Received: 17 November 2021 | Accepted: 14 August 2022
DOI: 10.1111/acer.14925

ORIGINAL ARTICLE

ALCOHOLISM CLINICAL & EXPERIMENTAL RESEARCH

Development and validation of the Alcoholic Beverage Identification Deep Learning Algorithm version 2 for quantifying alcohol exposure in electronic images

Abraham Albert Bonela[1,2] | Zhen He[2] | Thomas Norman[1,3] | Emmanuel Kuntsche[1]

[1]Centre for Alcohol Policy Research, La Trobe University, Melbourne, Victoria, Australia

[2]Department of Computer Science and Information Technology, La Trobe University, Melbourne, Victoria, Australia

[3]Australian Research Centre in Sex, Health and Society, La Trobe University, Melbourne, Victoria, Australia

Correspondence
Emmanuel Kuntsche, Centre for Alcohol Policy Research, La Trobe University, Building NR1, Bundoora 3086, Melbourne, Vic., Australia.
Email: e.kuntsche@latrobe.edu.au

Abstract

Background: Seeing alcohol in media has been demonstrated to increase alcohol craving, impulsive decision-making, and hazardous drinking. Due to the exponential growth of (social) media use it is important to develop algorithms to quantify alcohol exposure efficiently in electronic images. In this article, we describe the development of an improved version of the Alcoholic Beverage Identification Deep Learning Algorithm (ABIDLA), called ABIDLA2.

Methods: ABIDLA2 was trained on 191,286 images downloaded from Google Image Search results (based on search terms) and Bing Image Search results. In Task-1, ABIDLA2 identified images as containing one of eight beverage categories (beer/cider cup, beer/cider bottle, beer/cider can, wine, champagne, cocktails, whiskey/cognac/brandy, other images). In Task-2, ABIDLA2 made a binary classification between images containing an "alcoholic beverage" or "other". An ablation study was performed to determine which techniques improved algorithm performance.

Results: ABIDLA2 was most accurate in identifying Whiskey/Cognac/Brandy (88.1%) followed by Beer/Cider Can (80.5%), Beer/Cider Bottle (78.3%), and Wine (77.8%). Its overall accuracy was 77.0% (Task-1) and 87.7% (Task-2). Even the identification of the least accurate beverage category (Champagne, 64.5%) was more than five times higher than random chance (12.5% = 1/8 categories). The implementation of balanced data sampler to address class skewness and the use of self-training to make use of a large, secondary, weakly labeled dataset particularly improved overall algorithm performance.

Conclusion: With extended capabilities and a higher accuracy, ABIDLA2 outperforms its predecessor and enables the screening of any kind of electronic media rapidly to estimate the quantity of alcohol exposure. Quantifying alcohol exposure automatically through algorithms like ABIDLA2 is important because viewing images of alcoholic beverages in media tends to increase alcohol consumption and related harms.

KEYWORDS
alcohol exposure, alcoholic beverage recognition, artificial intelligence, deep learning, images, media

INTRODUCTION

As of January 2022, there are 4.95 billion global internet users (Kemp, 2022). Of these, 4.62 billion are active social media users—an increase of 424 million new users compared with January 2021 (Kemp, 2022). Internet users aged between 16 and 64 spend an average of 2 h 27 min daily on social media (Kemp, 2022). Social media platforms such as Facebook (2.9 billion), YouTube (2.6 billion), Instagram (1.5 billion) have at least 1 billion active users monthly (Kemp, 2022). Agents within the alcohol industry use social media to promote their brands by employing range of strategies through sharing different types of posts such as images, videos, links, and status updates, with the type of post depending on the platform and marketing strategy (McCreanor et al., 2013). One of the strategies employed by social media alcohol brand marketers is promoting their products through influencer posts (Alhabash et al., 2021). Alcohol brands tend to make the most of images—such as on Facebook or Instagram—to promote their products because social media followers are more engaged to images compared with videos, links, and statuses (Feehan, 2021).

Exposure to alcohol-related cues such as seeing an image of alcoholic beverage is likely to increase unconscious bias to focus on alcohol, increase craving (Schoenmakers et al., 2008) and impulsive decision-making, and impair inhibition drives that control consumption (Kuntsche et al., 2020). In both adolescents and adults, exposure to digital alcohol marketing was found to be associated with drinking initiation among non-drinkers (Anderson et al., 2009; Smith & Foxcroft, 2009), increased levels of consumption among drinkers (Anderson et al., 2009), and binge or hazardous drinking behavior (Noel et al., 2020). A recent study (Lauckner et al., 2019) found that 56% of young adults between the age of 18 and 25 report alcohol use, and 37% report binge drinking and are frequently exposed to alcohol images on social networking sites. To vie with the ever-growing degree of alcohol advertising and alcohol-related content on social networking sites, rapid screening tools need to be developed. Such alcohol screening tools can be used to monitor the quantity of alcohol exposure as well as filter unwanted exposure in both reactive and proactive ways (Norman et al., 2021). Although social networking sites and other sources of digital media serve as a platform for advertising and marketing in various ways, at the very core they still use images, text, and video representations. For the scope of this research article, we intend to focus primarily on working with one of the most fundamental visual representation types, i.e., images/photos, rather than a specific platform.

For this purpose, we developed the Alcoholic Beverage Identification Deep Learning Algorithm (ABIDLA) (Kuntsche et al., 2020) that can rapidly screen images in electronic media to identify alcoholic beverages within the images and can reduce the burden of manual annotation. ABIDLA (Kuntsche et al., 2020) was trained on 67,186 images to identify six beverage categories: Beer Cup (or Glass), Beer Bottle, Beer Can, Wine, Champagne, and Others. ABIDLA has a classification accuracy of 73.8% in identifying the above six categories and 85.2% in identifying between alcoholic beverage images (of beer/wine/champagne) and any other images (Kuntsche et al., 2020). The source code required to implement ABIDLA is publicly available at https://github.com/abrahamalbert18/ABIDLA and has been applied in recent research to identify alcoholic beverages in Instagram images taken at a music festival (Norman et al., 2021).

Although ABIDLA (Kuntsche et al., 2020) has demonstrated promising capabilities in identifying alcohol exposure in images with high accuracy, it has few limitations which requires the development of a more comprehensive—in terms of identifying more alcoholic beverages—and improved algorithm—in terms of performance. First, ABIDLA was trained to exclusively identify the most easily recognizable alcoholic beverages (i.e., beer and wine/champagne). Second, ABIDLA was trained using the images downloaded from a Google image search, and these images contain beverages portrayed in social contexts (such as a party setting). This is a limitation because training on one domain of images (such as Google image search results) leads to the introduction of bias towards that particular domain of images and an over-reliance on the contextual features of the image rather than the beverage itself (Norman et al., 2021).

To overcome the above mentioned limitations and make ABIDLA (Kuntsche et al., 2020) more versatile, as well as to improve its overall classification accuracy, we aimed to develop a new algorithm called ABIDLA2. First, we trained ABIDLA2 to be more comprehensive in terms of identifying a wider number of alcoholic beverages, including Cider, Cognac, Brandy, Cocktails, and Whiskey. However, due to high resemblance and chromatic similarity among the beverages, we merged some of these categories with those existing in ABIDLA—Beer/Cider Cup, Beer/Cider Bottle, Beer/Cider Can, Wine, Champagne, Others, Cocktails, and Whiskey/Cognac/Brandy. Second, we added "stand-alone" alcoholic beverage images—images of beverages without any background or context from Bing and Google Image Search engines. Third, we aimed to increase the classification accuracy of ABIDLA2 by using the following techniques: balanced data sampler—to address the class skew and produce uniform accuracy across categories (to prevent bias towards beverage category)—and self-training—to make use of a large secondary weakly labeled dataset (where search terms were used as the labels rather than the manual annotations), which is important because annotating several thousands of images is time and labour intensive.

MATERIALS AND METHODS

We addressed our first two aims by collecting additional training data and adding them to our dataset. Once the data were collected, we (i) used the supplemented dataset to develop our new algorithm using the techniques mentioned below, and then (ii) conducted data analyses to determine the model performance.

Data collection

Stage 1

This stage involved adding images of Cocktails, Whiskey, Cider, Cognac, and Brandy to our existing dataset. To achieve this, we followed the same data collection procedure as in ABIDLA (Kuntsche et al., 2020), i.e., scraping images from Google image search with a combination of two sets of keywords for each alcoholic beverage category: (i) drinking, sipping and (ii) relax, friends, social, celebrating, party, watching, movie, home, couch, pub, bar, restaurant, outdoors, park, cheerful, enjoy, depressed, sad, contextual, advertisement, marketing. One example search term is "cocktails drinking bar". In total, we downloaded 69,730 images of Cocktails, Whiskey, Cider, Cognac, and Brandy; these images were annotated by multiple annotators using the same annotation application developed for ABIDLA (Kuntsche et al., 2020). During the annotation process, annotators were provided a set of instructions on how to annotate the images using the annotation application and rules by expert on assigning an appropriate label in case of images with multiple beverages. The annotation rules are placed to maintain consistency across annotators labelling; for example, we included a rule that whenever an image contains multiple beverage categories such as two "Beer Cups" and a "Wine Bottle", the most prominent or easily identifiable beverage category should be annotated as label for that image. Annotation rules

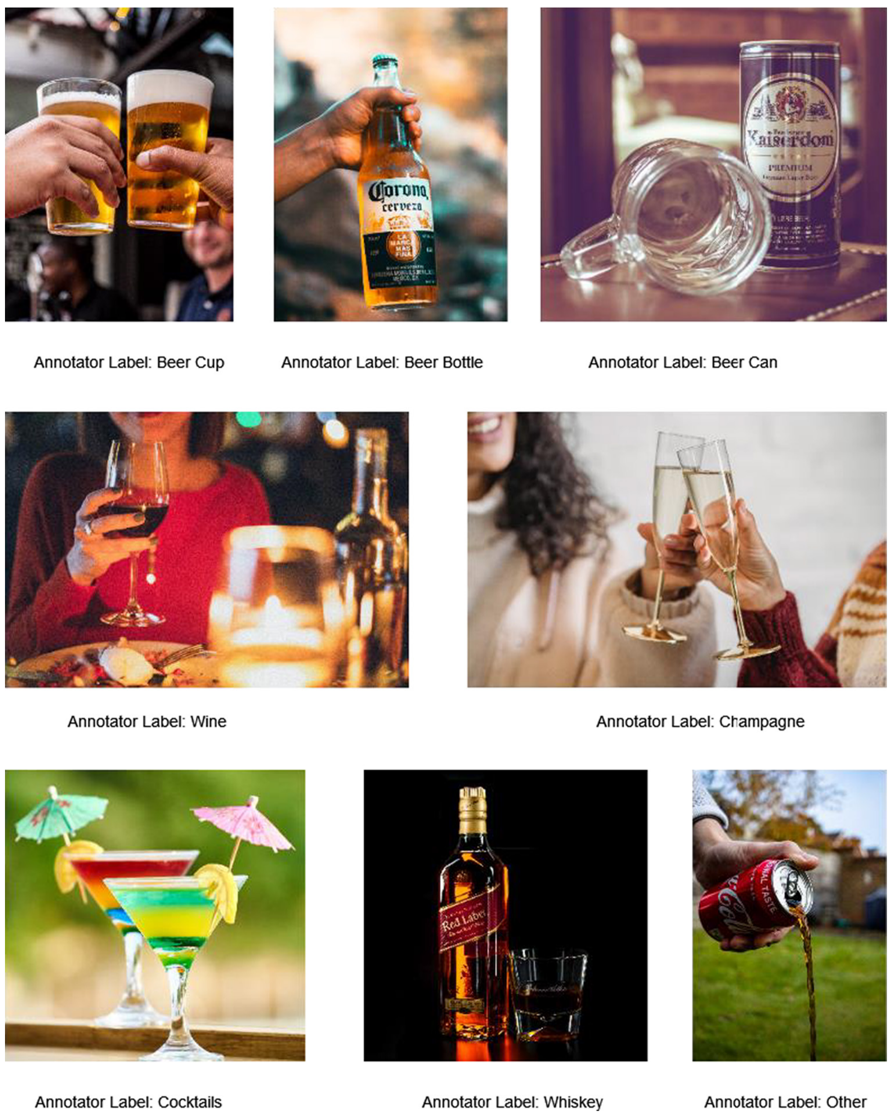

FIGURE 1 Shows example images of each category along with their annotations. All the images shown in this figure are downloaded from Pexels.com. These images are free to use and do not violate any copyright agreements (https://www.pexels.com/license/).

such as the one mentioned in the above example resulted in an inter-rater reliability of 0.903 between the annotations of the annotators and those of an expert (alcohol researcher with more than 20 years of experience) (Kuntsche et al., 2020). By the end of Stage 1 our dataset contained a total of 136,916 images (including the images already annotated for the original ABIDLA algorithm). Figure 1 shows example images of each category along with their annotations.

Stage 2

In Stage 2, we scraped 13,507 images from Bing Image Search to increase the variability of images in the dataset by using different sources. In addition to them, to avoid over-reliance on the contextual features of the image rather than the beverage itself (Norman et al., 2021) we downloaded 88,174 google images of "stand-alone" beverages (i.e., without background or context) using the keywords: "beer bottle", "wine", "champagne", etc. This was done to ensure that the algorithm learns more about the specific beverages of interest without the extraneous "noise" of surrounding imagery, and to improve the robustness of the algorithm. Having a proportion of alcoholic beverage images, collected from the above stages, substantially greater than the non-alcoholic beverage images can introduce a bias towards the alcoholic beverages in the algorithm predictions. To minimize this bias, we downloaded images from Bing and Google using the search terms: coke, water, and water bottle. Out of the downloaded 101,681 images, none of them were annotated manually by annotators because during the development of our algorithm we implemented a method called self-training (as described in the Self-training section) using weakly labeled images. The term "weakly labelled images" refers to images that are assigned labels based on the search terms (such as beer bottle and wine) used to find the images. The accuracy of these weak labels is not verified in anyway by humans. Important note: During our data collection, only publicly available images were scraped from the search engines: Google and Bing. Scraping the public data for research purposes does not violate the terms and policies of these search engines. None of the scraped images were published in this manuscript.

Dataset preparation

Initially, we split the total number of images into "training" (for training the model), "validation" (for finetuning the parameters of the model), and "testing" (for evaluating the model) datasets in a ratio of 70%, 10% and 20%, respectively. To strictly avoid any kind of bias towards a specific class of beverages during the testing, we maintained a uniform distribution in the testing dataset by maintaining 1762 images per beverage class and added the remaining images (resulting from the 20% split of the testing dataset) to training dataset. The class distribution and train, validation and test splits of the final dataset are shown in Table 1.

TABLE 1 Number of images in the training, validation, and testing dataset

Beverage class	Training dataset	Validation dataset	Testing dataset
Beer/Cider Cup	11,485	1016	1762
Beer/Cider Bottle	16,591	1138	1762
Beer/Cider Can	7721	535	1762
Wine	16,583	1279	1762
Champagne	10,207	606	1762
Cocktails	14,736	1006	1762
Whiskey/Cognac/Brandy	37,543	2117	1762
Other images	49,805	4822	1762
Total	164,671	12,519	14,096

Note: To maintain uniform distribution in the testing dataset, we kept 1762 images per beverage class and added the rest of the images to the training data.

In general, the annotated labels are used to train the models and verify the correctness of the predicted labels when testing the algorithm. During training we provide the model with images and their corresponding beverage category as labels. The models then learn to associate the input images with their corresponding labels. During testing, the model takes an input image and predicts the corresponding label. We then verify the correctness of the prediction by comparing the predicted label with the human annotated label. It is important to note that after the model is trained it can be used to predict the labels of a given image without knowing the corresponding label beforehand.

Development of the deep learning algorithm

In ABIDLA (Kuntsche et al., 2020), we trained three separate models, one for each task (classifying beverage images into six, three, and two categories). ABIDLA2 has been trained for two tasks: classifying an image into one of the eight categories (Beer/Cider Cup, Beer/Cider Bottle, Beer/Cider Can, Wine, Champagne, Cocktails, Whiskey/Cognac/Brandy, Other image; Task-1) and classifying an image into one of the two categories (Alcoholic beverage vs. Other image; Task-2). In ABIDLA2, instead of training multiple models for each task, we trained only one model for all the eight beverage categories and then used its learned parameters to make predictions on other tasks. This yielded approximately the same results as training a separate model for each task but with much less computation. The development of our algorithm was done entirely using PyTorch deep learning framework–an open-source library that contains many of the state-of-the-art deep learning algorithms that everyone can use to develop algorithms for customized applications.

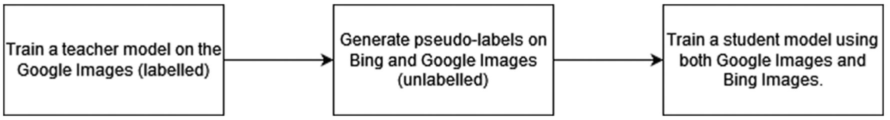

FIGURE 2 An illustration of our self-training approach. An illustration of our Self-training approach. Instead of using the entire unlabelled dataset, we used only the images with pseudo-labels that matches the original hashtag with a confidence higher than 70%.

Baseline model

We trained our Baseline algorithm on the dataset collected at the end of Stage 1. As for ABIDLA (Kuntsche et al., 2020), we chose Densely Connected Convolutional Network (Densenet-121) (Huang et al., 2017) architecture, a model architecture that contains 121 learnable layers, as our default model architecture. We finetuned a pretrained Densenet-121 model, which was initially trained on ImageNet dataset (Russakovsky et al., 2015), for 30 epochs using the Stochastic Gradient Descent Momentum Optimiser (momentum = 0.99) (Loizou & Richtárik, 2020) with an initial learning rate of 0.0001. We used a scheduler to decay the learning rate by a factor of 0.1 after every 10 epochs. An epoch is a complete forward pass through the entire dataset. Based on the empirical observations on validation loss and accuracy, we implemented early stopping by limiting the model training to 30 epochs. As overtraining the model with greater number of epochs may lead to overfitting, a scenario in which the model achieves higher accuracy in the training data (by memorizing the features of the training data) but fails to generalize well on the validation and test data. Along with early stopping, random horizontal flipping (flipping the images horizontally) and random vertical flipping (flipping the images vertically) data augmentation techniques were implemented to avoid overfitting.

Balanced sampler

Balanced sampler is a method we implemented during training that drew an approximately equal number of images per beverage category from the training dataset to maintain uniform distribution across beverage categories. In general, data collected in the real-world is imbalanced, leading to skewed class distribution making them major classes (a greater proportion of images) and minor classes (a smaller proportion of images). For example, as shown in Table 1 in our dataset, “Beer/Cider Can” (4.7%) and “Champagne” (6.2%) are the minor classes whereas “Other images” (30.2%) is the major class. This non-uniform distribution effects the performance of the model by introducing bias towards the major class (Kaur et al., 2019). During training, we implemented a balanced sampler strategy primarily to address the above issue of class imbalance in the training dataset. Our implementation of a balanced sampler involved assigning weights to each image in the dataset. The reason is to assign more weight to the minor class images and less weight to the major class images. We first computed the distribution of the entire dataset and computed the weights for each class of images and assigned these weights to every image in the entire training dataset using a weighted random sampler (https://pytorch.org/docs/stable/data.html#torch.utils.data.DataLoader). This weighted random sampler uses multinomial distribution for sampling images based on the weights resulting in oversampling the minor class by repeating and under-sampling the major classes based on the requirements.

Self-training

Self-training method involves the following three steps: (i) train a teacher model on the labeled dataset, (ii) use the trained teacher model to generate predictions (pseudo labels) on unlabelled images (of a different dataset), and finally, (iii) train a student model using a combination of labeled and pseudo labeled dataset (Xie et al., 2020). In our case, as illustrated in Figure 2, we initially trained a teacher model on the Google image dataset (i.e., dataset collected at the end of Stage 1, which was manually labeled by annotators) and used the teacher model to generate pseudo labels on Bing and Google Images (i.e., the dataset collected at the end of Stage 2, which is unlabelled data).

Analyses

We conducted an ablation study (Meyes et al., 2019) to determine how much each of the methods above contributed to the improvement of the algorithm's performance. In simple terms, an ablation study is an analysis that is performed by removing one method at a time from the algorithm while retaining the other methods and computing the performance of the algorithm; therefore, by conducting this study, we determined the contribution of each method to the algorithm performance (Meyes et al., 2019). We determined the performance of our algorithm by computing the classification accuracy that is the percentage of the total number of correct predictions made by the model divided by the total number of images in the testing dataset. In addition to computing the overall accuracy of the model for both tasks, we also computed the classification accuracy for each individual beverage category. We computed a confusion matrix of our algorithm on the test data to determine among which categories our algorithm gets confused and makes inaccurate predictions. We also computed a Kappa coefficient to determine the level of agreement between the human annotation and algorithm predictions—0.41 to 0.60 moderate; 0.61 to 0.80 substantial; and 0.81 to 1.0 almost perfect agreement (Landis & Koch, 1977). ABIDLA2 is publicly available free of charge. Appendix S1 contains the location where the algorithm is stored together with a

TABLE 2 Accuracy of the different ABIDLA2 versions in identifying images that contain alcoholic beverages

Algorithm	Validation accuracy (in %)	Test accuracy (Task-1) (in %)	Test accuracy (Task-2) (in %)
ABIDLA2	81.7	**77.0**	**87.7**
ABIDLA2 without self training	80.7	76.2 (−0.8)	87.5 (−0.2)
ABIDLA2 without balanced sampler	81.9	76.1 (−0.9)	86.7 (−1.00)
Baseline (without self-training and balanced sampler)	81.8	75.2 (−1.8)	85.6 (−2.1)
Original ABIDLA	—	73.8[a] (−3.2)	85.2 (−2.5)

Note: The values in the parenthesis represents the amount of contribution, in terms of accuracy (in %), by each technique or alternatively, values represent the difference between the accuracies of the best ABIDLA2 model and model without one technique at a time. Task-1 represents classifying eight beverage types. The bold values represents the accuracy values of the best algorithm.

[a]ABIDLA was trained to classify six alcoholic beverages which is a simpler task than ABIDLA2 Task-1 which is to classify eight beverages.

TABLE 3 Represents the comparison of performance between ABIDLA2 and ABIDLA for both Task-1 and Task-2

Categories/Algorithm	ABIDLA2	ABIDLA[a]
Task-1		
Overall accuracy	**77.0**	73.8
Beer/Cider Cup	77.4	73.6
Beer/Cider Bottle	78.3	50.5
Beer/Cider Can	80.5	77.4
Wine	77.8	78.9
Champagne	64.5	66.5
Cocktails	72.1	—
Whiskey/Cognac/Brandy	88.1	—
Other images	77.1	95.5
Task-2		
Overall accuracy	**87.7**	85.2
Alcoholic beverages	89.2	83.6
Other images	77.1	86.8

Note: The values represent the accuracy (in %) of algorithms for each category. ABIDLA2 has a kappa coefficient of 73.7% for Task-1. The bold values represents the accuracy values of the best algorithm.

[a]Values taken from Kuntsche et al. (2020).

step-by-step description about how researchers can apply the algorithm on their own datasets to perform analyses such as quantifying images of alcoholic beverages in any kind of electronic image.

RESULTS

ABIDLA2 with balanced sampler and self-training has achieved an accuracy of 77.0% and 87.7% for Task-1 and Task-2, respectively. We found that the balanced sampler method and self-training methods individually contributed to the improvement of the accuracy of ABIDLA2 by 0.9% and 0.8% respectively. However, when both balanced sampler and self-training methods are implemented simultaneously the resulting ABIDLA2 produced an accuracy of 1.8% higher than the baseline model as shown in Table 2. Our Baseline model has achieved an accuracy of 75.2% and 85.6% for Task-1 (classifying an image into one of the eight categories) and Task-2 (classifying an image into one of the two categories), respectively. Despite the more difficult task of identifying a greater number of beverage categories, ABIDLA2 also outperforms the original ABIDLA (Kuntsche et al., 2020) by 3.2% and 2.5% for Task-1 and Task-2, respectively.

As shown in Table 3, in Task-1, ABIDLA2 was most accurate in identifying Whiskey/Cognac/Brandy (88.1%) followed by Beer/Cider Can (80.5%), Beer/Cider Bottle (78.3%), Wine (77.8%), in this order. In Task-2, ABIDLA2 classified the alcoholic beverages with an accuracy of approximately 90%. The ABIDLA2 algorithm was 1.6 times more accurate than ABIDLA in identifying beer bottles. We found that ABIDLA2 was least accurate in classifying images of Champagne (64.5%) relative to the other beverage categories; however, this accuracy is more than five times higher than random chance, which is 12.5% (= 1/8 categories).

To better interpret ABIDLA2's performance, we computed a confusion matrix, which is a two-dimensional table that represents number of true positives, true negatives, false positives, and false negatives in each beverage category, to investigate where the algorithm made mistakes (Figure 3). Out of 1762 images per beverage in the testing dataset, we found that ABIDLA2 misclassified: 58 (3.3%) wine images as champagne and 182 (10.3%) as Others; 223 (12.7%) Champagne images as Wine; 305 (17.3%) Cocktail images as Others and 122 (6.9%) Other images as Cocktail images. ABIDLA2 has a kappa coefficient of 73.7% for Task-1.

DISCUSSION

In this article, we developed a revised version of our previous algorithm ABIDLA (Kuntsche et al., 2020) to identify 12 popular types of alcoholic beverages—Beer Cup, Beer Bottle, Beer Can, Wine Champagne, Cocktails, Whiskey, Cider Cup, Cider Can, Cider Bottle, Cognac, Brandy—in images with higher identification accuracy. Such an improvement was critical to efficiently recognize the most common alcoholic beverage images in a host of electronic media. In Task 2—classifying images into alcoholic beverages and other images—our results indicated that ABIDLA2 with an accuracy of 87.7%

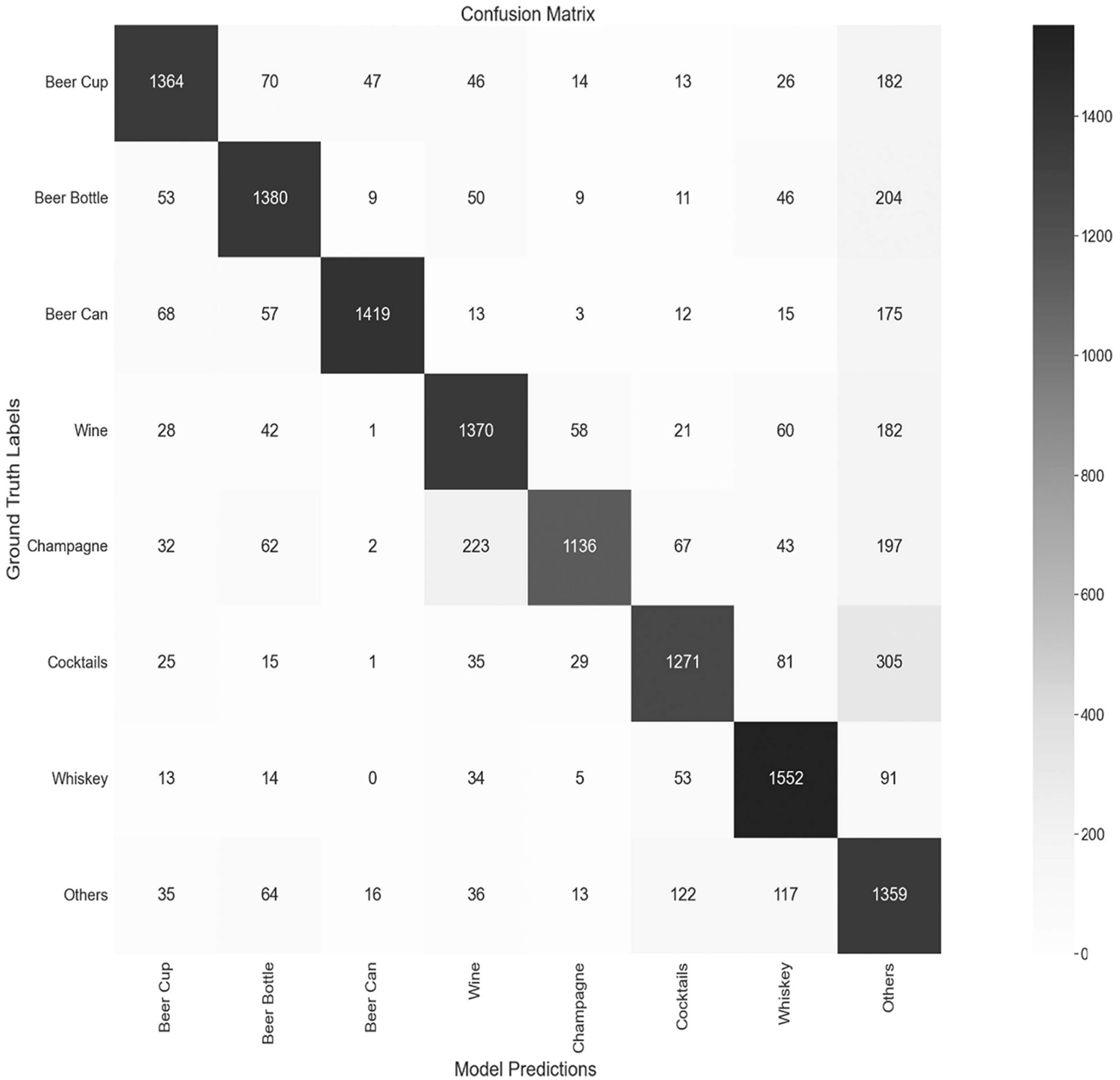

FIGURE 3 Represents the confusion matrix of our best ABIDLA-2 model when evaluated on the test data. The X-axis represents model predictions and y-axis represents ground truth (actual) labels—labels that were manually assigned by annotators. The diagnal values in the matrix represents the number of images in the test data that our algorithm correctly predicted as the ground truth labels and all other values in the matrix represents number of incorrect predictions. For simplicity we presented Whiskey/Cognac/Brandy as Whiskey and Beer/Cider Cup as Beer Cup etc.

outperformed the already high accuracy of ABIDLA (Kuntsche et al., 2020) of 85.2% by a further 2.5%. Our ablation study indicated that implementation of the balanced sampler strategy to address the skewness of the dataset, and self-training method to make use of large secondary weakly labeled dataset, contributed particularly to the improvement of ABIDLA2 performance.

Implementation of the balanced sampler strategy produced more uniform accuracies and thus helped to improve the accuracy of beverage categories previously underperforming in ABIDLA (Kuntsche et al., 2020) such as beer bottles. Implementation of the self-training method helped us discover that our algorithm can learn the essential features of beverages and improve its performance by itself, even from a weakly labeled dataset (i.e., a dataset that is not manually annotated by a human). One of the main benefits is that self-training considerably reduces annotation time (i.e., preparation for model training) when including new datasets. For instance, if we had annotated the unlabelled (Bing and Google images) dataset, it would have taken several weeks for multiple annotators to label the entire 101,681 images.

ABIDLA2 is most accurate in predicting images of "Whiskey/Cognac/Brandy" with an accuracy of 88.1% compared with other alcoholic beverages. This could be due to the clearly distinguishable features of this class of beverages—such as shape of the glasses and bottles, color of the beverage, etc. Compared with its previous version, ABIDLA2 classifies images of "Beer/Cider" and "Wine" more uniformly with an average accuracy of 77.3%. However, the images of "Champagne" had an accuracy (64.5%) that was 12.5% below its overall accuracy (77.0%) and 2% less than ABIDLA's champagne accuracy (66.5%). This decrease in accuracy in some beverage categories could be because ABIDLA2 had to learn to distinguish the features of beverages among a greater number of beverage categories compared with ABIDLA. The Confusion matrix (Figure 3) indicated that ABIDLA2 misclassified 223 (12.7%) images of "Champagne" as "Wine" but only 58 (3.3%) of "Wine" images as "Champagne". This is most likely due to champagne sharing features with white wine, such as the color, but wine in general (including red wine) do not necessarily share the features of champagne.

The highest confusion 17.3% (305/1762) was for images annotated to contain cocktails that ABIDLA2 misclassified as 'other', which may be due to the colorful and 'creative' appearance of these beverages. Among alcoholic beverages, 4.5% (81/1762) of the annotated cocktail images were misclassified by ABIDLA2 as containing whiskey/cognac/brandy, which can be the case when the cocktails are based on these spirits or share similar visual features. The misclassification of all other alcoholic beverage combinations was below 5%. The Kappa coefficient between the algorithm predictions and the annotator labels indicated that there is a substantial agreement between the algorithm predictions and annotators labels (Landis & Koch, 1977).

Strengths, limitations, and future work

One of the strengths of ABIDLA2 is that it can identify wider range alcoholic beverage categories in images than the previous version (Kuntsche et al., 2020). ABIDLA2 has the potential to learn from a weakly labeled dataset from a different source/domain with very minimal annotations, reducing the burden of coding. ABIDLA2 can potentially screen any kind of electronic media to quantify alcohol exposure in images and serve as the basis to create algorithms to quantify alcohol exposure in video data (e.g., movies, tv shows, streaming videos, etc.). In terms of application, there is the possibility to integrate ABIDLA2 into a web browser (as a plugin) or a mobile application to quantify and/or limit exposure to alcohol exposure depending on individuals' preferences. For example, people in alcohol-related rehabilitation or recovery may want to limit the alcohol exposure to reduce craving to minimum, whereas others may want to decrease the amount of alcohol advertisements. Such a plugin or application may also be useful for parents who wish to restrict alcohol-related content for their children when browsing online.

One main limitation of ABIDLA2 is the below average accuracy of "Champagne" and "Cocktail" images. Another limitation of ABIDLA2 is that it is trained to identify the most prominent alcoholic beverage in an image. Future work should aim to address this issue by developing an object detection algorithm which can identify the location of the beverages in the image. This can be achieved by clearly specifying the exact location of the beverages in the image by drawing bounding boxes around them and labelling them appropriately so that the algorithm will know exactly what to look for in any given new image. In addition to identifying the most prominent alcoholic beverage in an image, an object detection algorithm can also provide additional information such as the number of alcoholic beverages in the images, and how prominent the beverage is shown in the image. An alternative to annotating each beverage with a bounding box, we could pose the problem as a multi-label classification problem—i.e., predicting multiple beverage categories present in each image. Therefore, predicting the presence or absence of each beverage category independently. For example, label an image as containing both beer cups and wine. This approach imposes a smaller manual annotation burden compared with annotating bounding boxes while still giving the model the information that multiple beverage types exist simultaneously in an image.

Finally, our work did not include evaluation of our algorithm performance on images from an entirely new source/domain, although we included images from different sources during model training. Future work should test the robustness of the algorithm to identify alcoholic beverages in a completely new dataset that has been independently sourced (e.g., multiple social media platforms).

CONCLUSIONS

We have developed a more efficient and refined deep learning algorithm, ABIDLA2, which extends the capabilities of its predecessor by identifying 12 popular types of alcoholic beverages and can learn from a weakly labeled dataset to reduce the burden of coding. ABIDLA2 can rapidly screen any kind of electronic media—including Facebook, Instagram, Twitter, YouTube, and Netflix (by breaking the video material down into individual frames and treating each frame as a separate input image)—to provide an estimation of quantity of alcohol exposure through images. Algorithms like ABIDLA2 are important because of the enormous and still increasing popularity of electronic media, and the fact that exposure to alcohol images in such media has been demonstrated to be associated with drinking initiation among non-drinkers (Anderson et al., 2009; Smith & Foxcroft, 2009), increased levels of consumption among drinkers (Anderson et al., 2009), and binge or hazardous drinking behavior (Noel et al., 2020). ABIDLA2 is publicly available free of charge and can be applied to future scientific research or public health initiatives.

ACKNOWLEDGMENTS

We would like to thank Sarah Kuntsche, Zoë Kuntsche, and Geoffrey Leggat for their contribution to data annotation. We also would like to thank Prof. Ludwig Kraus for providing comments on the previous iteration of this article. Open access publishing facilitated by La

Trobe University, as part of the Wiley - La Trobe University agreement via the Council of Australian University Librarians. Open access publishing facilitated by La Trobe University, as part of the Wiley - La Trobe University agreement via the Council of Australian University Librarians.

CONFLICT OF INTEREST

There is no conflict of interest to declare.

ORCID

Abraham Albert Bonela https://orcid.org/0000-0002-8178-1640
Thomas Norman https://orcid.org/0000-0002-4773-4285
Emmanuel Kuntsche https://orcid.org/0000-0001-8931-0790

REFERENCES

Alhabash, S., Mundel, J., Deng, T., Mcalister, A., Quilliam, E.T., Richards, J.I. et al. (2021) Social media alcohol advertising among underage minors: effects of models' age. *International Journal of Advertising*, 40, 552–581.

Anderson, P., de Bruijn, A., Angus, K., Gordon, R. & Hastings, G. (2009) Impact of alcohol advertising and media exposure on adolescent alcohol use: a systematic review of longitudinal studies. *Alcohol and Alcoholism*, 44, 229–243.

Feehan, B. (2021) *2021 Social media industry benchmark report* [Online]. RivalIQ. Available from: https://www.rivaliq.com/blog/social-media-industry-benchmark-report/#title-alcohol [Accessed 3rd May 2021].

Huang, G., Liu, Z., van der Maaten, L. & Weinberger, K.Q. (2017) Densely connected convolutional networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708.

Kaur, H., Pannu, H.S. & Malhi, A.K. (2019) A systematic review on imbalanced data challenges in machine learning: applications and solutions. *ACM Computing Surveys (CSUR)*, 52, 1–36.

Kemp, S. (2022) *Digital 2022: global overview report* [Online]. Available from: https://datareportal.com/reports/digital-2022-global-overview-report [Accessed 15th February 2022].

Kuntsche, E., Bonela, A.A., Caluzzi, G., Miller, M. & He, Z. (2020) How much are we exposed to alcohol in electronic media? Development of the Alcoholic Beverage Identification Deep Learning Algorithm (ABIDLA). *Drug and Alcohol Dependence*, 208, 107841.

Landis, J.R. & Koch, G.G. (1977) The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174.

Lauckner, C., Desrosiers, A., Muilenburg, J., Killanin, A., Genter, E. & Kershaw, T. (2019) Social media photos of substance use and their relationship to attitudes and behaviors among ethnic and racial minority emerging adult men living in low-income areas. *Journal of Adolescence*, 77, 152–162.

Loizou, N. & Richtárik, P. (2020) Momentum and stochastic momentum for stochastic gradient, Newton, proximal point and subspace descent methods. *Computational Optimization and Applications*, 77, 653–710.

McCreanor, T., Lyons, A., Griffin, C., Goodwin, I., Moewaka Barnes, H. & Hutton, F. (2013) Youth drinking cultures, social networking and alcohol marketing: implications for public health. *Critical Public Health*, 23, 110–120.

Meyes, R., Lu, M., de Puiseau, C.W. & Meisen, T. (2019) Ablation studies in artificial neural networks. *arXiv preprint arXiv:1901.08644*.

Noel, J.K., Sammartino, C.J. & Rosenthal, S.R. (2020) Exposure to digital alcohol marketing and alcohol use: a systematic review. *Journal of Studies on Alcohol and Drugs, Supplement*, s19, 57–67.

Norman, T., Bonela, A., He, Z., Angus, D., Carah, N. & Kuntshe, E. (2021) Connected and consuming: adapting a deep learning algorithm to quantify alcoholic beverage exposure in user-generated Instagram images. *Drugs: Education, Prevention and Policy (IDEP)*, 1–8.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S. et al. (2015) Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115, 211–252.

Schoenmakers, T., Wiers, R.W. & Field, M. (2008) Effects of a low dose of alcohol on cognitive biases and craving in heavy drinkers. *Psychopharmacology*, 197, 169–178.

Smith, L.A. & Foxcroft, D.R. (2009) The effect of alcohol advertising, marketing and portrayal on drinking behaviour in young people: systematic review of prospective cohort studies. *BMC Public Health*, 9, 1–11.

Xie, Q., Luong, M.-T., Hovy, E. & Le, Q.V. (2020) Self-training with noisy student improves imagenet classification. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10687–10698.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Bonela, A.A., He, Z., Norman, T. & Kuntsche, E. (2022) Development and validation of the Alcoholic Beverage Identification Deep Learning Algorithm version 2 for quantifying alcohol exposure in electronic images. *Alcoholism: Clinical and Experimental Research*, 46, 1837–1845. Available from: https://doi.org/10.1111/acer.14925