# Multimodal Classification of Sexist Advertisements

Francesca Gasparini, Ilaria Erba, Elisabetta Fersini and Silvia Corchs
*Department of Informatics, Systems and Communication, University of Milano-Bicocca, Milan, Italy*

Keywords: Image Classification, Multimodal Classification, Sexist Advertising, Text Analysis.

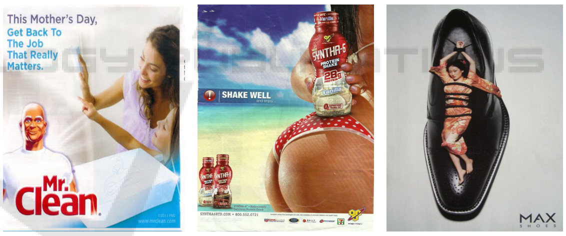
Abstract: Advertisements, especially in online social media, are often based on visual and/or textual persuasive messages, frequently showing women as subjects. Some of these advertisements create a biased portrays of women, finally resulting as sexist and in some cases misogynist. In this paper we give a first insight in the field of automatic detection of sexist multimedia contents, by proposing both a unimodal and a multimodal approach. In the unimodal approach we propose binary classifiers based on different visual features to automatically detect sexist visual content. In the multimodal approach both visual and textual features are considered. We created a manually labeled database of sexist and non sexist advertisements, composed of two main datasets: a first one containing 423 advertisements with images that have been considered sexist (or non sexist) with respect to their visual content, and a second dataset comprising 192 advertisements labeled as sexist and non sexist according to visual and/or textual cues. We adopted the first dataset to train a visual classifier. Finally we proved that a multimodal approach that considers the trained visual classifier and a textual one permits good classification performance on the second dataset, reaching 87% of recall and 75% of accuracy, which are significantly higher than the performance obtained by each of the corresponding unimodal approaches.

## 1 INTRODUCTION

Advertisements nowadays are more and more heavily based on surrounding messages of persuasion, especially making use of women as subjects. It is highly probable, especially in online social media, to look at an advertising with a female subject involved where the woman is portrayed in a highly sexualized manner with little connection to the brand being advertised (Zimmerman and Dahlberg, 2008). This type of communication, together with even more aggressive contents, can ultimately end up in misogynistic advertising.

Among the communication methods related to the advertising domain, both the image and the accompanying text, can encode several forms of sexism. Among them, we can highlight the most prevalent ones (Plakoyiannaki et al., 2008; Poland, 2016):

- Stereotype: women are typically portrayed as a good wives mainly concerned with tasks of housekeeping (see Figure 1(a));
- Objectification: women are presented as sex objects, even if sex is unrelated to the promoted product (see Figure 1(b));
- Dominance: the advertisement depicts women as physically or mentally dominated by men (see Figure 1(c)).



(a) Stereotype (b) Objectification (c) Dominance

Figure 1: Graphical examples Stereotype, Objectification and Dominance in advertisements.

These communication tools are detrimental to society because of the creation of biased portrays of women, resulting in unhealthy social and physical habits. Detecting, and taking actions against these forms of sexism, would demonstrate to women that they are valued in society. However, to the best of our knowledge, no research has been conducted in the literature to automatically detect sexist advertisements, neither analyzing the visual components nor the textual cues.

The problem of identifying misogynist language in online social media has recently attracted signifi-

Gasparini, F., Erba, I., Fersini, E. and Corchs, S.
Multimodal Classification of Sexist Advertisements.
DOI: 10.5220/0006859403990406
In *Proceedings of the 15th International Joint Conference on e-Business and Telecommunications (ICETE 2018) - Volume 1: DCNET, ICE-B, OPTICS, SIGMAP and WINSYS*, pages 399-406
ISBN: 978-989-758-319-3

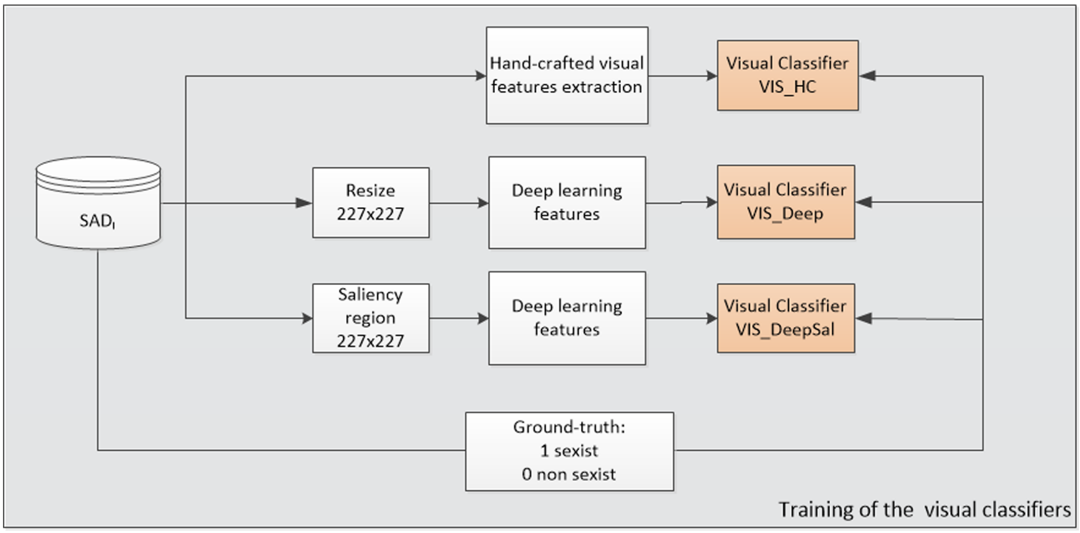

Hand-crafted visual features extraction
Visual Classifier VIS_HC
SAD$_I$
Resize 227x227
Deep learning features
Visual Classifier VIS_Deep
Saliency region 227x227
Deep learning features
Visual Classifier VIS_DeepSal
Ground-truth: 1 sexist 0 non sexist
Training of the visual classifiers

Figure 2: The three unimodal (visual) classifiers trained on the $SAD_I$ database.

cant attention. Social networks need to update their policy to address this issue and due to the high volume of texts shared daily, the automatic detection of misogynist and sexist text content is required. However, the problem of automatic misogyny identification from a linguistic point of view is still in its early stage. In particular, trivial statistics about the usage of misogynystic language in Twitter have been provided in (Hewitt et al., 2016), while in (Anzovino et al., 2018) a first tentative of defining linguistic features and machine learning models for automatically recognizing this phenomenon has been presented.

From the visual point of view, great efforts in the literature were devoted in defining appropriate filters to automatically detect digital pornographic images from non-pornographic ones. These filters are mainly based on color features and geometric relations of different body parts in naked skin regions (Ries and Lienhart, 2014; Zaidan et al., 2013). Automatic recognition of pornographic Web pages has also been considered. In particular classification strategies that consider text and image features have been proposed (Hu et al., 2007), showing that a fusion of text and image classification results outperforms unimodal classifiers. The improvement in classification performance adopting multimodal features has also been demonstrated recently by Corchs et al.(Corchs et al., 2017). In their work, integrating visual and textual information permits to increase the performance of emotion classification of social images.

In order to give a first insight in the field of automatic detection of sexist multimedia content, and taking into account the promising classification results of multimodal approaches, several contributions are proposed in this work:

- A unimodal classifier based only on visual features to automatically detect sexist visual content;
- A multimodal approach where both visual and texture features are analyzed to classify sexist advertisements;
- $SAD_I$: An image dataset of 423 advertisements[1], which have been considered sexist (or non sexist) with respect to the visual content. Up to our knowledge, this dataset is the first resource available in the literature concerned with sexist advertisements;
- $SAD_{IT}$: A dataset of 192 advertisements1] with sexist (and non sexist) contents labeled considering the visual and/or textual aspects. This dataset includes both images and related texts.

The paper is organized as follows. In section 2, the proposed unimodal and multimodal approaches for recognizing sexist advertisements are introduced. In section 3, the feature space derived for characterizing both visual and linguistic features associated to advertisements is described. In section 4, the proposed database of sexist advertisements is presented. In section 5 experimental results are reported. Finally, in section 6 conclusions and future work are discussed.

## 2 CLASSIFICATION APPROACH

Staring from the $SAD_I$ dataset we first tackle the task of classifying sexist contents from the visual point of view. The visual features considered for this issue

[1] The proposed dataset will be made available at publication time.

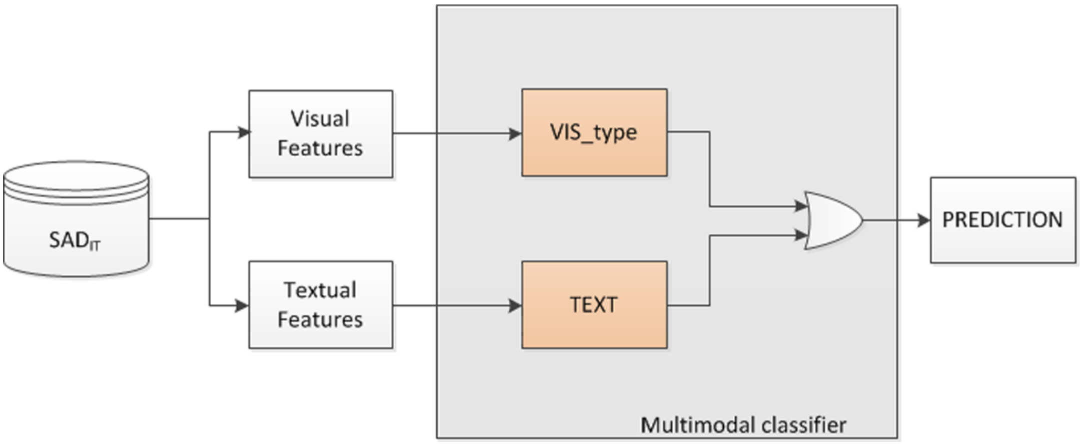


Figure 3: Multimodal classifier defined for the $SAD_{IT}$ database.

are both hand-crafted and deep ones. We extract the deep features with two different strategies: from a resized version of the input images and from the most salient region of them. In this way we train three different visual classifiers, labeled respectively $VIS\_HD$, $VIS\_Deep$, and $VIS\_DeepSal$, as reported in Figure 2. As unimodal classification models, we here consider and compare three state of the art models of classifier: Nearest Neighbor (NN), Decision Tree (DT), and Support Vector Machine (SVM).

Concerning the advertisements of the $SAD_{IT}$ dataset, they are classified from both the visual and textual point of view combining a visual classifier and a textual one, following a late fusion approach. In this case, as visual classifier, we consider and compare each of the three ones obtained in the training phase, as depicted in Figure 2, (VIS_type, where type={HC, Deep, DeepSal}). For what concerns the textual contents, we adopt a classifier trained not on advertisements, but using social network misogynist contents (Anzovino et al., 2018). In particular, we exploited a model trained on Twitter messages collected according to three main strategies: (1) using a set of representative keywords that are likely associated to sexist contents, (2) monitoring potential victims and (3) downloading the timeline tweets given by selected misogynist users. The model was trained by using both misogynistic and non-misogynistic texts (manually labeled), and exploiting the feature set presented in section 3.2. For the $SAD_{IT}$ dataset, we employed a late fusion approach, depicted in Figure 3, where the final classification $C_j$ of the $j^{th}$ advertisement is obtained as follows:

$$C_j = VIS\_type_j \vee TEXT_j \quad (1)$$

where $VIS\_type_j$ is the class assigned by the visual classifier considered, and $TEXT_j$ is the class assigned by the textual classifier.

# 3 FEATURE SPACE

## 3.1 Visual Features

To define our visual classifier we adopt both hand-crafted visual features, as well as a deep learning representation.

### 3.1.1 Hand-crafted Visual Features

We consider several hand-crafted visual features trying to take into account from low level to high level properties of visual content.

- Low level, grayscale features: Coarseness, Contrast, Directionality, Linelikeness, Roughness, (Tamura et al., 1978), Edge density (Mack and Oliva, 2004), Entropy (Schettini et al., 2010), Measure of Enhancement (Schettini et al., 2010), Local Binary Pattern (LBP) (Ojala et al., 1996), and Histogram of Oriented Gradients (HoG), developed by Ludwig et al. (Junior et al., 2009). All these features are 1-Dimensional (1-D), except LBP (2891-D) and HoG (1296-D).
- Low level colored Features: Chroma Variance (1-D) (Ciocca et al., 2016), Number of Regions (1-D) (Comaniciu and Meer, 2002), Colorfullness (1-D) (Hasler and Suesstrunk, 2003), Color Histogram in the HSV color space (32-D), color simple statistics (mean and standard deviation) in the RGB color space, (6-D), Auto-correlogram obtained quantizing the RGB color space in 64 colors, (64-D).
- Photographic, and aestetic features, and features related to visual perception: Feature Congestion and Subband Entropy (Rosenholtz et al., 2007), image complexity (Corchs et al., 2016), a measure

of the degree of focus (Minhas et al., 2009), all of them 1-D and a 113-D aesthetic feature vector (Bhattacharya et al., 2013).

- Features related to semantic concepts: a percentage of skin measure (Gasparini et al., 2008) and the number of faces (Viola and Jones, 2001).

The size of the hand-crafted feature vector is 4418.

#### 3.1.2 Deep Visual Features

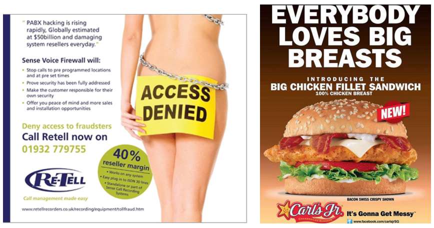
We extract a 4096-D feature vector from the last fully-connected layer, *L*7 of the pre-trained CNN AlexNet (Krizhevsky et al., 2012), without any fine tuning. We used the open-source deep learning library MatConvNet (Vedaldi and Lenc, 2015). This net requires input images of 227x227 pixels. In this work we consider two strategies to get this input size: a direct resize on the input image, and a strategy that resized the most salient region. The salient regions are obtained binarizing the saliency map (Itti and Koch, 2001).

### 3.2 Textual Features

A sexist language can be represented by several features that can be extracted from the text enclosed into the advertisements. In particular, the following subset of features, initially introduced in (Anzovino et al., 2018), have been considered:

- N-grams: we considered both character n-grams and token n-grams, according to their forms as unigrams, bigrams and trigrams.
- Syntactic: we considered Bag Of Part Of Speech, i.e. unigrams, bigrams and trigrams of Part of Speech tags extracted from the text contained in the advertisement;
- Metadata: such as the number of adjectives, personal pronouns, adverbs and possessive pronouns;
- Embedding: we created a high level representation of words contained into the advertisements using a model pre-trained on short text[2]. The purpose of this type of feature is to represent the text of an advertisement as the average representation of all its words derived through the word2vec model (Mikolov et al., 2013).

The size of the textual feature vector is 142371.


" PABX hacking is rising rapidly, Globally estimated at $50billion and damaging system resellers everyday."
Sense Voice Firewall will:
• Stop calls to pre programmed locations and at pre set times
• Prove security has been fully addressed
• Make the customer responsible for their own security
• Offer you peace of mind and more sales and installation opportunities
Deny access to fraudsters
Call Retell now on
01932 779755
RE-TELL
Call management made easy
www.retellrecorders.co.uk/recording/equipment/tollfraud.htm
ACCESS DENIED
40% reseller margin
• Works on any system
• Easy plug in to ISDN 30 lines
• Standalone or part of Sense Call Recording Systems
EVERYBODY LOVES BIG BREASTS
INTRODUCING THE
BIG CHICKEN FILLET SANDWICH
100% CHICKEN BREAST
NEW!
BACON SWISS CRISPY SHOWN
Carl's Jr.
It's Gonna Get Messy
www.facebook.com/carlsjrSG

(a) (b)

Figure 4: Sexist advertisements belonging to the $SAD_{IT}$ database, where the sexist aspect is mainly related to one of the two contents: visual (a), or text (b).

## 4 SAD: THE SEXIST ADVERTISEMENT DATABASE

Our SAD database is composed of two different datasets:

- $SAD_I$: a dataset of 423 images of advertisements, 215 labeled as sexist images and the remaining 208 labeled as non-sexist ones. The sexist images have been downloaded from an Italian Facebook group: *La pubblicitá sessista offende tutti (Sexist advertisement hurts everybody)*[3]. The images labeled as sexist were chosen among those validated by the members of this group. The non-sexist images were downloaded from another Italian Facebook group: *Pubblicitá creative e bizzarre (Creative and bizarre advertisements)*[4].
- $SAD_{IT}$: a dataset of 192 advertisements where both images and related texts have been downloaded. Within this dataset there are 107 sexist advertisements, and 85 non sexist ones. The 107 images are considered sexist either from the visual or textual point of view. Some of them are considered sexist from both these points of view together. The sexist advertisements were downloaded from the web using as keywords: "sexist ads", "woman objectification", "misogynyst ads". The non-sexists ones were also downloaded from the web and were chosen and validated by the authors of the paper.

  Some example of sexist advertisements are reported in Figure 5. It is important to highlight that there are advertisements where the visual content can be considered sexist (see Figure 4(a)), while the related text is neutral, while other show a neutral visual content, but contains a related sexist text (see Figure 4(b)). Examples of non sexist advertisements are depicted in Figure 6.

[2] https://www.fredericgodin.com/software/

[3] https://www.facebook.com/groups/pubblicitasessistaoffende/about/

[4] https://www.facebook.com/pubblicitacreativebizzarre

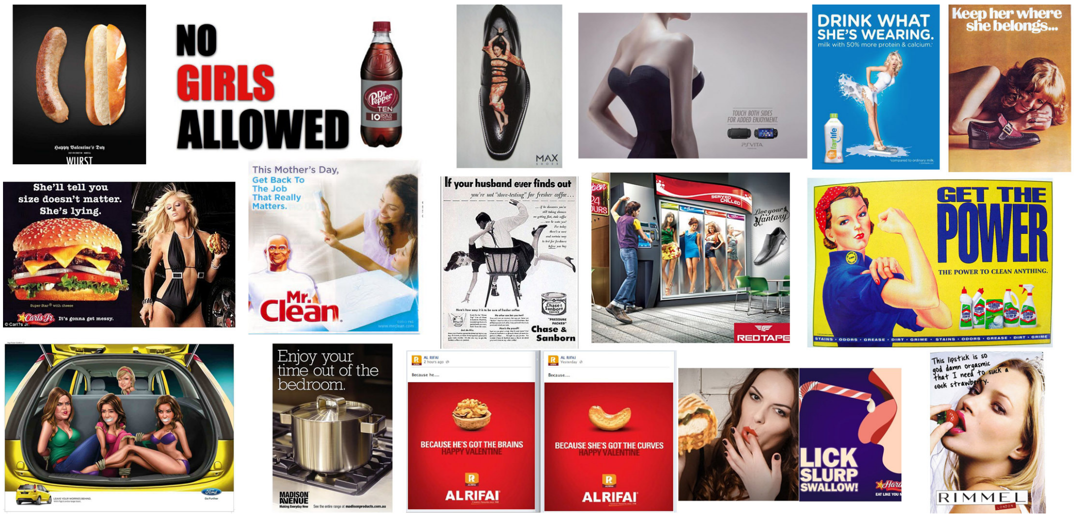

NO GIRLS ALLOWED
DRINK WHAT SHE'S WEARING.
Keep her where she belongs...
She'll tell you size doesn't matter. She's lying.
This Mother's Day, Get Back To The Job That Really Matters.
Mr. Clean
If your husband ever finds out
GET THE POWER
Enjoy your time out of the bedroom.
BECAUSE HE'S GOT THE BRAINS
BECAUSE SHE'S GOT THE CURVES
ALRIFAI
LICK SLURP SWALLOW!
RIMMEL

Figure 5: Examples of sexists advertisement belonging to the $SAD_{IT}$ database. Sexism can be related to the visual and/or textual content.

# 5 EXPERIMENTAL RESULTS

We here consider the results of the two different classification tasks addressed in this work. First we analyze the classification performance of the three visual classifiers here proposed (unimodal classification), to automatically detect sexist visual content in advertisements. Then we evaluate the performance of the multimodal classification of advertisements that are sexist in terms of text and/or visual content.

## 5.1 Unimodal Classification: Visual Content

We trained the three different visual classifiers described in Section 2 and depicted in Figure 2, on the $SAD_I$ database. For each of them, *VIS_HD*, *VIS_Deep*, and *VIS_DeepSal*, we considered three classifier models: NN, DT, and SVM and we evaluated the classification performance on the training dataset applying a five-fold cross validation strategy on these models, varying their parameters (number of neighbors for the k-NN, with the euclidean distance, pruning strategies for DT, and different kernels for SVM). We report the results in terms of accuracy, recall, precision and f-measure (i.e. the harmonic mean of recall and precision), with respect to the sexist class (positive class).

For all the three visual classifiers the best model in terms of f-measure is SVM with a radial basis function kernel. In Table 1 we report the performance of this model for each of the three types of visual features considered.

Table 1: Classification performance in terms of Accuracy (%), Precision (%), Recall (%) and F-measure (%) obtained by the best performing model (SVM), for each of the three types of visual classifiers on the training dataset.

| | **VIS_HC** | **VIS_Deep** | **VIS_DeepSal** |
|---|---|---|---|
| Accuracy | 72.10 | 78.72 | **80.85** |
| Precision | 70.46 | 73.86 | **79.91** |
| Recall | 77.67 | 82.79 | **83.26** |
| F-measure | 73.9 | 78.07 | **81.55** |

VIS_DeepSal is the best visual classifier on the training set, with respect to all the performance measures considered. It is not surprising that deep features extracted from a pre-trained net permit to achieve a high classification performance even in classification task not related to the one of the original net. It has been demonstrated by Razavian et al. (Sharif Razavian et al., 2014) that generic of-the-shelf CNN representation achieves better results in several computer vision classification tasks than other handcrafted visual features. Evaluating the deep features on the most salient region instead of on the whole image, seems to increase the classification performance.

## 5.2 Multimodal Classification: Visual and Textual Content

We here report the results of the classification on the $SAD_{IT}$ dataset, applying our multimodal classification strategy depicted in Figure 3, compared with the performance of the unimodal classifiers. Table 2 summarizes these results in terms of accuracy (%), recall (%), precision (%) and f-measure (%), with respect to the sexist class.

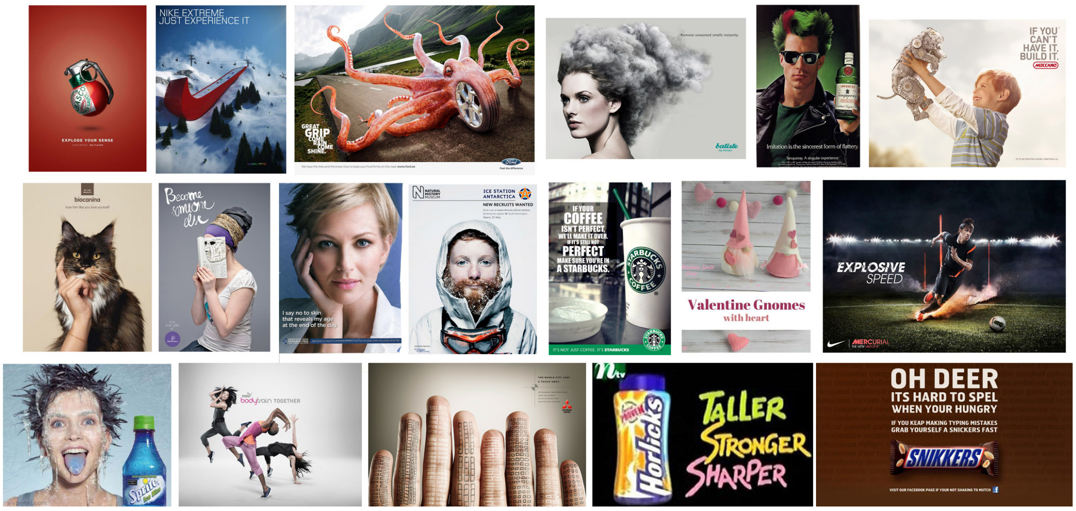

NIKE EXTREME
JUST EXPERIENCE IT
GREAT GRIP
IF YOU CAN'T HAVE IT, BUILD IT.
Become more alive
IF YOUR COFFEE ISN'T PERFECT, WE'LL MAKE IT OVER. IF IT'S STILL NOT PERFECT, MAKE SURE YOU'RE IN A STARBUCKS.
Valentine Gnomes
with heart
EXPLOSIVE SPEED
TALLER STRONGER SHARPER
OH DEER
ITS HARD TO SPEL WHEN YOUR HUNGRY
IF YOU KEAP MAKING TYPING MISTAKES GRAB YOURSELF A SNICKERS FAST
SNIKKERS

Figure 6: Examples of non sexist advertisement belonging to the $SAD_{IT}$ database. In this case neither the visual content nor the textual one are considered sexist.

Table 2: Performance of the different classifiers on the $SAD_{IT}$ dataset, in terms of Accuracy (%), Precision (%), Recall (%) and F-measure (%).

| | TEXT | VIS_HC | VIS_HC ∨ TEXT | VIS_Deep | VIS_Deep ∨ TEXT | VIS_DeepSal | VIS_DeepSal ∨ TEXT |
|---|---|---|---|---|---|---|---|
| Accuracy | 61.86 | 65.63 | 67.19 | **69.79** | 68.23 | 68.23 | 67.19 |
| Precision | **72.00** | 68.47 | 65.94 | 69.92 | 66.43 | 69.83 | 65.94 |
| Recall | 50.47 | 71.03 | 85.05 | 80.37 | **86.92** | 75.70 | 85.05 |
| F-measure | 59.34 | 69.72 | 74.29 | 74.78 | **75.30** | 72.65 | 74.29 |

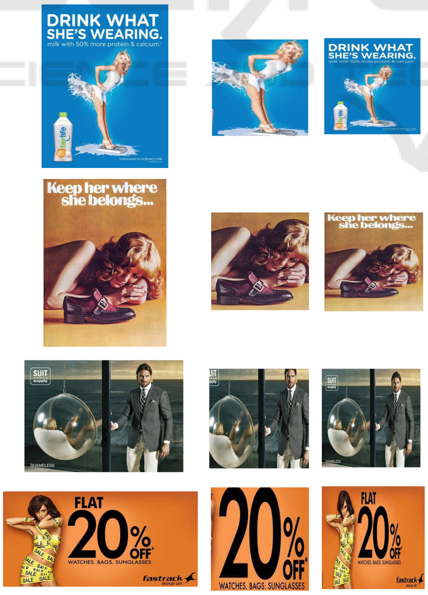

DRINK WHAT SHE'S WEARING.
milk with 50% more protein & calcium.
Keep her where she belongs...
SUIT supply
FLAT 20% OFF
WATCHES, BAGS, SUNGLASSES
fastrack

Figure 7: First column: original images; second column: corresponding most salient regions; third column: resampling of the original images. Observe that in some cases the most salient region corresponds to a significant portion of the original image (first and second rows), in some cases it is more or less equivalent (third row), while in other cases the extracted region does not provide adequate visual information (last row).

In the first column the performance of the unimodal text classifier is reported. We want to remark that this classifier was previously trained on a different classification task, by using both misogynistic and non-misogynistic texts from Twitter messages. However its performance when applied to detect sexist text in advertisements shows the highest precision when compared with all the other classifiers here considered.

To better evaluate the performance of the unimodal approaches on the $SAD_{IT}$ dataset we should recall that the elements of this dataset do not necessarily are contemporaneously sexist for both visual and textual contents, while the label of the classes are assigned in presence of at least one of the two contents. The multimodal classifiers always increase the performance in terms of recall and f-measure, when compared with the corresponding unimodal ones, while keeping high value of accuracy. Moreover, the contribution of the text classifier reduces the differences in performance adopting different visual features, in particular comparing hand-crafted versus deep ones.

The best classifier in terms of recall (86.92%) and f-measure (75.30%) is obtained combining VIS_Deep with TEXT. The deep features extracted on the most salient region do not permit on this dataset an improvement in classification performance. An analysis of the salient regions extracted on images belonging to the $SAD_{IT}$ dataset is reported in Figure 7. In this Figure, original images (left column) are compared with the corresponding salient regions (central column) and the resided version of the original images (right column). The first two rows represent images where the salient regions extracted significantly correspond to the salient visual content. Instead for images with a high dominance of text (fourth row) the salient region could not be related to the visual content, extracting prevalently the text. Finally for several images, the salient region is more or less equivalent to the original image itself (third row). This analysis can justify the small differences, in terms of performance, obtained on this dataset by applying the two different deep strategies.

## 6 CONCLUSIONS

Considering the task of classifying the sexist content of advertisements, we have proved that a multimodal approach that considers both visual and textual features permits good classification performance especially when compared with unimodal approaches. To best of our knowledge this is the first work that deals with this task. Within this context we provide a dataset of advertisements with images and text that will be available to the research community. Starting from the promising results here obtained, significant improvements can be reached with further analysis. In particularly, we plan to increase the dataset of both images and texts, and we plan to train a textual classifier on a more specific task. To automatically extract the texts, we plan to investigate different ocr, in fact within this preliminary analysis, texts have been manually extracted. The extraction of more significant salient regions in presence of text should also be investigated. Finally, other strategies to obtain a multimodal classification will be investigated, considering both early fusion and late fusion approaches.

## ACKNOWLEDGEMENTS

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Tesla K40 GPU used for this research.

## REFERENCES

Anzovino, M., Fersini, E., and Rosso, P. (2018). Automatic Identification and Classification of Misogynistic Language on Twitter. In *23rd International Conference on Natural Language & Information Systems*.

Bhattacharya, S., Nojavanasghari, B., Chen, T., Liu, D., Chang, S.-F., and Shah, M. (2013). Towards a comprehensive computational model foraesthetic assessment of videos. In *Proceedings of the 21st ACM international conference on Multimedia*, pages 361–364. ACM.

Ciocca, G., Corchs, S., and Gasparini, F. (2016). Genetic programming approach to evaluate complexity of texture images. *Journal of Electronic Imaging*, 25(6):061408–061408.

Comaniciu, D. and Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(5):603–619.

Corchs, S., Fersini, E., and Gasparini, F. (2017). Ensemble learning on visual and textual data for social image emotion classification. *International Journal of Machine Learning and Cybernetics*, pages 1–14.

Corchs, S. E., Ciocca, G., Bricolo, E., and Gasparini, F. (2016). Predicting complexity perception of real world images. *PloS one*, 11(6):e0157986.

Gasparini, F., Corchs, S., and Schettini, R. (2008). Recall or precision-oriented strategies for binary classification of skin pixels. *Journal of electronic imaging*, 17(2):023017–023017.

Hasler, D. and Suesstrunk, S. E. (2003). Measuring colorfulness in natural images. In *Electronic Imaging 2003*, pages 87–95. International Society for Optics and Photonics.

Hewitt, S., Tiropanis, T., and Bokhove, C. (2016). The problem of identifying misogynist language on twitter (and other online social spaces). In *Proceedings of the 8th ACM Conference on Web Science*, pages 333–335. ACM.

Hu, W., Wu, O., Chen, Z., Fu, Z., and Maybank, S. (2007). Recognition of pornographic web pages by classifying texts and images. *IEEE transactions on pattern analysis and machine intelligence*, 29(6):1019–1034.

Itti, L. and Koch, C. (2001). Computational modelling of visual attention. *Nature reviews neuroscience*, 2(3):194.

Junior, O. L., Delgado, D., Gonçalves, V., and Nunes, U. (2009). Trainable classifier-fusion schemes: An application to pedestrian detection. In *Intelligent Transportation Systems, 2009. ITSC'09. 12th International IEEE Conference on*, pages 1–6. IEEE.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105.

Mack, M. and Oliva, A. (2004). Computational estimation of visual complexity. In *the 12th Annual Object, Perception, Attention, and Memory Conference*.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119.

Minhas, R., Mohammed, A. A., Wu, Q. J., and Sid-Ahmed, M. A. (2009). 3d shape from focus and depth map computation using steerable filters. In *International Conference Image Analysis and Recognition*, pages 573–583. Springer.

Ojala, T., Pietikäinen, M., and Harwood, D. (1996). A comparative study of texture measures with classification based on featured distributions. *Pattern recognition*, 29(1):51–59.

Plakoyiannaki, E., Mathioudaki, K., Dimitratos, P., and Zotos, Y. (2008). Images of women in online advertisements of global products: does sexism exist? *Journal of Business Ethics*, 83(1):101.

Poland, B. (2016). *Haters: Harassment, Abuse, and Violence Online*. U of Nebraska Press.

Ries, C. X. and Lienhart, R. (2014). A survey on visual adult image recognition. *Multimedia tools and applications*, 69(3):661–688.

Rosenholtz, R., Li, Y., and Nakano, L. (2007). Measuring visual clutter. *Journal of vision*, 7(2):17–17.

Schettini, R., Gasparini, F., Corchs, S., Marini, F., Capra, A., and Castorina, A. (2010). Contrast image correction method. *Journal of Electronic Imaging*, 19(2):023005–023005.

Sharif Razavian, A., Azizpour, H., Sullivan, J., and Carlsson, S. (2014). Cnn features off-the-shelf: an astounding baseline for recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 806–813.

Tamura, H., Mori, S., and Yamawaki, T. (1978). Textural features corresponding to visual perception. *Systems, Man and Cybernetics, IEEE Transactions on*, 8(6):460–473.

Vedaldi, A. and Lenc, K. (2015). Matconvnet – convolutional neural networks for matlab. In *Proceeding of the ACM Int. Conf. on Multimedia*.

Viola, P. and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–I. IEEE.

Zaidan, A., Karim, H. A., Ahmad, N., Zaidan, B., and Sali, A. (2013). An automated anti-pornography system using a skin detector based on artificial intelligence: A review. *International Journal of Pattern Recognition and Artificial Intelligence*, 27(04):1350012.

Zimmerman, A. and Dahlberg, J. (2008). The sexual objectification of women in advertising: A contemporary cultural perspective. *Journal of Advertising Research*, 48(1):71–79.