# Machine Learning Prediction of Political Orientation using Portraits of US Congress Members

JONAH EDMUNDSON* and MADISON GREENOUGH*, University of British Columbia Okanagan, Canada

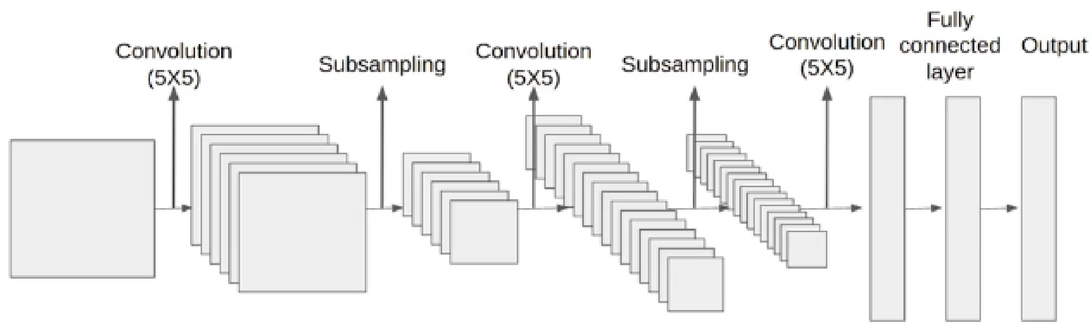


Fig. 1. LeNet-5 Architecture.

## ABSTRACT

Images of current and former U.S. Congress Members were gathered for the dataset to represent members of the Republican and Democrat parties. This report aims to answer the question - can a neural network model predict the political affiliation of an individual given simply their image? This leads to several other potential questions, as to whether political affiliation can be determined based on simply physical features, regardless of if a human can detect it or not. Michal Kosinski's 2021 paper inspired this research, as he conducted a large study and found that political orientation was classified in 72% of images. In this study, three different architectures were trained and tested for classifying images. This includes LeNet-5, DenseNet, and a custom architecture modified from LeNet. The LeNet-5 model outperformed the other two significantly, with test classification rates of 60.21%, 51.76%, and 51.41%, respectively. All models performed better than chance, 50%, but the models failed to perform as well as Kosinski's, likely due to dataset limitations and computational power. Future work includes sourcing larger, more representative datasets to obtain a better classification accuracy.

Additional Key Words and Phrases: datasets, neural networks, web-scraping, politics, facial analysis, CNN, classification

*Both authors contributed equally to this research.

Authors' address: Jonah Edmundson; Madison Greenough, University of British Columbia Okanagan, Kelowna, BC, Canada.


Manuscript submitted to ACM

**ACM Reference Format:**
Jonah Edmundson and Madison Greenough. 2023. Machine Learning Prediction of Political Orientation using Portraits of US Congress Members. 1, 1 (April 2023), 9 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

## 1 INTRODUCTION

Machine learning has been used for image classification in many revolutionizing ways, including classifying different species. There has also been moral controversy regarding big tech firms and their biased human image classification modelling. Google made headlines in 2015 for their perceived racist image labeling (Kayser-Bril). Image classification is difficult, especially when the task is to classify something that the naked human eye cannot predict accurately. Over the years, political affiliation has become a growing source of division and controversy. The two main political groups are liberals and conservatives, also known as members of the left and right wing, or in the United States, they are referred to as Democrats and Republicans. These two groups have very distinguishing behavioural features, but can there be distinguishing physical features as well?

In this report, the goal is to minimize all bias as absolutely necessary, to avoid the same pitfalls that Google experienced, while testing the possibility of artificial intelligence being able to predict underlying human behaviours and preferences from a simple image. Inspiration for this project came from this link. This would be a feat more capable than humans themselves. If successful, this model could open up a number of possibilities. Advertising companies could better target ads based on traditional behaviours of people based on their political affiliation. Voting modelling could better predict the outcome of elections by simply using images of citizens. There are numerous possibilities that would generate significant revenue and innovation for societies around the world.

## 2 BACKGROUND

### 2.1 Dataset

The dataset for this project was obtained from the US Congress Members archive. This includes images of 2,515 current and former members of the U.S. Congress spanning from 1973 to 2024, which includes the 93rd to 118th Congress. Representation from all US States were present. Photos of each member as well as their party affiliation (republican or democrat) were recorded. Members with an affiliation of "independent" were removed. After data collection, the rows were shuffled, and the dataset was split into training (80%), validation (5%) and testing (15%). The dataset contained a total of 1,889 rows. Further pre-processing was done and images were resized to 120 by 100 pixels (height by width) and converted to grayscale. Before CNN training, images were sequentially augmented thrice using the IMAGENET AutoAugment policy.

### 2.2 Assumptions

When conducting this analysis, the following assumptions were made in order to continue:

- "Democrats" and "republicans" are good proxies for "liberal" and "conservative", respectively.
- Members of US Congress are representative of democrats and republicans in general.
- Democrats and republicans have different facial attributes.

Fig. 2. Left: A subset of one batch before augmentation (64/128). Right: A subset of one batch after image augmentation (64/128).

These assumptions allowed us to collect the dataset mentioned above. It also allowed us to reasonably gather a dataset that could represent our two parties. Note that we did attempt to gather data of ordinary public members by scraping Facebook profile IDs from new members of 2 Facebook "Groups": republicans from Republicans in Texas, and democrats from Democratic Voices for Biden/Harris 2024. This method could not complete a large enough dataset for this analysis due to Facebook permissions, despite repeated attempts.

## 3 METHODS

### 3.1 Architecture

Three different architectures were trained and tested for classifying the images. These were LeNet-5, DenseNet, and a custom architecture (created from a modified LeNet). The specific architectures, including features and classifiers, of these NNs can be seen below in tables 1 to 3. The DenseNet architecture can be further investigated by visiting this GitHub repo. Specifics regarding performance comparison between the three main methods can be found in the results section.

### 3.2 Tuning

Many attempts were made to tune the each of the models in order to raise the classification accuracy for the test data set. Learning rate, batch size, number of epochs, dropout rate, number of features, weight decay, and depth (found only in DenseNet) were systematically altered individually. Many iterations of each of the

Table 1

| **LeNet-5** | | |
|---|---|---|
| *Features* | | |
| Convolutional Layer | | Kernel Size = 4 |
| Tanh Activation | | |
| Max Pooling Layer | | Kernel Size = 2 |
| Convolutional Layer | | Kernel Size = 14 |
| Tanh Activation | | |
| Max Pooling Layer | | Kernel Size = 4 |
| *Classifier* | | |
| Linear | 1408 In | 120 Out |
| Tanh Activation | | |
| Dropout | | 0.7 |
| Linear | 120 In | 84 Out |
| Tanh Activation | | |
| Dropout | | 0.8 |
| Linear | 84 In | 2 Out |

Table 2

| **DenseNet** | | |
|---|---|---|
| *Features* | | |
| Dense Block | | |
| Transition Block | | |
| Dense Block | | |
| Transition Block | | |
| Dense Block | | |
| Batch Normalization | | |
| *Classifier* | | |
| ReLU Activation | | |
| Linear | 24 In | 2 Out |

three models were run with these various hyperparameter values. The best values for each hyperparameter were discovered for each of the models, and can be seen in table 4. These final values each improved the model accuracy for all three approaches.

## 4 RESULTS

The LeNet-5 model outperformed DenseNet and the custom LeNet by a significant margin, even after extensive tuning. See the comparison in test error between the three models in this plot (fig. 6). LeNet-5 obtained a test classification rate of 60.21%. DenseNet and custom LeNet achieved classification rates of 51.76% and 51.41%, respectively. This shows that LeNet-5 was significantly better than both of the two remaining models by approximately 9%. It is interesting to note that the DenseNet and custom LeNet performed very similarly with respect to their classification rate, despite being very different models. These

Table 3

| **Modified LeNet-5 (Custom)** | | |
|---|---|---|
| *Features* | | |
| Convolutional Layer | | Kernel Size = 3 |
| Sigmoid Activation | | |
| Max Pooling Layer | | Kernel Size = 2 |
| Convolutional Layer | | Kernel Size = 5 |
| Sigmoid Activation | | |
| Max Pooling Layer | | Kernel Size = 2 |
| Convolutional Layer | | Kernel Size = 8 |
| Sigmoid Activation | | |
| Max Pooling Layer | | Kernel Size = 2 |
| *Classifier* | | |
| Linear | 1120 In | 120 Out |
| Tanh Activation | | |
| Dropout | | 0.8 |
| Linear | 120 In | 84 Out |
| Tanh Activation | | |
| Dropout | | 0.6 |
| Linear | 84 In | 2 Out |

Table 4

| **Hyperparameters** | LeNet-5 | DenseNet | Custom |
|---|---|---|---|
| *Learning Rate* | 0.001 | 0.0012 | 0.0015 |
| *Batch Size* | 128 | 128 | 128 |
| *Epochs* | 12 | 15 | 15 |
| *Drop out* | 0.7, 0.8 | 0 | 0.8, 0.6 |
| *Features* | 12,000 | 12,000 | 12,000 |
| *Classes* | 2 | 2 | 2 |
| *Weight Decay* | 0.0002 | 0.0001 | 0.0001 |
| *Depth* | *NA* | 4 | *NA* |

last two models perform only slightly better than random chance, as random chance is given by a 50% classification accuracy, while basic LeNet-5 performs decently well, but there is still questions regarding its reliability. The 60% mark is better than random chance, and possibly better than the naked human eye, but it is not a strong enough result to create a reliable model. Our goal at the beginning was to be able to better predict political affiliation compared to humans, which you could assume is no better than chance, so it is still a possibility that this goal has been reached. The convolutional filters of both convolutional layers for LeNet-5 can be seen in fig. 3.

Interestingly, figs. 4 and 5 show that while all the models were able to converge to a solution on the training data, this convergence did not lead to better classification for the validation set. For the training data, all three models were able to converge to the minimum cost in only two epochs. In contrast, for the validation data, the models just seem to jump around without improving for all 15 epochs.

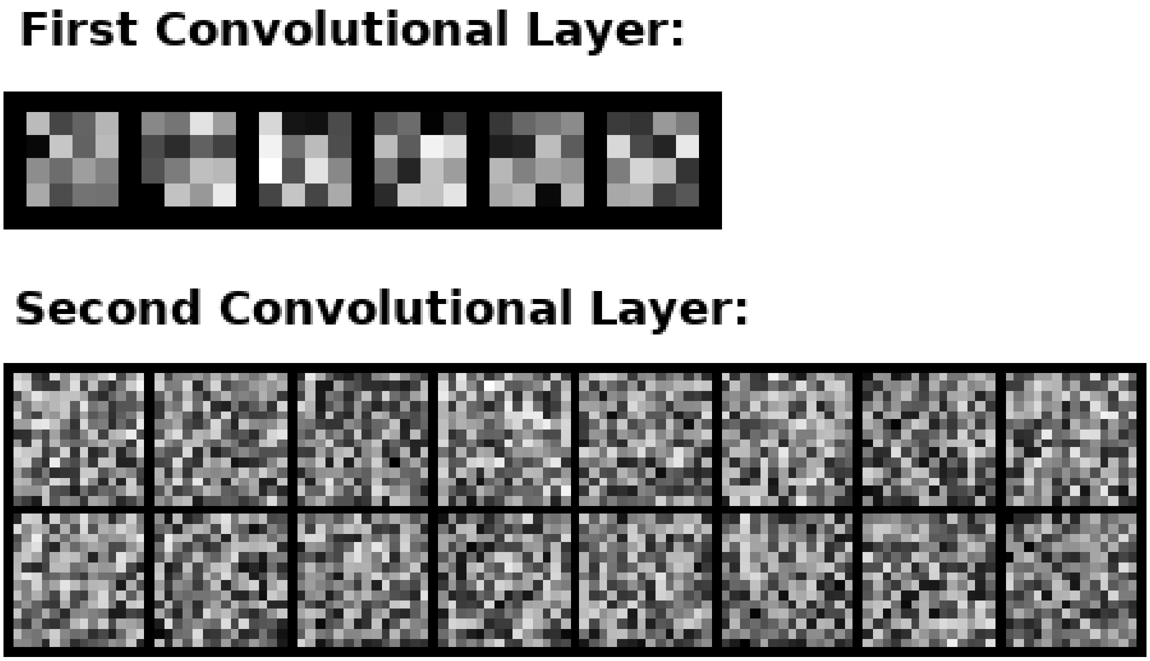


Fig. 3. The convolutional filters of the first and second convolutional layers.

## 5 DISCUSSION

Upon reflection of the performance of the various models tested, it is worth noting that the test classification rate was higher than the train classification rate. Although not impossible, we were quite perplexed as to why this was the case. After careful consideration, it was speculated that this is occurring because the training data is augmented whereas the testing data is not, so the model is able to more easily classify the test set images because of this, which resulted in the higher rate for the test set rather than the train set.

Although the model performed better than chance, and potentially better than an impartial human being could guess, the performance was not at the level that we would have liked. The poor classification rate of the NN may have occurred because of these potential reasons listed below:

(1) There could be no actual difference in facial features between democrats and republicans.
(2) There are facial differences between democrats and republicans, but this dataset is too small to pick up on these subtle facial differences. Or, these specific formal images of the congress members may not represent the facial differences the best
(3) The classifier section of the NN works well, but the convolution section is not accurately picking up the meaningful facial features that are able to distinguish between the two parties.
(4) The convolution section of the NN is able to pick up on some meaningful signals, but the classification section is not performing well.
(5) The NN is only picking up features in the background of the image (and not the facial features), which are not predictive nor meaningful.

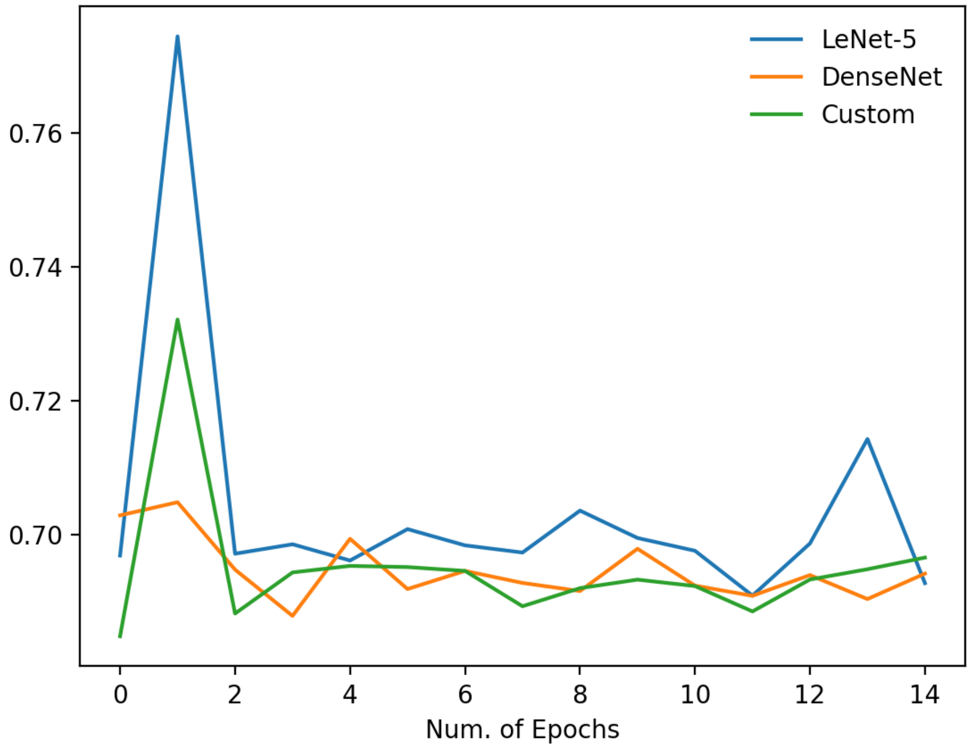


Fig. 4. The training cost over 15 epochs for the 3 architectures considered.

Out of all of these potential reasons, we believe that the fifth option is a more likely situation as to why the model is not performing as well as it could.

## 6 LIMITATIONS & FUTURE WORK

In order to rectify the low classification rate of the NN, as well as to iron out the assumptions listed earlier in this report, a new dataset should be collected. Ideally, the dataset would be collected from public social media profiles such as Facebook or dating websites, as per our original intentions. This new dataset should also be much larger, ideally by several orders of magnitude. Once this dataset is collected, more complex NN architectures could be attempted. If this were to be done, we speculate that the classification rate would increase, as was seen in the Kosinski (2021) paper.

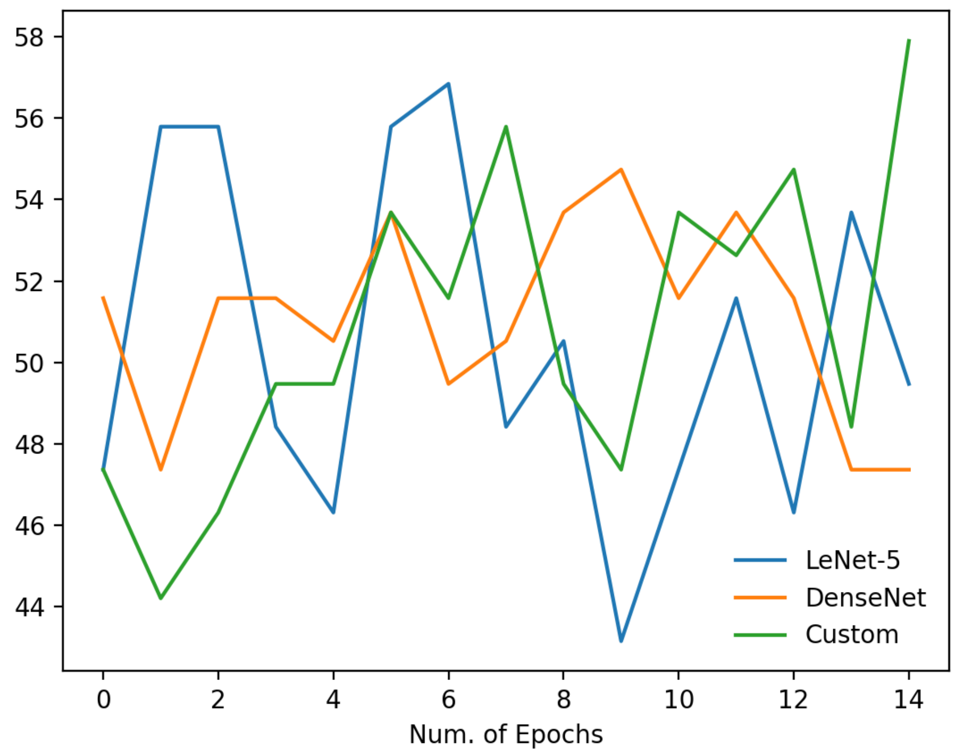


Fig. 5. The validation error over 15 epochs for the 3 architectures considered.

## 7 REFERENCES

Andreasveit. (n.d.). Andreasveit/densenet-pytorch: A pytorch implementation for densely connected Convolutional Networks (DenseNets). GitHub. Retrieved April 21, 2023, from https://github.com/andreasveit/densenet-pytorch

Congress.Gov. (n.d.). Members of the U.S. congress | Congress.gov | Library of Congress. Members of the U.S. Congress. Retrieved April 21, 2023, from https://www.congress.gov/members

Kayser-Bril, N. (n.d.). Image classification algorithms at Apple, Google still push racist tropes. AlgorithmWatch. Retrieved April 20, 2023, from https://algorithmwatch.org/en/apple-google-computer-vision-racist/

Kosinski, M. (2021, January 11). Facial recognition technology can expose political orientation from naturalistic facial images. Nature News. Retrieved April 21, 2023, from https://www.nature.com/articles/s41598-020-79310-1

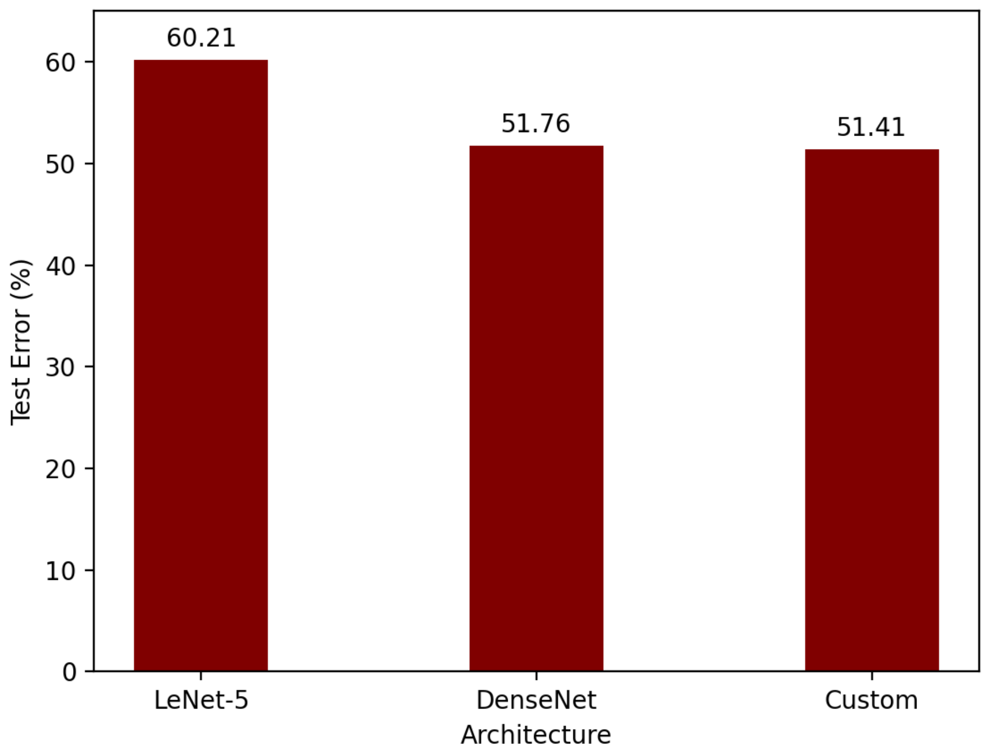


Fig. 6. The final testing error for the 3 architectures considered.