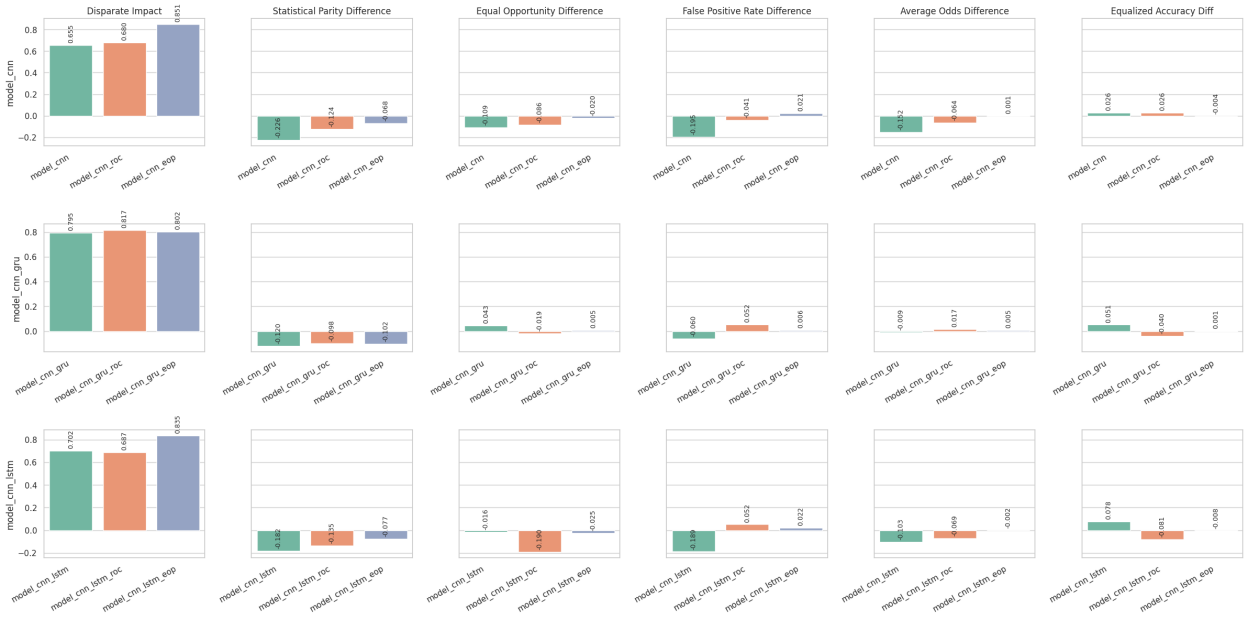


Figure 9: Post-Processing Bias Mitigation Result

Table 3: Summary of Model Performance and Fairness Metrics

Model	Acc.	Prec.	Rec.	F1.	DI	SPD	EOD	FPRD	AOD	EAD
model_cnn	0.791	0.781	0.746	0.763	0.655	-0.226	-0.109	-0.195	-0.152	0.026
model_cnn_gru	0.916	0.898	0.918	0.908	0.795	-0.120	0.043	-0.060	-0.009	0.051
model_cnn_lstm	0.811	0.787	0.799	0.793	0.702	-0.182	-0.016	-0.189	-0.103	0.078
model_cnn_roc	0.740	0.765	0.730	0.730	0.680	-0.124	-0.086	-0.041	-0.064	0.026
model_cnn_gru_roc	**0.920**	**0.920**	**0.925**	**0.925**	**0.817**	**-0.098**	**-0.019**	**0.052**	**0.017**	**-0.040**
model_cnn_lstm_roc	0.770	0.790	0.760	0.760	0.687	-0.135	-0.190	0.052	-0.069	-0.081
model_cnn_eop	**0.760**	**0.760**	**0.750**	**0.760**	**0.851**	**-0.068**	**-0.020**	**0.021**	**0.001**	**-0.004**
model_cnn_gru_eop	0.900	0.895	0.895	0.890	0.802	-0.102	0.005	0.006	0.005	0.001
model_cnn_lstm_eop	0.780	0.785	0.775	0.780	0.835	-0.077	-0.025	0.022	-0.002	-0.008
model_cnn_aug	0.818	0.817	0.769	0.792	0.803	-0.123	0.013	-0.094	-0.040	0.039
model_cnn_gru_aug	0.906	0.884	0.910	0.897	0.758	-0.143	0.062	-0.134	-0.036	0.096
model_cnn_lstm_aug	0.791	0.773	0.761	0.767	0.942	-0.033	0.245	-0.168	0.039	0.196
model_cnn_mas	0.626	0.658	0.358	0.464	1.226	0.157	0.171	0.206	0.189	-0.077
model_cnn_gru_mas	0.778	0.833	0.634	0.720	1.141	0.088	0.099	0.232	0.165	-0.095
model_cnn_lstm_mas	0.636	0.630	0.470	0.538	1.244	0.148	0.085	0.291	0.188	-0.136
model_cnn_rew	**0.684**	**0.669**	**0.590**	**0.627**	**1.131**	**0.075**	**0.075**	**0.179**	**0.127**	**-0.070**
model_cnn_gru_rew	0.768	0.828	0.612	0.704	1.183	0.114	0.099	0.281	0.190	-0.121
model_cnn_lstm_rew	0.545	0.333	0.007	0.015	1.017	0.016	0.018	0.014	0.016	-0.131
model_cnn_reg	0.822	0.783	0.836	0.809	0.642	-0.215	-0.166	-0.091	-0.128	-0.038
model_cnn_gru_reg	**0.949**	**0.928**	**0.963**	**0.945**	**0.834**	**-0.095**	**0.091**	**-0.038**	**0.026**	**0.068**
model_cnn_lstm_reg	0.764	0.776	0.672	0.720	0.776	-0.149	0.013	-0.181	-0.084	0.069

For the CNN model, the fairest results were achieved using the Equalized Odds Postprocessing (EOP) and Reweighing strategies, with EOP providing the best overall outcome. These strategies maintained good accuracy, and all fairness metrics remained within acceptable ranges—defined as 0.80 to 1.20 for Disparate Impact (DI), and –0.10 to 0.10 for the other fairness metrics. For the CNN-GRU model, the best fairness performance was obtained through Regularization, Reject Option Classification (ROC), and Reweighing,

with Regularization and ROC showing the most promising results. These strategies preserved high accuracy (around 90%) while keeping the fairness metrics within the acceptable bounds. In contrast, for the CNN-LSTM model, none of the applied bias mitigation strategies were effective. The resulting fairness metrics fell outside the acceptable range, indicating that these methods did not perform well with this particular model architecture.

6 Conclusion

In this study, we evaluated multiple bias mitigation strategies to improve fairness in EEG-based mental health disorder classification across different deep learning architectures. Some methods, like data augmentation and reweighting, improved fairness but slightly reduced accuracy. In contrast, Regularization and Reject Option Classification (ROC) notably enhanced both accuracy and fairness in the CNN-GRU model. The CNN model achieved its best fairness with Equalized Odds Postprocessing (EOP) and Reweighing. However, none of the strategies effectively improved fairness in the CNN-LSTM model, indicating architecture-specific limitations.

Future work should focus on expanding fairness evaluations beyond the gender dimension. Specifically, fairness concerning age groups present in the dataset warrants further investigation, as age-related bias can be equally critical in clinical settings. Additionally, exploring more advanced in-processing bias mitigation techniques, such as Adversarial Debiasing, may offer better trade-offs between fairness and performance, especially for models like CNN-LSTM where current strategies are less effective.

References

[1] U. Rajendra Acharya et al. "Automated EEG-based screening of depression using deep convolutional neural network". In: *Computer Methods and Programs in Biomedicine* 161 (2018), pp. 103–113. DOI: 10.1016/j.cmpb.2018.04.012. URL: https://doi.org/10.1016/j.cmpb.2018.04.012.

[2] H. Cai et al. "MODMA dataset: A Multi-modal Open Dataset for Mental-disorder Analysis". In: *arXiv preprint* (2020). eprint: arXiv:2002.09283.

[3] H. Dibeklioglu, Z. Hammal, and J.F. Chon. "Dynamic multimodal measurement of depression severity using deep autoencoding". In: *IEEE Journal of Biomedical and Health Informatics* 22 (2017), pp. 525–536.

[4] K. Elnaggar, M.M. El-Gayar, and M. Elmogy. "Depression Detection and Diagnosis Based on Electroencephalogram (EEG) Analysis: A Comprehensive Review". In: *Diagnostics* 15.2 (2025), p. 210. DOI: 10.3390/diagnostics15020210.

[5] Moritz Hardt, Eric Price, and Nati Srebro. "Equality of opportunity in supervised learning". In: *Advances in neural information processing systems*. 2016, pp. 3315–3323.

[6] M. He, E. M. Bakker, and M. S. Lew. "DPD (DePression Detection) Net: a deep neural network for multimodal depression detection". In: *Health Information Science and Systems* 12.1 (Nov. 2024), p. 53. DOI: 10.1007/s13755-024-00311-9.

[7] Faisal Kamiran, Asim Karim, and Xiangliang Zhang. "Decision theory for discrimination-aware classification". In: *2012 IEEE 12th International Conference on Data Mining*. IEEE. 2012, pp. 924–929.

[8] S. Khan et al. "A Machine Learning Based Depression Screening Framework Using Temporal Domain Features of the Electroencephalography Signals". In: *PLoS ONE* 19.3 (Mar. 2024), e0299127. DOI: 10.1371/journal.pone.0299127.

[9] Angus Man Ho Kwok et al. *Machine Learning Fairness for Depression Detection using EEG Data*. 2025. arXiv: 2501.18192 [cs.CV]. URL: https://arxiv.org/abs/2501.18192.

[10] C. D. Mathers and D. Loncar. "Projections of global mortality and burden of disease from 2002 to 2030". In: *PLOS Medicine* 3 (Nov. 2006), pp. 1–20.

[11] P. Olejniczak. "Neurophysiologic basis of EEG". In: *Journal of Clinical Neurophysiology* 23 (2006), pp. 186–189.

[12] Luca Oneto and Silvia Chiappa. "Fairness in Machine Learning". In: *CoRR* abs/2012.15816 (2020). arXiv: 2012.15816. URL: https://arxiv.org/abs/2012.15816.

[13] F. Ringeval et al. "AVEC 2019 Workshop and Challenge: State-of-Mind, Detecting Depression with AI, and Cross-Cultural Affect Recognition". In: *Proceedings of the 9th International on Audio/Visual Emotion Challenge and Workshop*. New York, NY, USA: Association for Computing Machinery, 2019, pp. 3–12.

[14] D. Santomauro et al. "Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic". In: *The Lancet* 398 (Oct. 2021).

[15] K. Smith, P. Renshaw, and J. Bilello. "The diagnosis of depression: Current and emerging methods". In: *Comprehensive Psychiatry* 54 (Aug. 2012).

[16] Zhuozheng Wang et al. "A Depression Diagnosis Method Based on the Hybrid Neural Network and Attention Mechanism". In: *Brain Sciences* 12.7 (2022), p. 834. DOI: 10.3390/brainsci12070834. URL: https://doi.org/10.3390/brainsci12070834.

[17] Qingchen Zhang, Laurence T. Yang, and Zhikui Chen. "Deep Computation Model for Unsupervised Feature Learning on Big Data". In: *IEEE Transactions on Services Computing* 9.1 (2016), pp. 161–171. DOI: 10.1109/TSC.2015.2497705.