

# 420

The first CNN for cannabis category recognition

---



420.

---

---

# Introduction

## Why this project ?

In France, cannabis isn't legal yet, which means that **you aren't able to know the quality or the category of the cannabis you're buying**. I used to smoke a lot, and I've always been angry about the fact that I wasn't able to know in advance what will be the effects of the cannabis I was smoking. Even though I don't smoke anymore, this problem is still annoying me.

## What do you mean by cannabis category ?



There are **3 main cannabis categories** : Indica, Sativa and Hybrid. Each of these category has different effect on the body. As found on [Leafly](#) :

- **Indicas** strains are believed to be physically sedating, perfect for relaxing with a movie or as a nightcap before bed.
- **Sativas** tend to provide more invigorating, uplifting cerebral effects that pair well with physical activity, social gatherings, and creative projects.

- 
- **Hybrids** are thought to fall somewhere in between the indica-sativa spectrum, depending on the traits they inherit from their parent strains.

You understood... Let's say it's sunday night, and you want to rest to be in shape for the coming week, so you smoke some cannabis. Sadly, you bought some sativa without knowing it. Resting will be complicated.

### How to solve the problem ?

Some firm have already address this problem by creating high-tech cannabis analyzer.



But the price of these analyzer is way to high for the common mortal, the prices range from \$15,000 to \$25,000. Even if these device are very accurate (thanks god, in view of the price), nobody will pay these prices to gain a better understanding over the cannabis they're buying.

### What if a simple picture was enough ?

Imagine taking a picture of your cannabis and getting back an analysis of it, for free. With today technologies, it isn't a dream anymore. That's why I've built the first CNN for cannabis category recognition. My CNN, called 420, is able to take an image of cannabis as input and gives back its category (sativa / indica / hybrid) as output. Welcome to deep learning.

---

## Dataset

### Where did you find it ?

I found all of the images directly on [Leafly](#), this website is incredible ! There are a lot of public resources on this website such as informations, images or reviews about cannabis. I've gathered every cannabis images URL in a CSV file called ***images-url.csv***.

At the end I've gathered more than 21,000 images of cannabis. More than 8,000 of hybrid, 8,000 of indica and 5,000 of sativa.

I've downloaded all of these images using the Requests Python library. The code can be found in the ***data-extraction.ipynb*** notebook.

Once I've downloaded all the images I've reorganized them by training, validation and testing set. The code can be found in the ***data-segmentation.ipynb*** notebook.

## CNN Architecture

### Which architecture did you chose ?

In order to get the best results I decided to use transfer learning. I've tried all the greatest architecture including Xception, VGG16, VGG19, ResNet50 and InceptionV3

I've frozen all of the layers except the last one in order to preserve the knowledge gained on imagenet.

```
for layer in model.layers[:-1]:  
    layer.trainable = False
```

---

I've then added 2 dense layer of 1024 nodes and a Dropout layer between them, in order to don't get the weights stuck in a certain part of the network. I've then added an output layer with a softmax function in order to get an estimation of the cannabis category. This way we will be able to tell our users how much they can trust our predictions.

```
x = model.output
x = Flatten()(x)
x = Dense(1024, activation="relu")(x)
x = Dropout(0.5)(x)
x = Dense(1024, activation="relu")(x)
predictions = Dense(3, activation="softmax")(x)
```

I've then used the Stochastic Gradient Descent as the optimizer of my model and I used a small learning rate (0.0001).

```
model_final.compile(
    loss = "categorical_crossentropy",
    optimizer = optimizers.SGD(lr=0.0001, momentum=0.9),
    metrics=["accuracy"]
)
```

## CNN Training

To train my model I first decided to use data augmentation. Data augmentation helps our algorithms to be more adaptive to new images in the future. By rotating, distorting, or rescaling our images, the algorithms gain in invariance.

```
train_datagen = ImageDataGenerator(
    rescale = 1./255,
    horizontal_flip = True,
    fill_mode = "nearest",
    zoom_range = 0.3,
    width_shift_range = 0.3,
    height_shift_range=0.3,
    rotation_range=30
)

test_datagen = ImageDataGenerator(
    rescale = 1./255,
    horizontal_flip = True,
    fill_mode = "nearest",
    zoom_range = 0.3,
    width_shift_range = 0.3,
    height_shift_range=0.3,
    rotation_range=30
)
```

I then used these generator in order to generate more images from my existing images.

```
train_generator = train_datagen.flow_from_directory(
    'datasets/train',
    target_size = (img_height, img_width),
    batch_size = batch_size,
    class_mode = "categorical"
)

validation_generator = test_datagen.flow_from_directory(
    'datasets/valid',
    target_size = (img_height, img_width),
    class_mode = "categorical"
)
```

---

I started first with 50 epochs and a batch size of 421 images. I used that to benchmark the different architecture. I found that ResNet50 was the best architecture for my model.

```
Epoch 1/50  
366/421 [=====>....] - ETA: 48s - loss: 1.1595 - acc: 0.3569
```

Despite all my efforts and different architecture my accuracy was not excellent.

## CNN Results

The accuracy of my model is only of 41%, at this point I think that the problem comes from the dataset. The image are very hard to labeled because it's very hard to recognize cannabis category even for humans.

So it made the process of assessing the quality of the images very hard.

My first fear with this project was to find correct labeled images. Even if leafly gives us access to photos we can't be sure that they are correctly labeled.

## Conclusion

I'm very happy of this project. I hope you'll find it cool too. Even if the conclusion is not very enthusiastic I hope it gave you the passion to go forward in computer vision. Even if the problems are not always solvable, it worth trying.

To conclude we saw the differences in cannabis category, define the problem and formulate an answer to it. We then tried to answer the problem with images of cannabis extracted from the web but the results were not satisfying. In the future I'll try to see first if a human can do it before asking an algorithm to do it.

---

## References

- <https://www.leafly.com/>
- <https://keras.io/applications/>
- <https://medium.com/@14prakash/transfer-learning-using-keras-d804b2e04ef8>