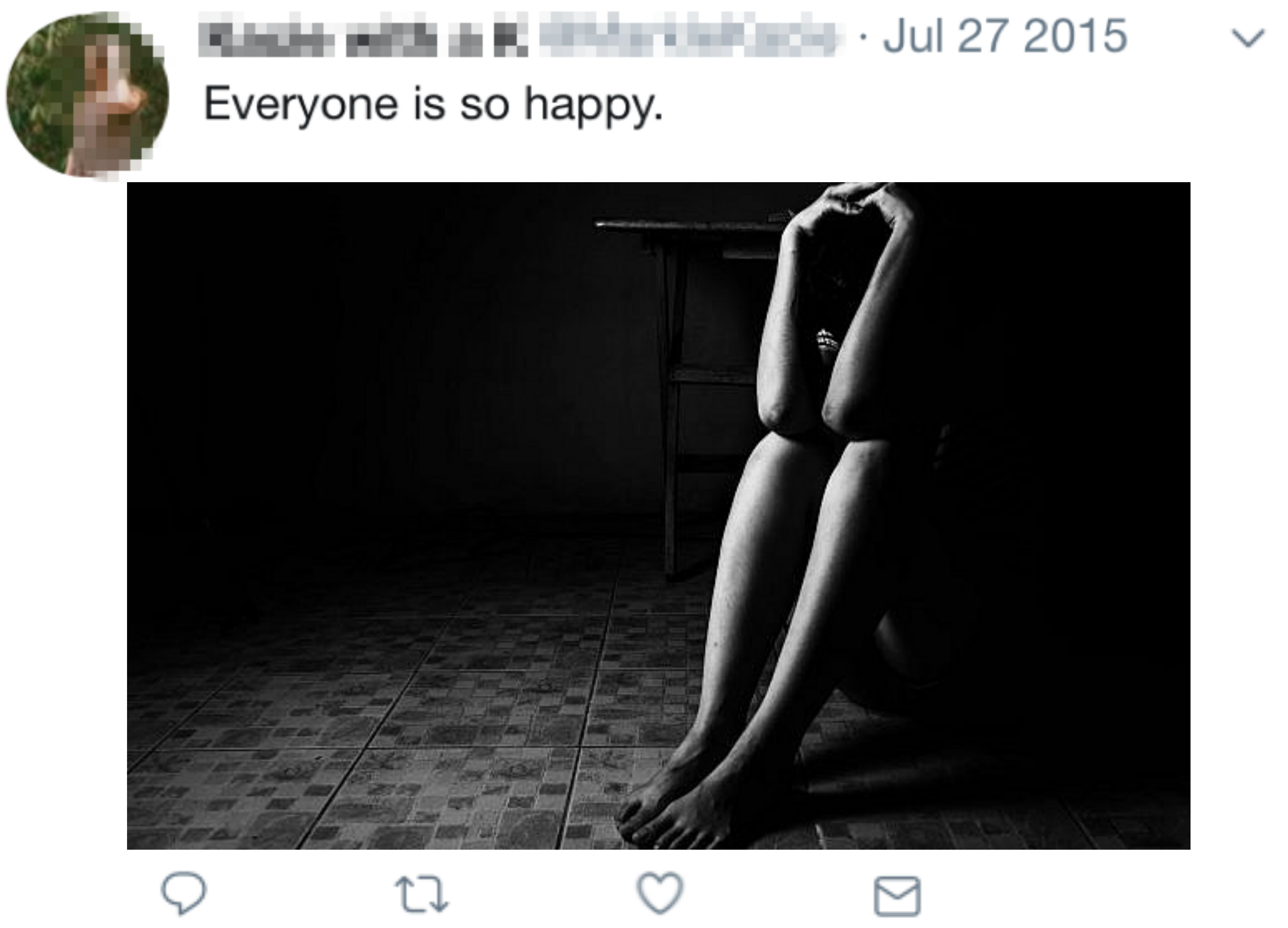
Cooperative Multimodal Approach to Depression Detection in Twitter

Tao Gui*[1], Liang Zhu*[1], Qi Zhang[1], Minlong Peng[1], Xu Zhou[1], Keyu Ding[2], Zhigang Chen[2]

[1]*Shanghai Key Laboratory of Intelligent Information Processing, Fudan University*
[1]*Shanghai Insitute of Intelligent Electroics Systems*
[2]*iFlytek Co., Ltd.*
*Equal contribution.

Introduction

The advent of social media has presented a promising new opportunity for the early detection of depression. To do so effectively, there are two challenges to overcome. The first is that textual and visual information must be jointly considered to make accurate inferences about depression. The second challenge is that due to the variety of content types posted by users, it is difficult to extract many of the relevant indicator texts and images. In this work, we propose the use of a novel cooperative multi-agent model to address these challenges. From the historical posts of users, the proposed method can automatically select related indicator texts and images.


[redacted] · Jul 27 2015

Everyone is so happy.

Figure 1: An example of a multimodal tweet. If we consider only the textual content "*Everyone is so happy,*" we cannot easily determine the actual feelings of the author. Images posted by users can provide a wealth of information for detecting depression.

Datasets

Dataset		# Users	# T	# T + I
D₁	Depressed	1,402	292,564	-
	Non-Depressed	5,160	3,953,183	-
D₂	Depressed	1,402	251,834	40,730
	Non-Depressed	5, 160	3,302,366	650,817

Table 1: Statistical details of the datasets used in our experiments, where **# T** and **# T + I** represent the number of tweets that contain only texts and that contain both text + image pairs, respectively.

Results

Methods	Training Data	Accuracy	Precision	Recall	F1
NB (Pedregosa et al. 2011)	Various Features	0.724	0.727	0.728	0.728
MSNL (Song et al. 2015)		0.818	0.818	0.818	0.818
WDL (Rolet, Cuturi, and Peyré 2016)		0.768	0.769	0.768	0.768
MDL (Shen et al. 2017)		0.848	0.848	0.850	0.849
GRU (Chung et al. 2014)	Text	0.824	0.825	0.823	0.824
GRU + Random sampling		0.760	0.760	0.757	0.756
VGG-Net (Simonyan and Zisserman 2014)	Image	0.702	0.703	0.702	0.702
VGG-Net + Random sampling		0.642	0.643	0.642	0.643
GRU + VGG-Net	Text+Image	0.845	0.843	0.847	0.845
GRU + VGG-Net + Random sampling		0.811	0.811	0.810	0.810
Co-Attention (Lu et al. 2016)		0.866	0.871	0.863	0.865
Dual-Attention (Nam, Ha, and Kim 2017)		0.848	0.848	0.848	0.848
Modality Attention (Moon, Neves, and Carvalho 2018)		0.866	0.868	0.862	0.864
GRU + VGG-Net + Unified advantages (Egorov 2016)		0.866	0.866	0.865	0.865
GRU + VGG-Net + COMMA (text + image)		**0.900**	**0.900**	**0.901**	**0.900**

Table 2: Comparison of performances in terms of four selected measures.

Method: COMMA

We propose COMMA policy gradients, which adopt a centralized training framework with decentralized execution by applying a centralized critic and differentiated advantages, as shown in Figure 2. Both the text and image selectors are policy gradient agents, which take the text and image features as inputs and determine whether to select the features. The selectors are trained by following the different gradients estimated by the critic. The differentiated advantages are shaped rewards that compare the current global reward to those received when each agent's action is replaced with an opposite action (misoperation). The text and images features are extracted by GRU and pretrained VGG-Net, respectively, and then the classifier uses the features selected by agents to detect depression.

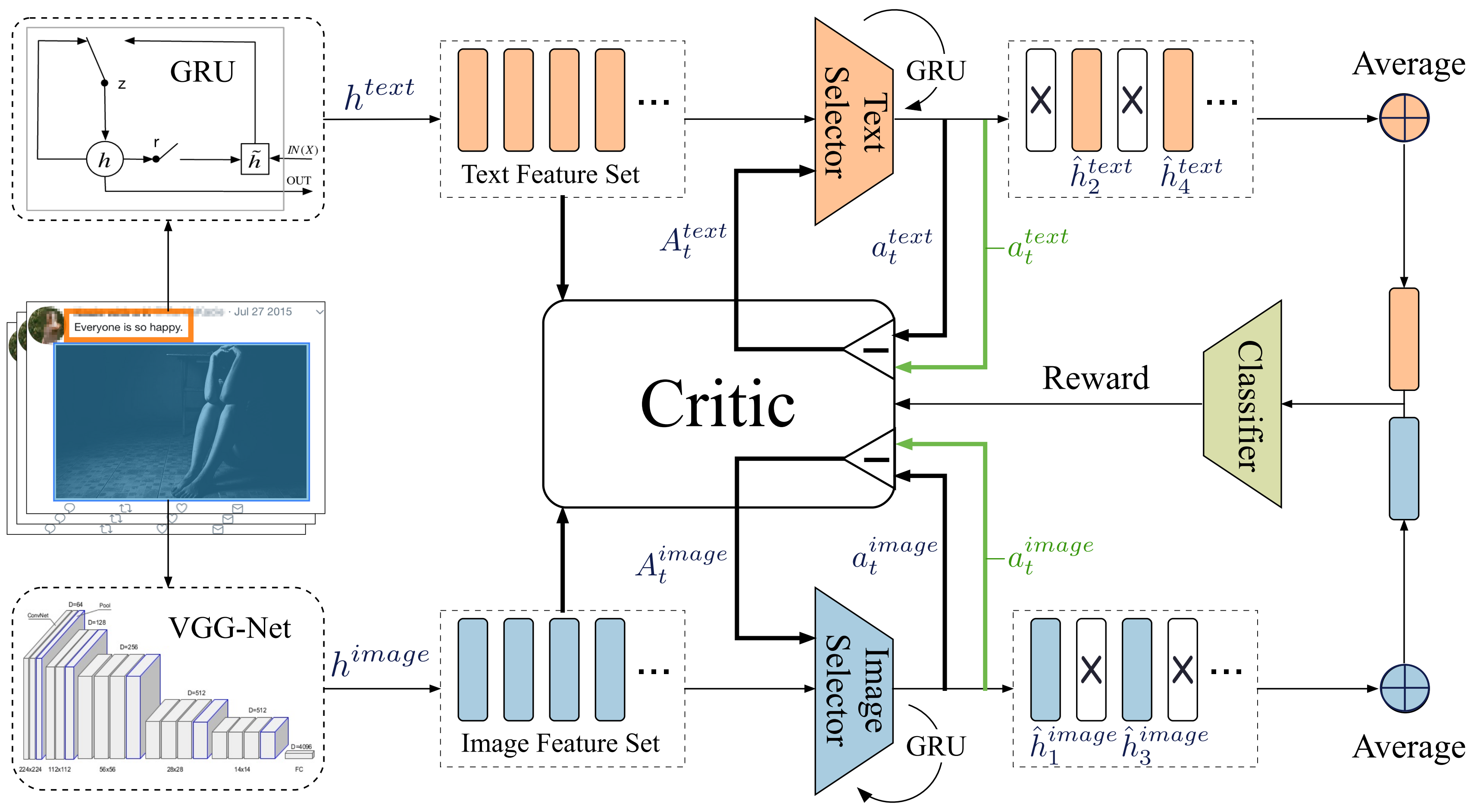


Figure 2: Architecture of the proposed model. At each time step t, the advantage A_t^e of selector e is given by comparing the current global reward to the reward received when that agent's action is replaced with an opposite action $-a_t^e$.

Analysis

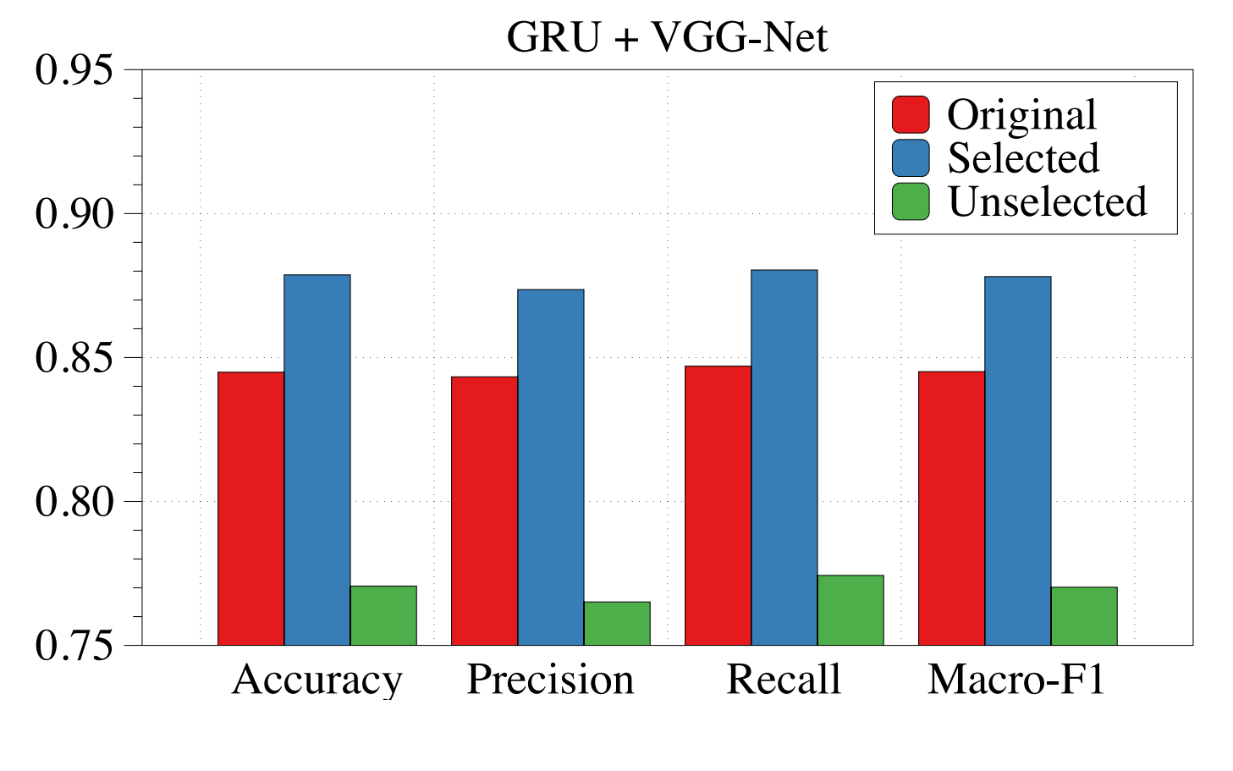


Figure 3: Comparison of models trained on original posts, selected posts, and unselected posts.

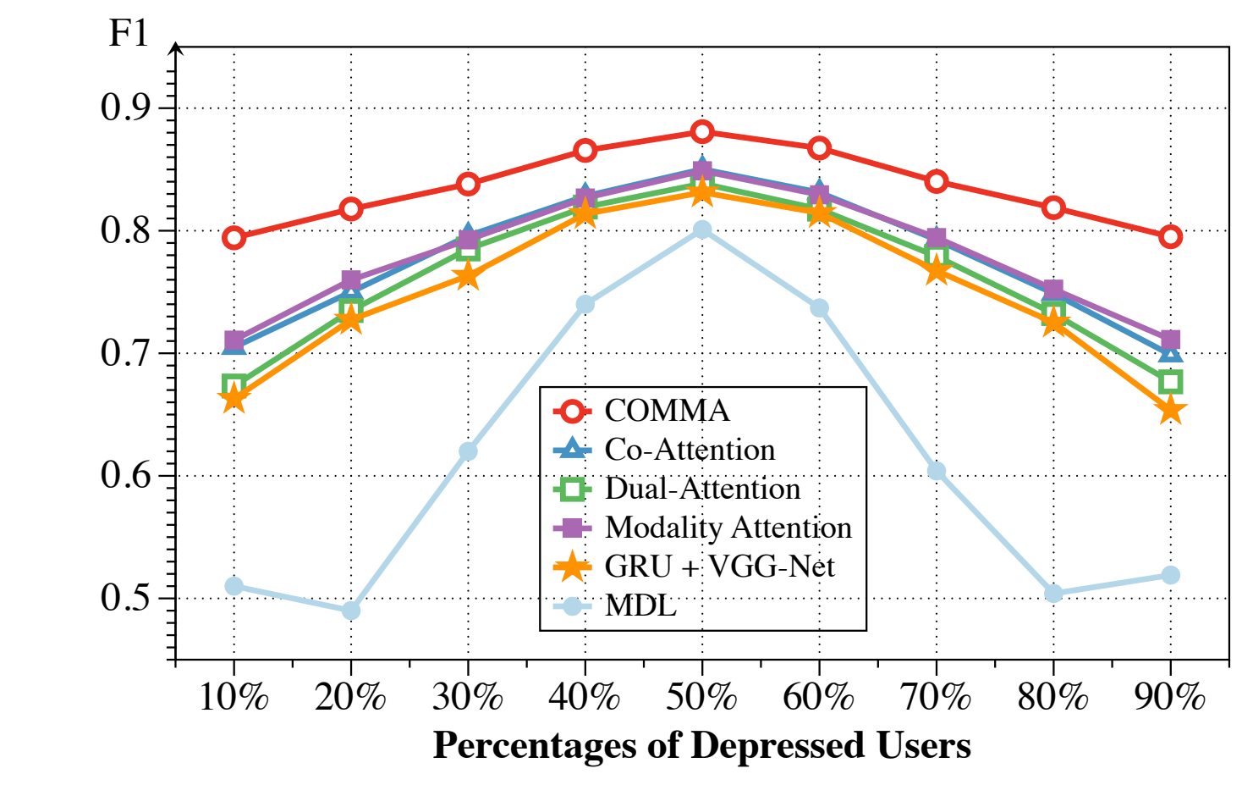


Figure 4: Comparison of the models trained on the datasets with different percentages of depressed users. The total number of users is 1,500.

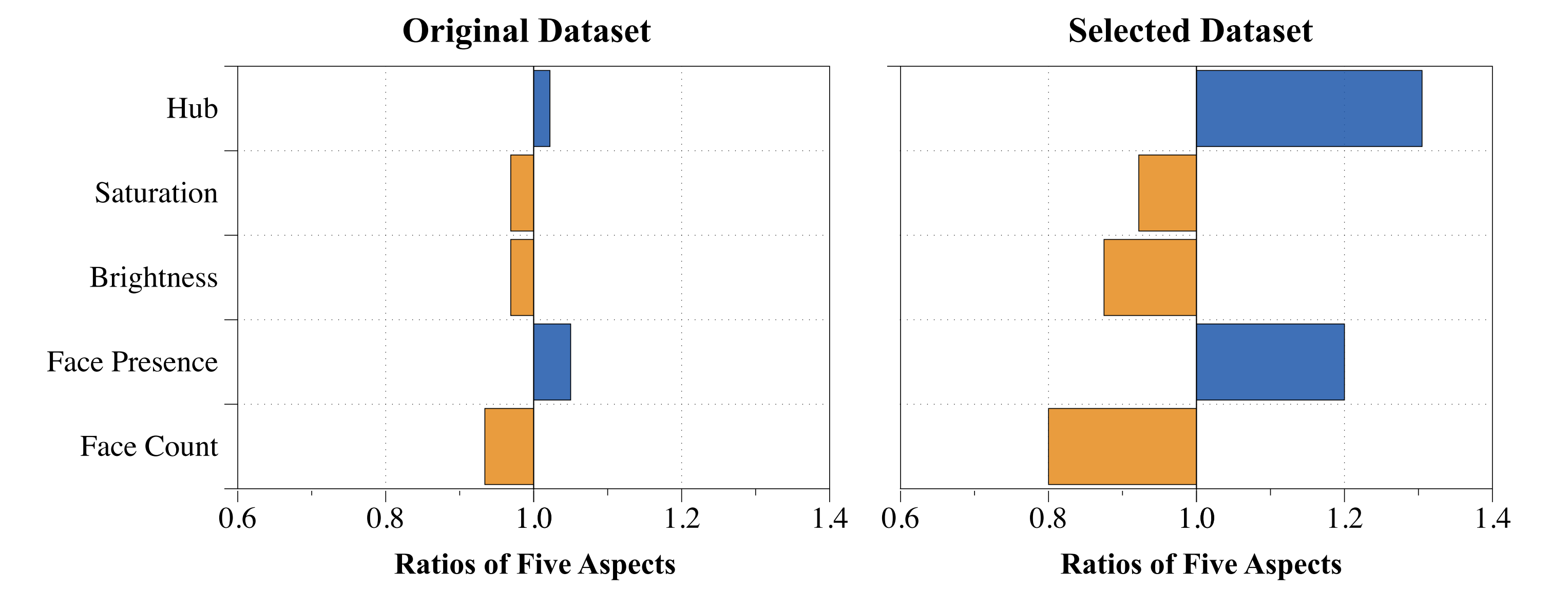


Figure 5: Comparison of original and selected posts. The y-axis values show the five aspects of each image, and the x-axis values are the ratios of these five aspect values of depressed users to those of non-depressed users.