Gender Identification from Community Question Answering Avatars

BILLY PERALTA1, ALEJANDRO FIGUEROA2, ORIETTA NICOLIS.3, AND ÁLVARO TREWHELA4

1Departamento de Ciencias de la Ingeniería, Facultad de Ingeniería, Universidad Andres Bello, Antonio Varas 880, 8370146 Santiago, Chile; (e-mail: billy.peralta@unab.cl)
2Departamento de Ciencias de la Ingeniería, Facultad de Ingeniería, Universidad Andres Bello, Antonio Varas 880, 8370146 Santiago, Chile; (e-mail: alejandro.figueroa@unab.cl)
3Departamento de Ciencias de la Ingeniería, Facultad de Ingeniería, Universidad Andres Bello, Antonio Varas 880, 8370146 Santiago, Chile; (e-mail: orietta.nicolis@unab.cl)
4Departamento de Ciencias de la Ingeniería, Facultad de Ingeniería, Universidad Andres Bello, Antonio Varas 880, 8370146 Santiago, Chile; (e-mail: a.trewhela@uandresbello.edu)

Corresponding author: Billy Peralta (e-mail: billy.peralta@unab.cl).

ABSTRACT

There are several reasons why gender recognition is vital for online social networks such as community Question Answering (cQA) platforms. One of them is progressing towards gender parity across topics as a means of keeping communities vibrant. More specifically, this demographic variable has shown to play a crucial role in devising better user engagement strategies. For instance, by kindling the interest of their members for topics dominated by the opposite gender.

However, in most cQA websites, the gender field is neither mandatory nor verified when submitting and processing enrollment forms. And as might be expected, it is left blank most of the time, forcing cQA services to infer this demographic information from the activity of their users on their platforms such as prompted questions, answers, self-descriptions and profile images.

There is only a handful of studies dissecting automatic gender recognition across cQA fellows, and as far as we know, this work is the first effort to delve into the contribution of their profile pictures to this task. Since these images are an unconstrained environment, their multifariousness poses a particularly difficult and interesting challenge. With this mind, we assessed the performance of three state-of-art image processing techniques, namely pre-trained neural network models. In a nutshell, our best configuration finished with an accuracy of 81.68% (Inception-ResNet-50), and its corresponding Grad-Cam maps unveil that one of its principal focus of attention is determining silhouettes edges. All in all, we envisage that our findings are going to play a fundamental part in the design of efficient multi-modal strategies.

INDEX TERMS

Community Question Answering, Image Processing, Social Computing, User Demographic Analysis, Computers and Information Processing, Data Systems, Digital Systems, Artificial Intelligence

I. INTRODUCTION

DEMOGRAPHIC variables are often used as proxy measures, when factors of interest are more difficult to identify, conceptualize, or quantify. In many cases, these variables are used as the first set of informative features to take into account in predictive analysis. Here, aspects such as age and gender are known to be reliable predictors of shifting preferences, for example. Business strategies are employed in accordance with the different segments defined by these variables, since it is normally assumed that consumers with similar demographic characteristics have similar preferences (e.g., interests, needs, values, incomes, and buying patterns).

By examining these cohorts, one can not only forecast, but also understand how these differences evolve in a life span [I]. Good examples are changes in personal expenditures as we age, where older people spend half as much on nightlife, entertainment and apparel when compared to younger individuals. Understanding their evolution also helps to design targeted advertisement, where the content of billboards can be visualized based on the demographics of pedestrians (e.g.,
In effect, this evolution in people interests is also reflected in the different social media outlets through their existence, especially in the specific topics they take part and in their sentiments towards them. It goes without saying, people profit from each kind of online service for different purposes, but at the end of the day, the online activities of their members surround and evolve in conformity to their daily life interests and worries.

As opposed to social networks such as Facebook [2] and Twitter [3–8], there is a small amount of studies touching on demographics across cQA sites [9–12]. Even rarer is finding works using these variables as proxies for modeling factors, for which their explicit formulation and accurate forecasting is hard (e.g., user interest or desire to answer a specific question). The reason to this is two-fold: a) some cQA websites do not collect this kind of information, for example Stack Exchange does not store genders and does not enforce a “real name policy” [12, 13]; and b) if asked when enrolling, demographic variables are typically optional and with a rising interest in personal privacy many people choose not to provide this information.

Human gender classification plays a pivotal role in a notable number of real-world applications, including forensic science, vending machines and human–computer interaction systems, in general. Speaking of cQA services, gender identification is essential for several tasks including recognizing malicious activity, deception, filtering and banning fake profiles. Furthermore, it is vital for diversifying and boosting the dynamicity of their platforms, since it can be integrated into tasks such as question routing, expert finding, personalization and dedicated displays [10, 14]. Evidently, presenting diverse outputs aims in part at stirring up the interest of community fellows to acquire knowledge by exploring new topics. In this way, for instance, cQA sites can mitigate gender disparity across topics dominated by men or women [13].

By and large, there is a strong need for research on how gender orientation affects individual participation/contribution in this sort of social network [14].

Although gender is normally left blank when signing up, it can be guessed on the basis of cues derived from word patterns, textual semantic analysis, names and profile pictures [10, 13]. A pioneer work on this is due to [13], who assessed the performance of several combinations of gender recognizers for facial images on cQA profile pictures. The techniques explored in this work fail when images do not contain a human face, or when faces are not shown frontally. Broadly speaking, facial image gender classification is a challenging task due not only to various changes in viewing angles or facial expressions, but also to other factors such as extreme poses and backgrounds. Simply put, more unconstrained conditions entail a more difficult task [15].

Different from its forerunners, our work capitalizes on current state-of-the-art image processing, namely neural network architectures, for guessing genders from cQA profile pictures (avatars for short). Needless to say, this is a sophisticated task, due to the multifariousness of these avatars (see Fig. [1] [9–12]. In brief, our contributions to this body of knowledge are as follows:

1) Contrary to previous investigations, our subject of study are avatars instead of traditional facial profile images. Since cQA websites allow the use of almost any class of picture for this purpose, avatars are more widely diverse, e.g., facial and non-facial images, landscapes, signs, colors, some are automatically generated as well (see Fig. [1] [9–12].

2) We conducted experiments on a massive dataset comprising 186,224 elements. This material was automatically compiled and labelled via matching name catalogues and n-gram expressions closely linked to a specific gender. When tagging an avatar, we benefitted exclusively from its respective textual inputs, namely questions, answers and nicknames, self-descriptions as well.

3) By means of this large-scale corpus, we trained Convolutional Neural Networks (CNNs) classifiers and fine-tuned state-of-the-art pre-trained image recognition architectures.

4) In order to discover discriminative visual patterns, we examined the Grad-Cam heat maps generated by our best configuration [16].

In a statement, our best performance was achieved by an Inception-ResNet-50 that seeks, in part, to delineate body silhouettes. The article is structured as follows. Section II presents an overview of related work, and subsequently section III outlines our research questions. Section IV details our approach, the acquisition and the annotation process of our corpus. In section V the experiments are introduced, followed by a detailed description and analysis of the results. Finally, section VI summarizes the main aspects of the paper and outlines some future directions.

II. RELATED WORK
As far as we can tell, our work is the first to make a focused effort into designing avatar-based machine learning models to automatically detect genders across cQA members. As evidenced by numerous recent comprehensive surveys [14], [17–19], the subject of demographic variables in this field is largely unexplored.

Essentially, our work is at the crossroad of two developing topics within social networks (i.e., cQA) and image processing: demographic user analysis and gender recognition on unconstrained multifarious profile images.

A. GENDER IDENTIFICATION FOR COMMUNITY QUESTION ANSWERING
The work of [10] pioneered efforts to automatically distinguishing genders across cQA questions authors based on their texts, asker demographics, question metadata and web
searches. By using supervised learning models trained with a large-scale dataset, they built several high-dimensional spaces in conformity to different levels of accessible information (e.g., question titles, bodies, metadata and web searches). They discovered that age, industry and second-level question categories were good indicators of the asker gender. When these characteristics were inaccessible, linguistic traits were used in an attempt to deduce them from textual sources, especially semantic and dependency analysis.

Earlier than [10], the research of [11] superficially touched on gender demographics, when looking into the impact of sentiment analysis on cQA websites. More concretely, they focused on the impact of gender on the attitude (i.e., inclination towards positive or negative sentiments) and sentimentality (i.e., amount of sentiments). As a result, their study revealed that a) women are more sentimental when answering; b) in terms of attitude, men are more neutral, whereas women are more positive in their answers; and c) they show a similar behavior across their posted questions.

From another angle, [13] assessed a couple of semi-automatic/heuristics for automatically identifying genders on Stack Overflow. First, they tried to connect cQAs profiles to other social networks where genders are provided. Recall here that Stack Overflow does not explicitly record gender and does not enforce the “real name policy”. Second, they tested the performance of facial recognition tools on 900 manually annotated avatars. Unsurprisingly, this approach performs poorly when non-facial avatars are being analyzed. They also found out that; a) in some cases, it is impossible for humans to determine the gender by an eyeball inspection; and b) some women pretend to be men and the other way around.

Recent works have focused on mitigating the peer disparity in terms of gender across some cQA communities. For instance, [20] discovered that women who encountered other women are more likely to engage sooner than those who did not in Stack Overflow. Later, [21] revealed that women tend to ask more questions while men are likely to provide more answers. As a logical consequence, women have much higher reputation scores on average (less likely to get upvotes). They diminished this gender gap by designing a reputation strategy that rewards points for asking and answering to the same level. All in all, these pieces of research highlight the relevance to user engagement of automatically recognizing genders across cQA websites.

B. GENDER CLASSIFICATION ON IMAGES

By studying a corpus of 19,000 facial images of male, female and children, [22] proposed two deep learning-based methods for gender classification. Out of two strategies, a Convolutional Neural Networks (CNNs) outperformed an Alex Net by 2% (i.e., 92% to 90% accuracy). By the same token, [23] capitalized on VGG-16 and ResNet-50 neural network architectures for automatically recognizing gender across facial images from Malaysians and Caucasians people. In their case, the former finished with an accuracy of 88%, whereas the latter 85%. Likewise, [24] compared the performance of VGG-16 and Alex Net for gender classification on a collection of pictures belonging to women, men, old, young, children, and babies. Their experimental results show that VGG-16 outclasses Alex Net by a large margin.

With the same aim, the works of [25], [26] devised CNNs for age and gender classification, specifically [26] obtained an accuracy of up to 98.5%. Along the same lines, [27] performed gender recognition via using CNNs and Local Recipient Areas Excessive Learning Machine models. Experiments on 11,000 facial images from the Adience dataset showed that the former technique accomplishes 80% accuracy, while the latter 87.13% [28]. In a similar spirit, [29] also benefitted from CNNs for automatic age and gender classification on the same database.

HyperFace was designed by [30] to perform simultaneous facial recognition, pose prediction, and gender recognition using CNNs. Notably, HyperFace-ResNet was based on the ResNet-101 model to enhance their system speed. It is also worth highlighting here the study of [31], who profited from Deep Learning methods for gender prediction across pedestrians. Interestingly enough, their approach segmented people from the image before performing their classification.

From an alternative viewpoint, the work of [32] seeks to segment face images into the following parts: mouth, hair, eyes, nose, skin, and back; and then, it performs gender identification by means of Random Decision Forest classifiers. Closer to our work, [33] introduced a CNN to distinguish between images of human faces from computer generated avatars as part of the ICMLA 2012 Face Recognition Challenge.

C. IMAGE-BASED GENDER RECOGNITION ON OTHER SOCIAL NETWORK PLATFORMS

While most research into gender prediction on online social networks analyze texts, some image-based approaches have come forth in recent years. For instance, both [34], [35] present distinct techniques to combine texts and images for inferring genders on Twitter.

In the same vein, [36] studied several machine learning approaches for gender identification on Twitter users. Their method employed several features related to user profile picture and description, nickname as well. They concluded that they can achieve a classification rate of 82% with a minimum expenditure of resources. On the same subject, [37] trained four distinct classifiers by taking advantage of usernames, nicknames, descriptions and pictures, textual content as well. With regard to profile pictures, they carried out facial recognition by means of Face++ on the whole, a classifier that aggregates these four predictions finished with 93.2% accuracy for English.
III. RESEARCH QUESTIONS

This work enhances the existing body of knowledge in cQA platforms by exploring state-of-the-art supervised models for image-based gender recognition. To the best of our knowledge, this work takes the lead on discovering discriminative visual patterns of genders across their avatars.

These avatars are a challenging subject of study due to several reasons: a) they are small-sized low-resolution images (i.e., 96x96 or 128x128 pixels); b) avatars are very diverse in nature (e.g., facial and non-facial images, some are automatically generated as well); c) even focusing on facial-based recognition is hard due to occlusion or when faces are not shown frontally; and d) visual patterns connected to a specific gender might not be necessarily detected by an ocular inspection.

With prior works as the foundation, we advance this area of research by answering the following four main research questions:

- **RQ1**: How well do vanilla state-of-the-art image classification techniques perform on cQA avatars?
- **RQ2**: Can publicly available image databases cooperate on boosting the performance by pre-training avatar classifiers?
- **RQ3**: What are the limitations of automatically identifying genders across these avatars?
- **RQ4**: What visual patterns are informative of genders across avatars?

IV. GENDER IDENTIFICATION FROM COMMUNITY QUESTION ANSWERING AVATARS

In this section, we outline our corpus acquisition and annotation process, together with introducing the neural network architectures used in our experimental settings.

A. CORPUS ANNOTATION

In our study, we benefited from the corpus compiled by [9], which encompasses 657,805 community member profiles. Each of these records contains the corresponding sets of questions, answers, nicknames and self-descriptions (see Fig. 1). Thus, we capitalized on these textual inputs for automatically assigning each community peer one of two genders (i.e., male or female), whenever it was possible. It is worth noting here that we focused only on the 219,626 (33.39%) community fellows that provided a non-default avatar.

In the same spirit of [15], our automatic annotation process starts off by verifying if a nickname is contained in any of the following seven gender by name archives: a) Howarders [95,025]; b) Arun Babu [117,950]; c) Joerg Michael [48,527]; d) CMU [7,944]; e) WGNDE [177,043]; f) both Miguel Gil’s Spanish list [49,340]; and g) our own compilation of 58 highly recurrent nicknames, which could not be found across the six previous catalogues (e.g., sweetgirl, justagirl, guy and boy).

In general, nicknames can represent not only real names (e.g., devin espinoa and helen_robi), but also strings containing many different characters including numbers and math symbols. Consider the following illustrative examples: “*kimberley*125”, “woody <3’s skateboarding” and email addresses. Consequently, each nickname must be preprocessed before checking as to whether or not it is included in any of the seven name archives. With this aim, we determine and extract its “core substring”, i.e., the substring that denotes the name, by first cleaning and removing special characters.

More concretely, this preprocessing converts nicknames to lower case, and it then removes any character not belonging to the ASCII interval [97, 122] (i.e., [a,z]). Subsequently, it trims each nickname at the first space, hyphen, at, underscore or dot as a means of finding its “core substring”. For our working examples, this preprocessing produces “kimberley” and “woody”; in the case of email addresses, its outcome is the login/user name or its first part (e.g., “billy.peralta@unab.cl” → “billy”). If the resulting string does not match any list, we then start to systematically trim its end one character at a time until a match is found or its length is five characters. This reduction helps to remove some classes of suffixes typically used after names across social networks aliases (e.g., “lauraweird” → “laura”, “ulrich53” → “ulrich” and “dennis372006” → “dennis”). Since each of these databases can return male and/or female, we count the frequency of each possible gender to decide the final label.

At this point, we preliminarily tagged community members by assigning the highest frequent gender whenever there was one. Subsequently, we profit from these preliminary labels for finding gender indicative phrases across their questions, answers and self-descriptions. For this purpose, we took advantage of CoreNLP [10] for tokenizing and splitting sentences, and computing lowercased n-grams afterwards ($n = 2 \ldots 7$). It is worth noting here that we also capitalized on part-of-speech tagging for substituting numbers
FIGURE 1: Illustrative record excerpts corresponding to ten different community fellows. In bold red, phrases indicative of their respective gender. The first row contains self-descriptions, the next one question titles and bodies, and the last row answers.

with a placeholder. After this, these n-grams were ranked in conformity to their Entropy, and low-ranked elements were manually inspected in order to verify if each of them by itself suffices to make a good guess of the gender. Eventually, this process aided in compiling a collection of 1,486 gender indicative phrases (see Table 1).

Accordingly, the next step consists in revising all preliminary labelled and non-labelled community peers by searching for gender indicative phrases across their questions, answers and self-descriptions. Thus, their gender frequency counts are updated by adding to each community fellow his/her counts of male/female aligned phrases. Likewise, the corresponding highest frequent gender was attached to each user when possible.

On the whole, this automatic annotation process assisted in discovering the gender of 186,224 (84.79%) out of the 219,626 community members, previously associated with an non-default avatar. We resized all images to fit 90x90 pixels, and randomly split this dataset into 110,866 training, 37,965 evaluation and 37,393 testing samples. It is important to note here that held-out evaluations were conducted in all our experiments by keeping these splits unchanged. Also, it has to be clarified that we utilize the test dataset for providing an unbiased evaluation of a final model fit on the training/evaluation datasets. The overall distribution is as follows: 126,970 (68.18%) instances are female, whereas 59,254 (31.82%) males.

**Caveats**

Given the nature of our annotation strategy, community members are tagged according to the gender they identify themselves on the website. It goes without saying that some people might run fake profiles or pretend to belong to another gender, at least, from time to time. Deception brings about uncertainty not only to an automatic, but also to a manual, tagging process. Further, there are community fellows who can willingly and/or unwillingly lie, and some errors are due also to the inherent shallow nature of our tagging approach. It worth highlighting here that our preliminary manual inspection did not find that other sexual orientations take a
low participation, it is also difficult to compile a comprehensive list of their typifying names and phrases. Recall that this is a retrospective data collection, which covers the activity that took place before Sept. 2018. As a rule of thumb, we manually judged the agreement amongst the image, name and texts produced by one hundred randomly selected community peers. For 90% of these members, we found no evidence of discrepancy among these three sources.

B. MODELS

In this work, four competitive Deep Neural Networks approaches were tested: standard CNNs, VGG networks, Residual Networks (ResNet) and Inception-ResNet models. Next, we describe each of these architectures.

1) Convolutional Neural Networks

CNNs are comprised of a set of multiple layers which are usually grouped considering blocks of convolution, pooling and activation operations. The filters applied on the convolution operation correspond to local patterns of features. Each filter is represented by a matrix of numbers, which is typically consistent with a visual pattern. A generic forward propagation in a CNN layer consists of three phases. First, the layer performs multiple convolutions in parallel to produce a set of linear matrix transformations. A convolution is produced by applying the filter over the input data by all possible locations. Second, the layer applies a pooling function to reduce the output of the convolution operation. The pooling layer is used to down sampling feature maps by aggregating the presence of these features, where each filter is a pooling function. Typically, the outputs given by both the convolution and pooling operations are processed through a nonlinear activation function. This function conforms to sigmoidal, rectified linear or linear functions.

After parameter training, CNNs learn a series of convolutional filters, which are then optimized with respect to the cost function defined by the learning task (e.g. classification or regression). Fig. 2 details the architecture employed on our experiments. In the beginning, three consecutive convolutional blocks are used. Each of them is formed by a 3x3 kernel, but with a different number of filters: 16, 32 and 64, respectively. Then, a max pooling layer is applied. After passing through two fully connected layers, it is finally classified by the softmax function.

2) VGG Network

This model is a CNN constructed with very small convolutional filters. Its design corresponds to a Deep Neural Network, but with the restriction that its layers depth does not increase the computational complexity. Its original version, VGG-16, is constituted by thirteen convolutional and three fully connected layers.

In our empirical settings, we accounted for a generic implementation of this network (see Fig. 3). Initially, five convolutional blocks consisting of 2, 2, 3, 3 and 3 layers are employed. Each of them implements a 3x3 kernel, but they differ in the amount of filters: 64, 128, 256, 512 and 512, respectively. A max pooling layer is subsequently applied. The last three layers are then fine-tuned, and its outcome is connected to the global average pooling layer, then to a dense layer and finally to second dense layer which outputs the two classes.

3) ResNet model

A Deep Residual Network (ResNet) is a CNN that implements small convolution filters, making its architecture very simple. In contrast to classical CNNs and VGG, ResNet uses residual connections to reduce over-fitting and the gradient vanishing effect inherent of traditional deep networks. The original ResNet is constructed with sets of 1x1 and 3x3 convolutional layers. These layers reduce the representation complexity by extracting high level feature maps of training images.

The architecture of the ResNet assessed in our experiments is sketched in Fig. 4. At its first stage, this model utilizes a convolutional layer with a 7x7 kernel and 64 filters, and employs a 3x3 pooling layer afterwards. The next step consists in applying four convolutional blocks with residual connections. These four blocks are comprised of 3, 4, 6 and 3 convolutional sub-blocks, respectively. Inside each sub-block, three convolutional layers are used with kernels of sizes 1x1, 3x3, and 1x1, respectively. An average pooling layer is subsequently employed. Eventually, the final outcome of the ResNet network coincides with the output of the activation of a fully connected layer. In our experiments, we consider a ResNet model with fifty layers (ResNet-50), where the last layer is linked to the global average pooling layer, then to a dense layer which outputs the two classes.

4) Inception-ResNet model

The Inception-ResNet is a CNN based on the integration of a Deep Residual Network and an improved version of

---

| Gender | i n | Indicative Phrase | Gender | i n | Indicative Phrase | Gender | i n | Indicative Phrase |
|--------|-----|------------------|--------|-----|------------------|--------|-----|------------------|
| F      | 2   | i ovulated       | M      | 2   | my gf            | F      | 4   | i had a miscarriage |
| M      | 4   | up with my girlfriend | F      | 4   | i took the pill  | M      | 5   | i am a warrior, with |
| F      | 3   | my wife           | M      | 3   | amount of a light moisturizer | F      | 5   | i have a crush on a guy |
| M      | 5   | i liked           | M      | 5   | i’ve never had a girlfriend | M      | 7   | so me and my girlfriend have been |

TABLE 1: Sample of phrases signalling gender (i n stands for the number of tokens).
Inception Neural Network [41]–[43]. The Inception family of models are characterized for their multi-branch architectures. Their blocks are constituted by a battery of diverse filters (e.g., 1x1, 3x3 and 5x5) that are concatenated in each branch. In the last dense layers, they have a powerful representational ability due to their procedures based on splitting, transformation, and merging. The residual model enables the training of very deep Inception-ResNet architectures.

In our work, we use the variant named InceptionResNet-v2 network, which efficiently uses residual connections (see Fig. 4). Our implementation first applies a stem block formed by multiple convolutional layers. Next in the pipeline comes a block of five Inception-ResNet-A and one Reduction-A. Subsequently, this model employs ten Inception-ResNet-B and one Reduction-B modules. Then, it utilizes five Inception-ResNet-C and one Reduction-C modules. It is worth noting here that all Inception-ResNet modules are multi-branch convolutional blocks with residual connections of different structures, whereas all Reduction modules are multi-branch convolutional blocks that reduce the complexity of the input. In the next stage, an average pooling layer acts as drop-out. Eventually, the last layer is fully connected and it is in charge of generating the final predictions.

V. EXPERIMENTS

This section shows the experiments conducted to evaluate the performance of the previously described Deep Neural Networks models on gender identification from avatar pictures. We compare the results obtained by pre-trained and non-pre-trained models as well as analyze the most relevant
A. RQ1 AND RQ2: ANALYSIS OF CLASSIFICATION PERFORMANCE.

Classification results were obtained by carrying out an stratified hold-out strategy, where the training, validation and test sets were built as described in Section IV-A. In all our configurations, these partitions were unchanged. Although an stratified cross-validation approach is feasible, we believe that the amount of available data is large enough to apply hold-out evaluations. In details, parameters were optimized in the training set. The final model of each neural network was chosen in conformity to its classification rate on the validation set. Note that the test dataset was used solely for providing an unbiased assessment of a final model fit on the training/evaluation datasets. Unless otherwise stated, all reported figures correspond to the respective evaluation on the test set.

Before conducting experiments on neural network models, we designed a sanity-check baseline. For this purpose, Liblinear classifiers were implemented with different intensity histogram features. We accounted for all classification models made available by this library, and took advantage of its algorithm for finding the best cost parameter (C). The best liblinear model achieved an accuracy of 72.57% by means of L2-regularized L2-loss support vector classification (primal) and a very small value of C (3.8147e-06). Additionally, we experimented with some basic image segmentation techniques including quantization and alternative representations [12]. From all these techniques, the HSV representations of avatars was the only fruitful one. When modelling pictures according to how their colors appear under light, we reaped a marginal improvement of 0.32% (72.89%, C=1.2207e-4).

As for neural network models, we used the implementa-

tions provided by the Keras toolkit [45]. Note that all three pre-trained networks were built on top of the ImageNet database. In all cases, we use a fine-tuning approach where we train the second to last dense layer, which comprises 1024 weights, and leave all other layers frozen with their respective pre-trained weights. It should also be pointed out that the same number of iterations (equal to 100) was set all the time, and we consider an early-stopping scheme based on the cost function on the validation set. Accordingly, we report the Accuracy, Precision, Recall and Macro F1-Score obtained by each model.

Overall, the Inception-Resnet and VGG-16 neural networks outperform the other two alternatives. Regarding Accuracy, Inception-Resnet obtains a marginal improvement over the nearest second, VGG. Both Recall and Macro F1-Score yield similar results, while Precision is slightly higher for VGG-16. It is worth emphasising that the performance of the most basic neural network, CNN, is notoriously worse than its rival models, due partly to the fact that it does not benefit from pre-trained weights (data). As a means of confirming this, additional experiments were carried out in all networks using a random initialization of weights. As a result, Accuracy ranged from 70% to 72% (close to our sanity-check baseline), thus supporting the relevance of the patterns inferred from large databases like ImageNet.

### TABLE 2: Results accomplished by each neural network on our avatar corpus (test set).

| Model          | Accuracy | Precision | Recall | MF1   |
|----------------|----------|-----------|--------|-------|
| CNN            | 78.73    | 0.7654    | 0.7231 | 0.7567|
| VGG-16         | 81.46    | 0.8039    | 0.7530 | 0.7696|
| ResNet-50      | 80.75    | 0.7924    | 0.7469 | 0.7621|
| Inception-ResNet-50 | 81.68 | 0.8024    | **0.7608** | **0.7774** |

In a complementary way, we inspected the corresponding confusion matrices in order to gain some insight into the performance by class (see Fig. 7). In light of the figures, we can draw the following conclusions:
1) The error rate for the female class ranges from 7.57% (VGG) to 9.82% (CNN), whereas from 39.51% (Inception-ResNet) to 45.56% (CNN). This suggests that the different inference processes of these neural network models are more sensitive to male avatars.

2) There is a correlation between the overall accuracy and the error rate across male avatars (see Table 2): 39.51% (Inception-ResNet, Acc. 81.68%), 41.83% (VGG, Acc. 81.46%), 42.15% (ResNet, Acc. 80.75%) and 45.56% (CNN, Acc. 78.73%).

In summary, confusion matrices reveal that the good performance accomplished by Inception-ResNet is due mainly to its more effective way of deducing patterns, especially those informative of male avatars.

Metrics like Accuracy and Precision are inherently sensitive to class skews. Since Receiver Operating Characteristics (ROC) are grounded on True Positive Rate (TPR) and False Positive Rate (FPR), they are concave downwards functions insensitive to changes in class distributions. In this graph, lines changing concavity (inflection point) closer to the upper-left corner are better because they correspond to classifiers with lower expected cost [46]. Essentially, Fig. 6 displays outcomes similar to the previous metrics in qualitative terms, that is to say Inception-ResNet performs slightly better than ResNet and VGG.

Given the fact that “more north-west” inflection points imply better expected performance, it is natural to think on the Area Under the Curve (AUC) as an extra metric to compare these classifiers. In this regard, all three pre-trained architectures achieve pretty tight results. In other words, they all have an almost equivalent probability of ranking a randomly chosen female sample higher than a randomly selected male instance.

In view that Inception-ResNet outputs probability distributions over the target classes, we can plot the amount of male and female samples at different confidence levels, i.e., likelihood of belonging to the positive (female) group (see
Fig. 8. On the one hand, this figure clearly shows that the chances of being an actual female systematically increases as long as this value goes up (true positive). And the other way around, as long as this score decreases, the higher the probability of being a genuine member of the male cohort (true negative). Note also that the largest fraction of confidence scores given to females concentrates around 0.85-0.97, while to men around 0.04-0.13. These two extreme and opposite concentrations signal that most of the time there is strong evidence for one gender only, and that Inception-ResNet is able to discover it consistently. However, on the other hand, Fig. 8 also depicts two slight upward trends of misclassified instances, one in each of these two concentrations, found at each extreme. We interpret this as errors stemming from our automatic tagging, or as a consequence of people that identify themselves and/or pretend to belong to the opposite gender when interacting in the community.

**FIGURE 8: Probability/Score Assigned to the Female Class vs. Number of Actual Female/Male Samples (Inception-ResNet-50).** In dark colors, correctly classified male (red) and female (blue). In light red and blue, misclassified male and female, respectively.

**TABLE 3: Significance tests (p-value [Acc > NIR]).**

| Model          | Acc.  | CI 95%      | NIR  | P-Value |
|----------------|-------|-------------|------|---------|
| CNN            | 78.73 | (0.7831, 0.7914) | 0.68 | <2.2e-16 |
| VGG-16         | 81.46 | (0.8106, 0.8185) | 0.68 | <2.2e-16 |
| ResNet-50      | 80.75 | (0.8035, 0.8115) | 0.68 | <2.2e-16 |
| Inception-ResNet-50 | 81.68 | (0.8128, 0.8207) | 0.68 | <2.2e-16 |

The overall Accuracy rate is significantly greater than the largest class for each model, meaning that the models are working well with unbalanced data.

**B. RQ3: Qualitative Analysis of Classification Models.**

In this section, we qualitatively analyze the automatic classifications performed on our avatar database. More precisely, the focus of this analysis is on the outcomes produced by Inception-Resnet, since it was the model that achieved the best prediction rate. Here, we globally examined its classifications by viewing the avatar images within the test set. Accordingly, Figures 9-12 highlight some random instances of pictures within both categories sorted by their respective hits and errors. Some findings are as follows:

1) Visual analysis reveals that Inception-Resnet performs better when avatars exhibit patterns, which intuitively correspond to their proper cohorts. In particular, we observe that errors occur in cases where users display avatars portraying someone of the opposite gender, e.g. a male user displaying a female body on his picture. This is along the lines with our conclusions drawn from Fig. 8.

2) More concretely, avatars corresponding to images of virtual women (Figures 9a-9d and 9i) and of real women (Figures 9e-9f) are usually correctly recognized.

3) Further, there are less obvious patterns that this network is able to infer such as woman’s eyes (Fig. 9g) and beaches (Fig. 9k) as well as sweet animals (Fig. 9f). Generally speaking, we observed that landscapes are strongly connected to female members.

4) Likewise, avatars portraying virtual men (Figures 10a-10d and 10f) and real men (Figures 10e-10i and 10j) are typically correctly identified by the neural network. On top of that, it is also capable of learning some abstract masculine patterns as sketched in Fig. 10g-10h.

5) Some error cases within the female category are highlighted in Figure 11. We observe that in the event of avatars with pictures of men (see Fig. 11a-11c and 11d), the network is usually wrong. Errors can also occur in ambiguous images such as couples (Fig. 11e), a woman near a car (Fig. 11f), or symbols (Fig. 11g).

6) Similarly to the female class, this neural network misclassified pictures showing virtual women tagged as male (Fig. 12a - 12d). In addition, there are some images with ambiguous patterns like the animals in Fig. 12b-12d.

7) In some cases, this network got it right, and it was also easy for an ocular inspection to agree with the model predictions, but their corresponding automatic annotations claimed the contrary label (see Fig. 11b, 11h, 11l, 12e, 12i). This can be a consequence of misannotations, of someone members portraying the picture of the opposite gender, impersonating, running fake profiles, or of people identifying themselves as
FIGURE 9: True positives. Normally, female appearances were correctly guessed as females.

FIGURE 10: True negatives. In general, male-looking avatars were correctly classified.
FIGURE 11: False positives.

FIGURE 12: False negatives.
if they were part of the opposite cohort when sharing content on the platform.

In a statement, Inception-Resnet was able to deduce effective visual patterns that are informative of both genders across cQA avatars. Concerning errors, it will be always possible for community fellows to capitalize on images related to the opposite gender, because these pictures display their friends, relatives, or due to any other unidentifiable reason. Since this would be unavoidable, in such cases, we conjecture that a visual classification algorithm has a certain error rate that is not feasible to improve without the help of other sources of information including texts and/or activity graph patterns. Nevertheless, it is crystal clear that one of the great advantages of visual over textual gender classification it is that the former is a language independent approach.

### C. RQ4: UNDERSTANDING PREDICTIONS VIA GRAD-CAM HEAT MAPS.

In this section, we analyze the interpretability of the model learnt by Inception-Resnet as it relates to how it manages to discriminate by gender. To do this, we obtain the heat maps of relevance according to the Grad-Cam network applied to a randomly selected battery of avatars, both correctly and wrongly classified [48].

The Grad-Cam neural network generates a heat map, where the most relevant parts of a classified image are shown. It produces this heat map every time the classification is made considering the accumulated error in a typically global pooling layer. In our work we use a convolutional layer activation output as input to the Grad-Cam network, specifically the Keras activation \_74 Inception-Resnet network layer [45].

Like Sec. V-B Fig. 13 - 16 present instances for both categories sorted by both hits and errors. To facilitate this analysis, we provide the heat maps for the same arrays of random pictures shown in that section.

By inspecting these heat maps, we discover that informative regions often vary on a case-by-case basis. Nevertheless, these maps reveal one prominent pattern: body silhouettes. The network focuses on delineating these edges/contours regardless if it is dealing with virtual (Fig. 13a - 13d - 13i) or real (Fig. 13e - 13f) avatars. It is key to note here that a greater variety of body silhouette patterns was observed in the case of male pictures (Fig. 14a - 14i). Interestingly enough, in the case of an eye, it also tends to focus on its contour and pupil (Fig. 13g); while in the case of a beach, the vegetation is in its spotlight (Fig. 13h). Other interesting contrasts are due to images such as a) a cat, where the grid is centered on the object itself (Fig. 14g); and b) a guitar, where the network targets at the outline and the background (Fig. 14h).

With regard to misclassifications, we found out that the network follows patterns similar to the hits. For example, in the case of virtual or real pictures of people (Fig. 15a - 15d - 15i), the model also aims at body silhouettes. In like manner, the network centers on the contour when dealing with the couple, but this time following an unclear pattern (Fig. 15e), while in the image of a woman and a car, it focuses on the intersection of both elements (Figure 15f).

On the other hand, in more complex pictures such as a man in a library (Fig. 15d) and the logo (Fig. 15g), the network directs its attention to much of the image. Lastly, some pictures do not show a well-defined pattern like animals (Fig. 16a - 16b).

To sum this up, we found out that one of the focal points of this model is the outline of objects of interest. This is also an aspect that human users would use. However, it is surprising that the network does not target specifically at avatar faces. We conjecture that this is due to the high variance intrinsic of avatar images, which makes it difficult to obtain visual patterns. On the other hand, this observation suggests as an alternative the incorporation of face detection technology to enhance the classification rate.

### VI. CONCLUSION

As far as we know, this work leads the way on the contribution of profile pictures to identify genders across cQA members. In so doing, the performance of three state-of-art pre-trained neural network models was evaluated. In short, the best system finished with an accuracy of 81.68% (Inception-ResNet-50). A remarkable finding is that Grad-Cam heat maps disclosed that this model directs its attention to delineate body silhouettes.

Since avatar pictures are an unconstrained environment, their multifariousness poses a particularly difficult and interesting challenge. We envisage that capitalizing on frontier facial recognition technology will bring about notable improvements. We also envision that pre-training distinct neural networks architectures on top of a massive corpus of avatars will lead to substantial gains in performance.

### REFERENCES

[1] S. A. Mirlohi Falavarjani, F. Zarrinkalam, J. Jovanovic, E. Bagheri, and A. A. Ghorbani, "The reflection of offline activities on users' online social behavior: An observational study," Information Processing & Management, vol. 56, no. 6, p. 102070, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0020593619305981

[2] H. A. Schwartz, J. C. Eichstaedt, M. L. Kern, L. Dzurzynski, S. M. Ramones, M. Agrawal, A. Shah, M. Kosinski, D. Stillwell, M. E. Selgman et al., “Personality, gender, and age in the language of social media: The open-vocabulary approach,” PloS one, vol. 8, no. 9, p. e73791, 2013. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0073791

[3] F. Rangel, P. Rosso, M. Koppell, E. Stamatatos, and G. Inches, “Overview of the author profiling task at pan 2013,” in CLEF Conference on Multilingual and Multimodal Information Access Evaluation. CELCT, 2013, pp. 352–365.

[4] F. Rangel, P. Rosso, I. Chugur, M. Potthast, M. Trenkmann, B. Stein, B. Verhoeven, and W. Daelemans, “Overview of the 2nd author profiling task at pan 2014,” in CLEF 2014 Evaluation Labs and Workshop Working Notes Papers, Sheffield, UK, 2014, pp. 1–30.

[5] F. Rangel, P. Rosso, B. Verhoeven, W. Daelemans, M. Potthast, and B. Stein, “Overview of the 4th author profiling task at pan 2016: cross-genre evaluations,” in Working Notes Papers of the CLEF 2016 Evaluation Labs, CEUR Workshop Proceedings/Balog, Kristian [edit.]; et al., 2016, pp. 750–784.

[6] F. M. Rangel Pardo, F. Celli, P. Rosso, M. Potthast, B. Stein, and W. Daelemans, “Overview of the 3rd author profiling task at pan 2015,” in CLEF
FIGURE 13: Inception-Resnet heat maps for true positives. In general, avatar silhouettes are its focal point.

FIGURE 14: Inception-Resnet heat maps for true negatives. At large, body contours are in the spotlight of this model.
FIGURE 15: Inception-Resnet heat maps for false positives.

FIGURE 16: Inception-Resnet heat maps for false negatives.
[18] S. Tilki, H. Dogru, A. Hameed, A. Jamil, J. Rasheed, and E. Alimovski, “Understanding the reputation differences between women and men on stack overflow,” in 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2015, pp. 34–42.

[19] D. Ford, “Recognizing gender differences in stack overflow usage: Applying the bechdel test,” in 2016 IEEE Symposium on Virtual Languages and Human-Centric Computing (VL/HCC), 2016, pp. 264–265.

[20] B. Lin and A. Serebrovnik, “Recognizing gender of stack overflow users,” in Proceedings of the 13th International Conference on Mining Software Repositories, ser. MSR ’16. New York, NY, USA: Association for Computing Machinery, 2016, p. 425–429. [Online]. Available: https://doi.org/10.1145/2845295.2845371

[21] A. Ahmad, C. Feng, S. Ge, and A. Yousif, “A survey on mining stack overflow: question and answering (qa) community,” Data Technologies and Applications, vol. 52, no. 2, pp. 190–247, 2018.

[22] M. Islam, N. Tasnim, and J. H. Baek, “Human gender classification using transfer learning via parece frontier cnn networks,” Inventions, vol. 5, no. 2, 2020. [Online]. Available: https://www.mdpi.com/2411-5134/5/2/16

[23] C. Szegedy, W. Jomah, A. Gachagan, and S. Marshall, “Appearance based pedestrians’ gender recognition by employing stacked auto encoders in deep learning,” Future Generation Computer Systems, vol. 88, pp. 38–39, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167812517322388

[24] E. Eidinger, R. Enbar, and T. Hassner, “Age and gender estimation of unfiltered faces,” IEEE Transactions on Information Forensics and Security, vol. 9, no. 12, pp. 2170–2179, 2014.

[25] G. Levi and T. Hassner, “Age and gender classification using convolutional neural networks,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, June 2015.

[26] S. Arora and M. Bhatia, “A robust approach for gender recognition using deep learning,” in 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2018, pp. 1–6.
REFERENCES

[1] S. A. Mirlohi Falavarjani, F. Zarrinkalam, J. Jovancovic, E. Bagheri, and A. A. Ghorbani, “The reflection of online activity on users’ online social behavior: An observational study,” Information Processing & Management, vol. 56, no. 6, p. 102089, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306437719309681

[2] H. A. Schwartz, J. C. Eichstaedt, M. L. Kern, L. Dziurzynski, M. E. J. listings, “Personality, gender, and age in the language media: The open-vocabulary approach,” PloS one, vol. 8, no. 9, p. e73791, 2013. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0073791

[3] F. Rangel, P. Rosso, M. Koppell, E. Stamatakis, and G. Inches, “Overview of the author profiling task at pan 2013,” in CLEF Conference on Multi-lingual and Multimodal Information Access Evaluation. CELECT, 2013, pp. 352–365.

[4] F. Rangel, P. Rosso, I. Chugur, M. Potthast, M. Trenkmann, B. Stein, B. Verhoeven, and W. Daleemens, “Overview of the 2nd author profiling task at pan 2014,” in CLEF 2014 Evaluation Labs and Worksho... Findings, in Working Notes Papers of the CLEF 2016 Evaluation Labs, CEUR Workshop Proceedings/Balog, Krisztian [edit.]; et al., 2016, pp. 750–784.

[6] F. M. Rangel Pardo, F. Celli, P. Rosso, M. Potthast, B. Stein, and W. Daleemens, “Overview of the 3rd author profiling task at pan 2015,” in CLEF 2015 Evaluation Labs and Worksho... Papers, Sheffield, UK, 2014, 2014, pp. 1–30.

[7] A. Cutola, N. Kumar, and J. Cutler, “Predicting the demographics of twitter users from website traffic data,” in AAAI, 2015.

[8] A. Cutola, N. K. Ravi, and J. Cutler, “Predicting twitter user demographics using indirect supervision from website traffic data,” Journal of Artificial Intelligence Research, vol. 49, p. 389, 2014.

[9] A. Figueroa, B. Peralta, and O. Nicolis, “Coming to grips with age: A survey of approaches for community question answering,” ACM Trans. Web, vol. 10, no. 3, Aug. 2016. [Online]. Available: https://doi.org/10.1145/2934887

[10] J. M. Jose and J. Thomas, “Finding best answer in community question answering sites: a review,” in 2018 International Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET), Dec 2018, pp. 1–5. [Online]. Available: https://ieeexplore.ieee.org/document/8821219

[11] D. Ford, A. Harkins, and C. Parvin, “Someone like me: How does peer influence participation of women on stack overflow?” in 2017 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC), 2017, pp. 239–243.

[12] Y. Wang, “Understanding the reputation differences between women and men on stack overflow,” in 2018 25th Asia-Pacific Software Engineering Conference (APSEC), 2018, pp. 436–444.

[13] S. Tilki, H. Dogru, A. Hameed, A. Jani, J. Rasheed, and E. Alimovski, “Gender classification using deep learning techniques,” 05 2021.

[14] T. V. Janahuraman and P. Subramaniam, “Gender classification based on asian faces using deep learning,” in 2019 IEEE 9th International Conference on System Engineering and Technology (ICSET), 2019, pp. 84–89.

[15] G. Guanzo and I. H. Fernández, “Detergender algorithm: kurnar görtününtu cinsiyeti tanımı,” 2019.

[16] G. Levi and T. Hassner, “Age and gender classification using convolutional neural networks,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, June 2015.

[17] S. Arora and M. Bhatia, “A robust approach for gender recognition using deep learning,” in 2018 9th International Conference on Computing, Communication and Networking Technologies (CCNT), 2018, pp. 1–6.

[18] Y. Akbulut, A. Sengür, and S. Ekiçi, “Gender recognition from face images with deep learning,” in 2017 International Artificial Intelligence and Data Processing Symposium (IDAP), 2017, pp. 1–4.

[19] E. Eidinger, R. Enbar, and T. Hassner, “Age and gender estimation of un-filtered faces,” IEEE Transactions on Information Forensics and Security, vol. 9, no. 12, pp. 2170–2179, 2014.

[20] G. Levi and T. Hassner, “Age and gender classification using convolutional neural networks,” in 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2015, pp. 34–42.

[21] R. Ranjan, V. M. Patel, and R. Chellappa, “Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 1, pp. 121–135, 2019.

[22] M. Raza, M. Sharif, M. Yasmin, M. A. Khan, T. Saba, and S. L. Fawwaz, “Appearance based pedestrians’ gender recognition by employing stacked auto encoders in deep learning,” Future Generation Computer Systems, vol. 88, pp. 28–39, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167739X17322288

[23] K. Khan, M. Attique, I. Syed, and A. Gul, “Automatic gender classification through face segmentation,” Symmetry, vol. 11, no. 6, 2019. [Online]. Available: https://www.mdpi.com/2078-2489/12/2/48

[24] J. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016.

[25] M. Vicente, F. Batista, and J. P. Carvalho, “Gender detection of twitter users based on multiple information sources,” in Interactions Between Users and Internet Systems: A Case Study on Twitter User Gender Recognition, 2016, pp. 1–6.

[26] S. Arora and M. Bhatia, “A robust approach for gender recognition using deep learning,” in 2018 9th International Conference on Computing, Communication and Networking Technologies (CCNT), 2018, pp. 1–6.

[27] Y. Akbulut, A. Sengür, and S. Ekiçi, “Gender recognition from face images with deep learning,” in 2017 International Artificial Intelligence and Data Processing Symposium (IDAP), 2017, pp. 1–4.

[28] E. Eidinger, R. Enbar, and T. Hassner, “Age and gender estimation of un-filtered faces,” IEEE Transactions on Information Forensics and Security, vol. 9, no. 12, pp. 2170–2179, 2014.

[29] G. Levi and T. Hassner, “Age and gender classification using convolutional neural networks,” in 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2015, pp. 34–42.

[30] R. Ranjan, V. M. Patel, and R. Chellappa, “Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 1, pp. 121–135, 2019.

[31] M. Raza, M. Sharif, M. Yasmin, M. A. Khan, T. Saba, and S. L. Fawwaz, “Appearance based pedestrians’ gender recognition by employing stacked auto encoders in deep learning,” Future Generation Computer Systems, vol. 88, pp. 28–39, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167739X17322288

[32] K. Khan, M. Attique, I. Syed, and A. Gul, “Automatic gender classification through face segmentation,” Symmetry, vol. 11, no. 6, 2019. [Online]. Available: https://www.mdpi.com/2078-2489/12/2/48

[33] B. He, J. Liu, and Y. Xu, “Convolutions applied to human face classification,” in 2012 11th International Conference on Machine Learning and Applications, vol. 2, 2012, pp. 580–583.

[34] S. Sakaki, Y. Miura, X. Ma, K. Hattori, and T. Okuhma, “Twitter user gender inference using combined analysis of text and image processing,” in Proceedings of the Third Workshop on Vision and Language. Dublin, Ireland: Dublin City University and the Association for Computational Linguistics, Aug. 2014, pp. 54–61. [Online]. Available: https://aclanthology.org/W14-5408

[35] L. Geng, K. Zhang, X. Wei, and X. Feng, “Soft biometrics in online social networks: A case study on twitter user gender recognition,” in 2017 IEEE Winter Applications of Computer Vision Workshops (WACVW), 2017, pp. 1–8.

[36] D. Fernandez, D. Moctezuma, and O. S. Siordia, “Features combination for gender recognition on twitter users,” in 2016 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC), 2016, pp. 1–6.

[37] M. Vicente, F. Batista, and J. P. Carvalho, “Gender detection of twitter users based on multiple information sources,” in Interactions Between Users and Internet Systems: A Case Study on Twitter User Gender Recognition, 2016, pp. 1–6.

[38] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016. [Online]. Available: http://www.mdpi.com/2078-2489/12/2/48

[39] G. E. Newland, W. L. S. Ngah, and S. Marshall, “Activation functions: comparison of trends in practice and research for deep learning,” in 2nd International Conference on Computational Sciences and Technology, 2021, pp. 124–133.
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3130078, IEEE Access

PERALTA et al.: Gender Identification from Community Question Answering Avatars

K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv preprint arXiv:1409.1556, 2014.

K. He, X. Zhang, S. Ren, and J. Sun, “Identity mappings in deep residual networks,” in European conference on computer vision. Springer, 2016, pp. 630–645.

C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, “Inception-v4, inception-resnet and the impact of residual connections on learning,” in Thirty-first AAAI conference on artificial intelligence, 2017.

C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2818–2826.

T. Fawcett, “An introduction to ROC analysis,” Pattern Recognition Letters, vol. 27, no. 8, pp. 861–874, 2006, rOC Analysis in Pattern Recognition. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0262885603000696

F. Chollet et al., “Keras,” https://keras.io, 2015.

Y.-C. Cheng and S.-Y. Chen, “Image classification using color, texture and regions,” Image and Vision Computing, vol. 21, no. 9, pp. 759–776, 2003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0262885603000696

K. Max, “Building predictive models in r using the caret package.” Journal of Statistical Software, vol. 28, 11 2008.

Y. Bengio, “Deep learning,” Scholarpedia, vol. 7, no. 8, 2012.

K. He, X. Zhang, S. Ren, and J. Sun, “Identity mappings in deep residual networks,” in European conference on computer vision. Springer, 2016, pp. 630–645.

BILLY PERALTA received the M.S and Ph.D. degree in Computer Sciences from Pontificia Universidad Católica, Chile, in 2008 and 2013, respectively. From 2004 to 2018, he was an Assistant Professor with the Universidad de Temuco, Chile. Since 2018 he has been an Assistant Professor with the Universidad Andres Bello, Chile. He is the author of 30 articles, and he is member of IEEE Society and Chilean Society of Computer Science. His research interests lie in the area of machine learning and computer vision.

ÁLVARO TREWHELA was born in Santiago, Chile in 1990. Since 2018 he has been studying a B.S. in computer science at the Andrés Bello University, Chile. His research interests include natural-language processing, image processing, and machine learning, in general.

ALEJANDRO FIGUEROA is an associate professor with the Faculty of Engineering, Universidad Andres Bello, Santiago, Chile. His research interests include natural-language processing, machine learning, context grounding and multimodality in question-answering systems, information retrieval as well. Figueroa received a Ph.D. in computational linguistics from the Universitats des Saarlandes, Saarbruecken, Germany.

ORIETTA NICOLIS earned a degree in Economics in 1995 at the University of Verona, Italy. She obtained her Ph.D degree in Statistics in 1999 at the University of Padua (Italy) and a postdoctoral fellowship in statistics at the University of Brescia (Italy) in the following two years. From 2002 to 2012, she has worked as Researcher and Aggregate Professor of Statistics at the University of Bergamo (Italy). From 2012 to 2018 she was working at University of Valparaíso where she was Director of the PhD Program in Statistics. Since August 2018, she is Full Professor at the Engineering Faculty of the University Andres Bello en Vinà del Mar (Chile) where she is director of the Master in Computation Sciences. She is also responsible of national projects on artificial intelligence and statistical models. Dr. Nicolis is author of more than 60 international publications and reviewer of several international scientific journals. Her research interests include the study of spatio-temporal models, machine learning methods, deep learning, big data, computer science, wavelet-transforms, fractional, and multifractal processes.