Online Appendix to:

Hidden in Plain Sight? Detecting Electoral Fraud Using Statutory Results*

Zach Warner† J. Andrew Harris‡ Michelle Brown§
Christian Arnold¶

July 13, 2020

*We thank the National Democratic Institute for generously sharing their data. We also thank our research assistants, Henri Böschen, Karunakar Reddy Mannem, and Jeff Morgan, for their contributions. Authors are listed in reverse-alphabetical order.

†Postdoctoral Research Fellow, Cardiff University. email: WarnerZ@cardiff.ac.uk, web: http://www.zachwarner.net.

‡Assistant Professor, New York University, Abu Dhabi. email andy.harris@nyu.edu, web: http://www.jandrewharris.org/.

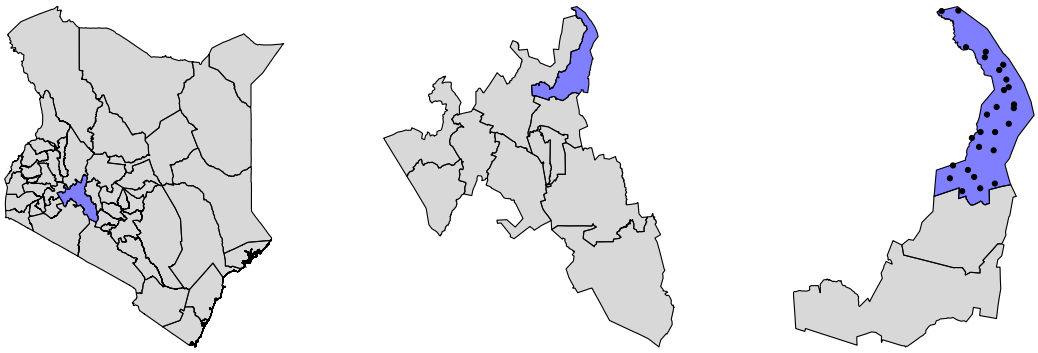
§Senior Advisor, Elections and Political Processes, National Democratic Institute.

¶Lecturer, Cardiff University. email: ArnoldC6@cardiff.ac.uk, web: http://christianarnold.org/.

Contents

A.1 Kenya's internal political boundaries

Kenya's 2010 Constitution reorganized the country's internal political boundaries. With the former provinces eliminated, 47 counties became the new top-level administrative units. Nested within these counties are 290 parliamentary constituencies, which are further subdivided into county assembly wards. Figure A1 visualizes these subdivisions.



(a) Counties **(b)** Constituencies **(c)** Wards

Figure A1: Kenyan territorial divisions. Polling centers—which can contain multiple polling stations (or "streams")—are indicated with dots. From left to right, Nakuru county, Subukia constituency, and Waseges ward are highlighted.

A.2 Details about hyperparameter tuning

A.2.1 Model architectures

We tune across five model architectures: {inception, average, wide, deep, hard}.[1] Note that throughout, we refer to a "block" as the combination of a 3-pixel empty border padding layer (which helps prevent the edges near the border getting ignored during training), a convolutional layer, a batch normalization, an exponential linear unit layer, and a pooling layer (in that order). The details of each model are as follows.

The inception model's base model is the InceptionV3 model (Szegedy et al. 2016), trained on the ImageNet database (Deng et al. 2009). We take the top layer of this model (its output), add a pooling layer, a batch normalization layer, a dense layer with 1,024 exponential linear units (ELUs), and then a 50% dropout layer before the activation layer using a sigmoid function. We train the model such that all layers are trainable, including those in the underlying InceptionV3 model.[2]

The "average" model's feature extraction step consists of three layers: a block in which the convolutional step has 32 filters, another block with 64 filters, and a third block with 128 filters. The prediction step is also grouped into three layers. In the first fully-connected layer we have a 10% dropout step, a flattening step, a dense step with 1,024 ELUs, a batch normalization, and an activation step. In the second fully-connected layer we have another 10% dropout step followed by a dense step with 512 ELUs, batch normalization, and then activation. The final layer is another 10% dropout, a dense step reducing to the final predictions, which are then batch normalized and activated using the sigmoid function.

The "wide" model builds on the average model. The first difference is that the feature extraction step includes a fourth block where the convolutional step has 256 filters. The second difference is that in the prediction step, the number of ELUs in the first fully-connected layer is 4,096 (still reducing down to 512).

The "deep" model differs from the average model in that the feature extraction step has 7 blocks, wherein the convolutional layers have 16, 16, 16, 16, 32, 32, and 64 filters. The prediction step has three fully-connected layers instead of the average model's two, with 2,048, 1,024, and 256 ELUs, respectively.

Finally, the "hard" differs from the average model in that the feature extraction step has 5 blocks, wherein the convolutional layers have 32, 64, 64, 128, and 256 filters. The prediction step has three fully-connected layers instead of the average model's two, with 4,096, 2,048, and 512 ELUs, respectively.

1 . The inception model is named for the base model we use; average because it is a reasonable starting point for a classification task of average difficulty; wide because the model appears wide if visualized vertically (due to the large number of ELUs in the first fully-connected layer); deep because the model appears tall if visualized vertically (due to the number of blocks in the feature extraction step); and hard because it is a complex model for difficult classification tasks.

2 . Usually only the top few layers of a transfer learning model need to be retrained, with the rest kept "frozen." However, since our application is somewhat different than typical object-detection tasks—e.g., we are looking for blank spaces within a table instead of identifying faces—our models perform better when all layers are kept trainable.

A.2.2 Data augmentations

All images, irrespective of data augmentation, undergo the following preprocessing steps. First, images are cropped to the region of the form where the relevant feature is consistently found. For instance, the QR code is always in the top-left corner of the form, so we automatically zoom in on this region for the QR code irregularity. Second, the cropped images are resized to be square, so that the pixel width and pixel height are equal (the precise size is determined by the tuning parameters discussed below). This step increases computational stability and is standard in deep learning models. Third, all images are denoised using a median filter. This technique reduces noise while preserving edges better than other filters, e.g. a Gaussian filter, do. Finally, image pixel values are rescaled from RGB coordinates ranging over $[0, 255]$ so that they are expressed as a proportion between 0 and 1.

After these preprocessing steps, the images are then augmented. The Python module we use to fit our deep learning modules, keras (Chollet 2015), provides a number of built-in functions for augmentation through its ImageDataGenerator class. The augmentations relevant to our data and model are as follows:

1. samplewise_center: Set the global mean across all pixels to 0.
2. samplewise_std_normalization: Divide all pixels by the standard deviation across all pixels.
3. horizontal_flip: Flip the image horizontally with 50% probability.
4. vertical_flip Flip the image vertically with 50% probability.
5. brightness_range: Shift the brightness of all pixels by a random value within the range given.
6. zoom_range: Zoom in or out of the image by a random value up to the value given.
7. channel_shift_range: Shift red, blue, and green channels of all pixels by a random value up to the value given.
8. rotation_range: Rotate the image in either direction by a random value up to the value given.
9. width_shift_range: Shift the image horizontally by a random value up to the value given.
10. height_shift_range: Shift the image vertically by a random value up to the value given.
11. shear_range: Shear the image (i.e., fix one axis and rotate the other axis) by a random value up to the value given.

For all augmentations that reveal pixels outside the existing boundaries (e.g., by shifting or rotating the image), empty pixel values are filled using the value of the nearest neighbor, the default setting in keras.

For the purposes of tuning our models, we use three settings: {none, low, high}. In the none condition, none of these data augmentations are applied. In the low setting, we turn on samplewise centering, samplewise scaling, horizontal flipping, and vertical flipping; set the brightness range and zoom range to $[0.95, 1.05]$; set the channel shift range to 5; and set the rotation range, width shift range, height shift range, and shear range to $[-0.05, 0.05]$. In the high setting, we turn on samplewise centering, samplewise scaling, horizontal flipping, and vertical flipping; set the brightness range and zoom range to $[0.70, 1.30]$; set the channel shift range to 30; and set the rotation range, width shift range, height shift range, and shear range to $[-0.30, 0.30]$.

We note here that keras's data augmentation is not replication-safe because it does not respect random seeds set at the beginning of a script call. We therefore implement a custom ImageDataGenerator class and a custom Generator sequence to call the class, bypassing this problem. This code can be found in our replication archive.

A.2.3 Tuning procedures

For each of our 10 irregularities, we tune over the 60 models in a $\{5 \times 3 \times 2 \times 2\}$ grid. This grid consists of:

- Model architecture: inception, average, wide, deep, hard;
- Data augmentation: none, low, high;
- Image size: 256 x 256 pixels, 512 x 512 pixels; and
- Batch size: 16, 32.

These 600 models were run using slurm arrays on Cardiff's high-performance computing cluster, Supercomputing Wales, the national supercomputing research facility for Wales. Each model was allowed to run until either: its out-of-sample F_1 achieved 0.99, it reached 300 epochs, or it timed out after 48 hours (the maximum run-time allowed on the cluster).

A.3 Diagnostic plots, classification model training

Below are training diagnostic plots for each of the ten (tuned) classification models we estimate. Each plot provides the accuracy, loss, and mean squared error (MSE) for the training and test data. Higher accuracy values and lower loss and MSE values indicate better performance.

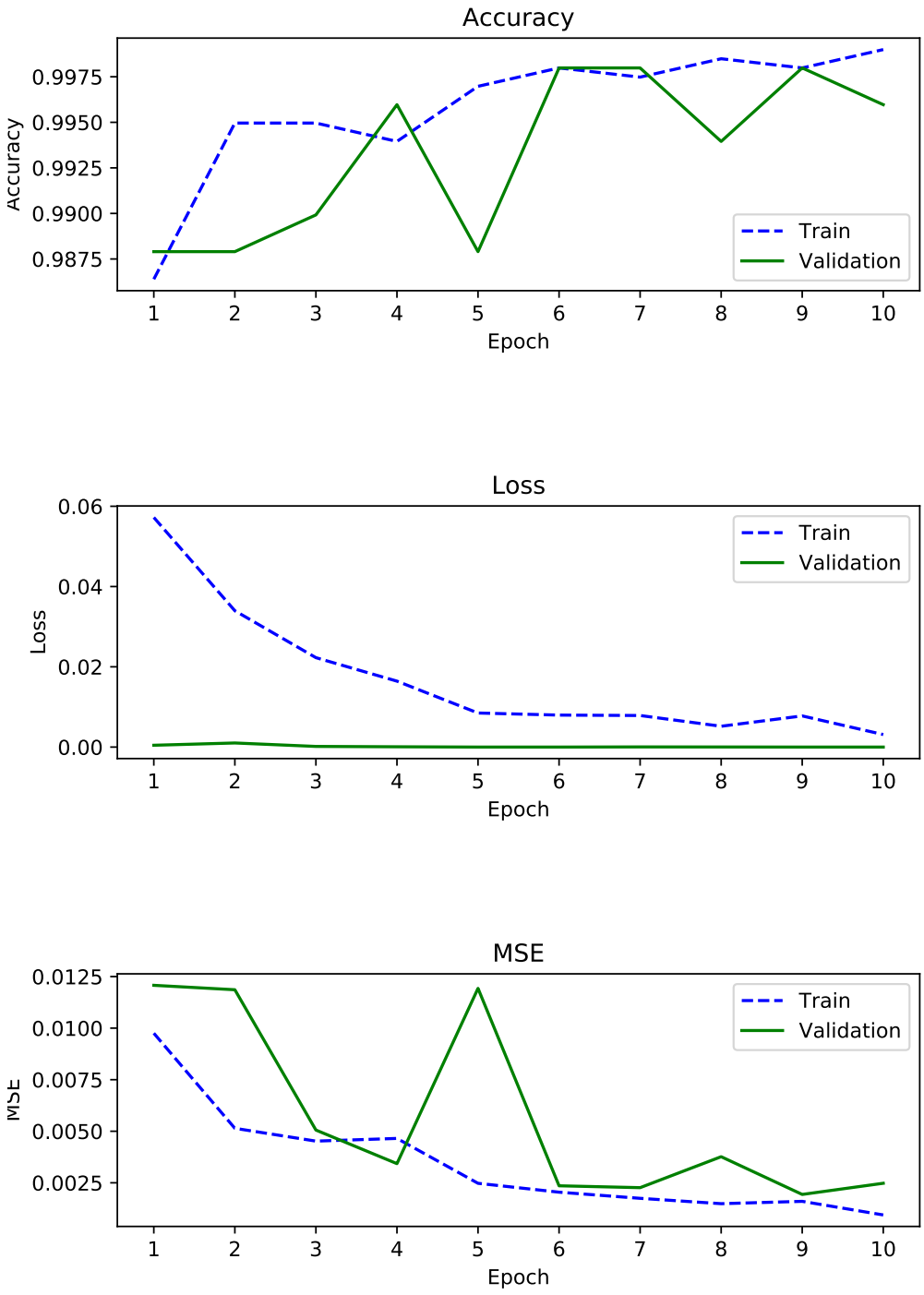


Figure A2: Training diagnostic plots, QR code missing

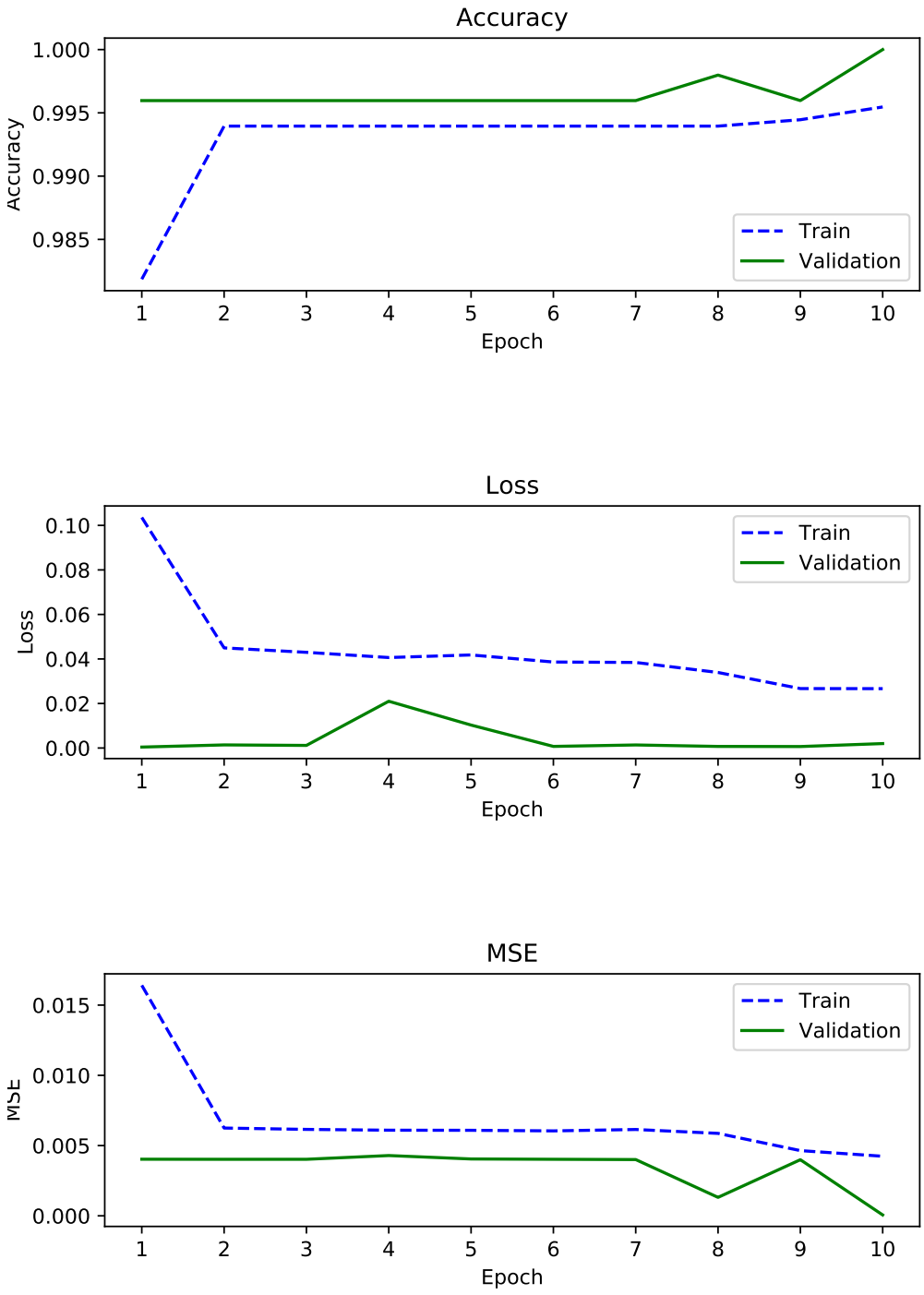


Figure A3: Training diagnostic plots, poor scan quality

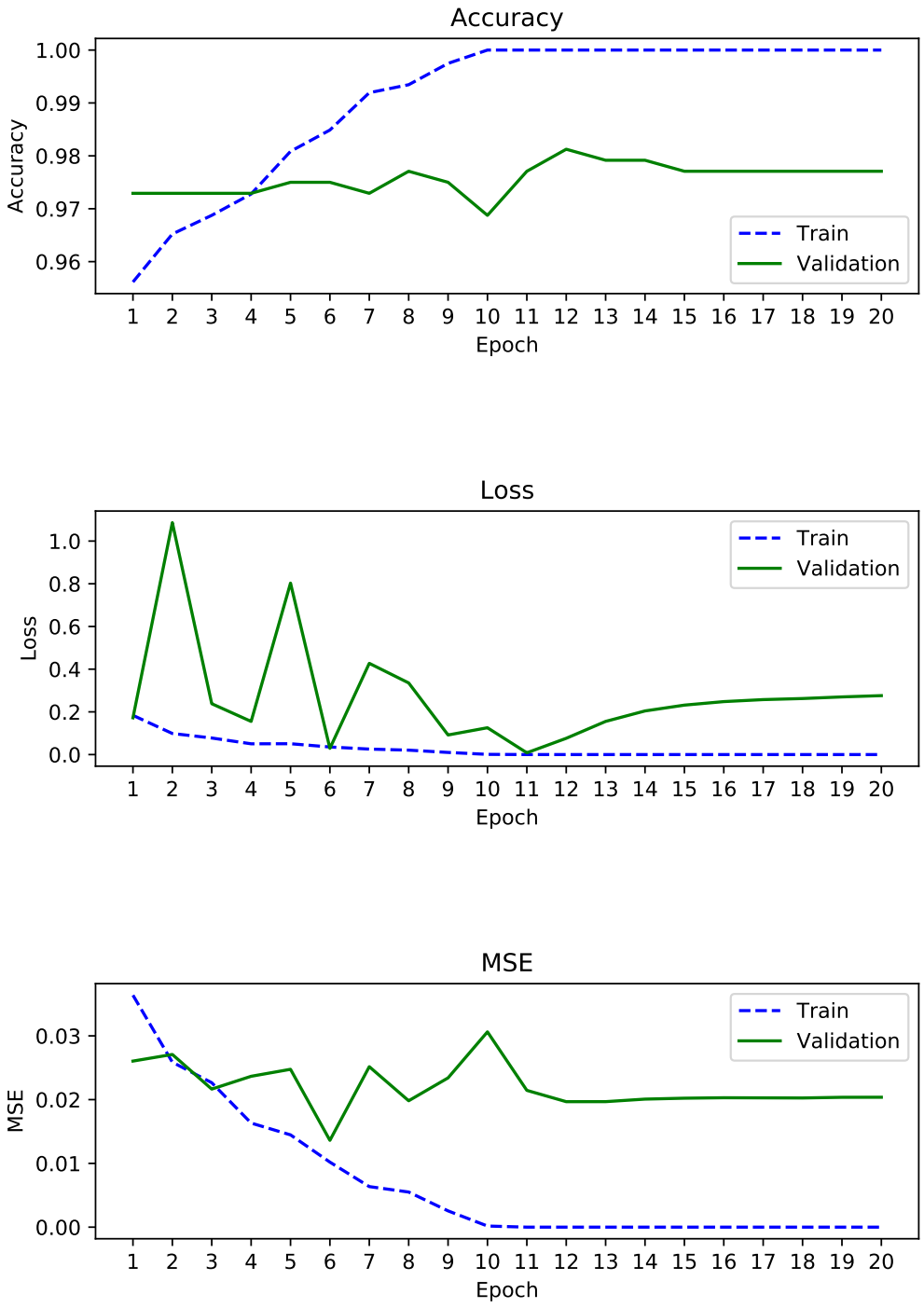


Figure A4: Training diagnostic plots, form not stamped

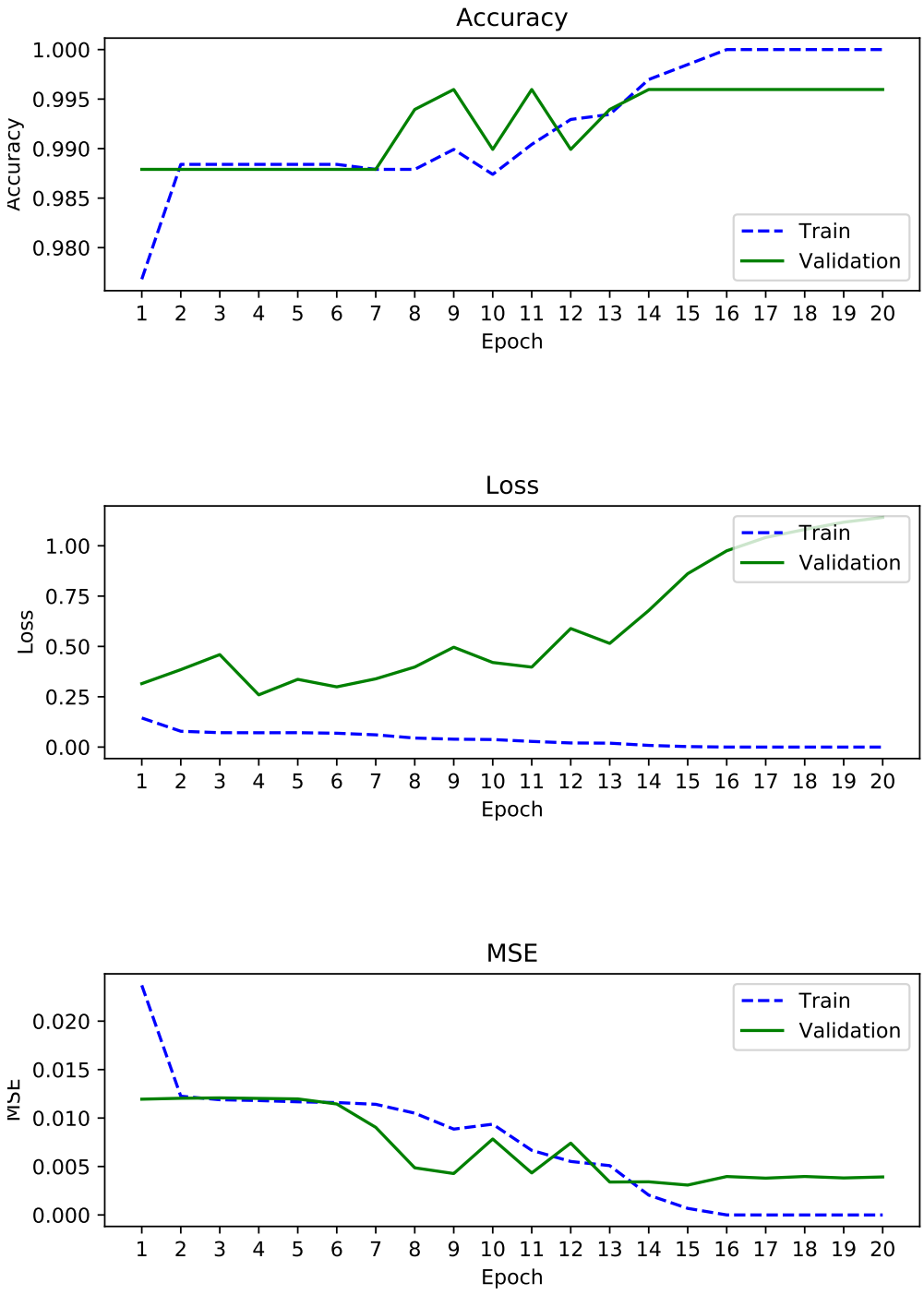


Figure A5: Training diagnostic plots, presiding officer did not sign

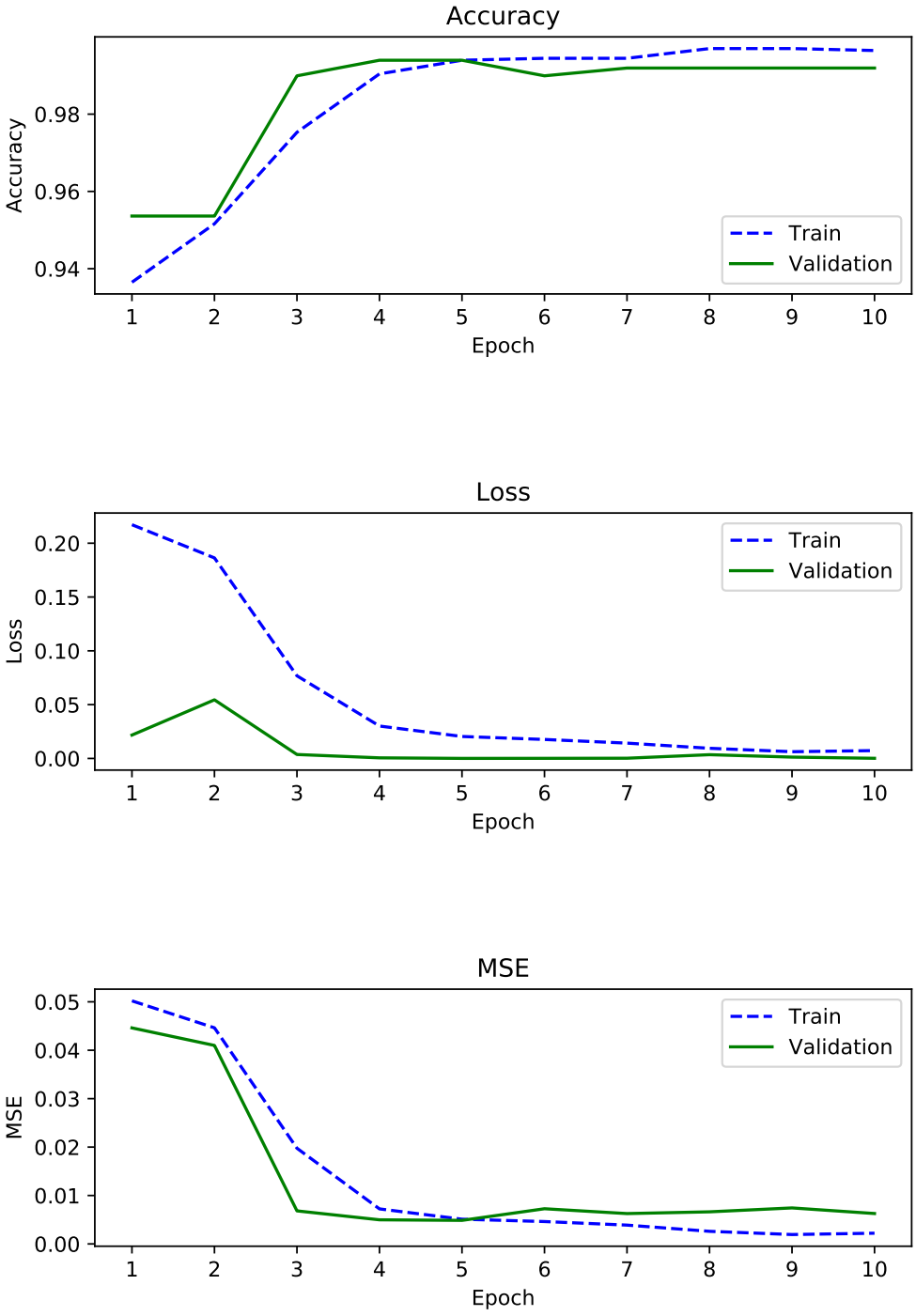


Figure A6: Training diagnostic plots, no agents listed

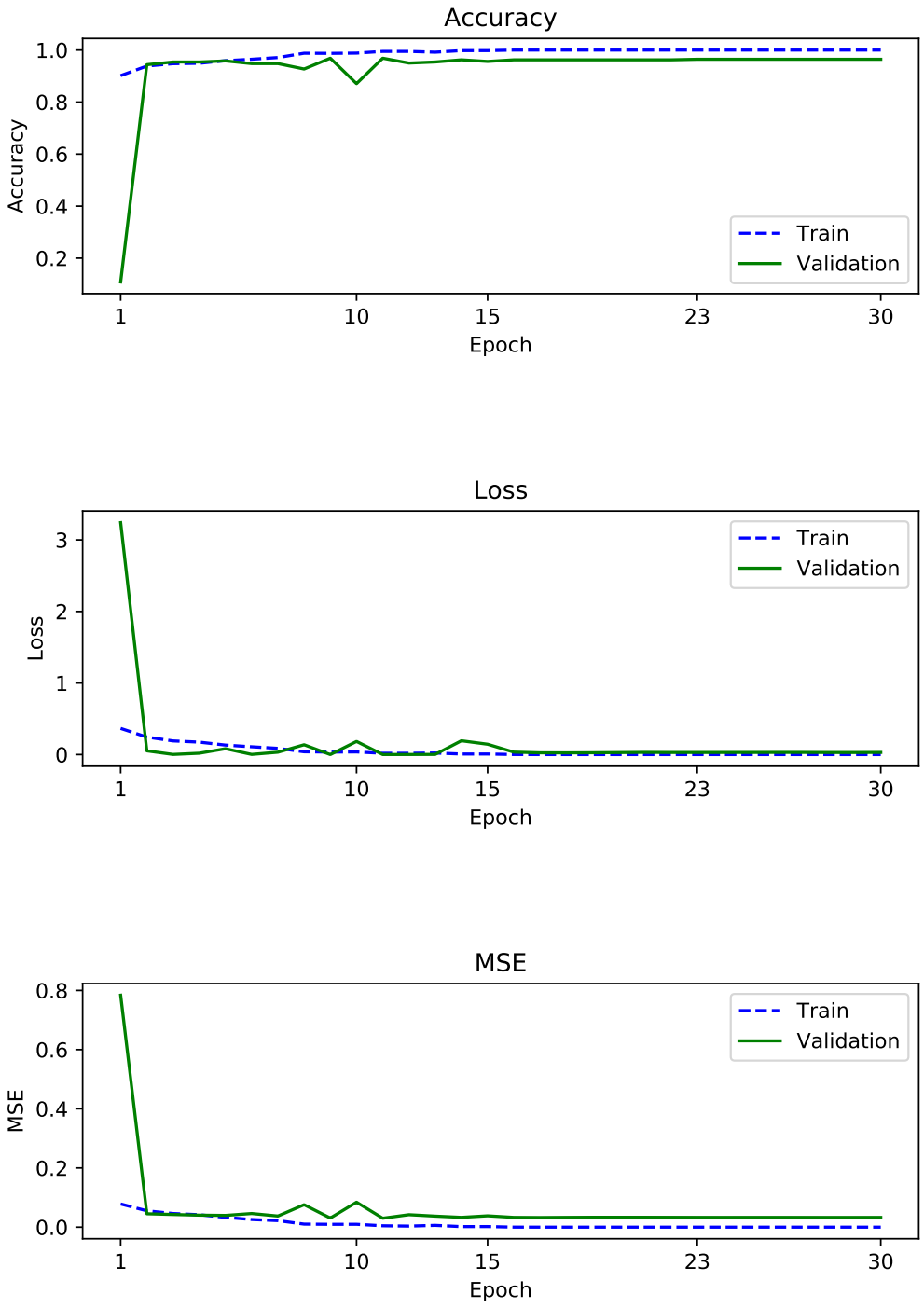


Figure A7: Training diagnostic plots, any agent did not sign

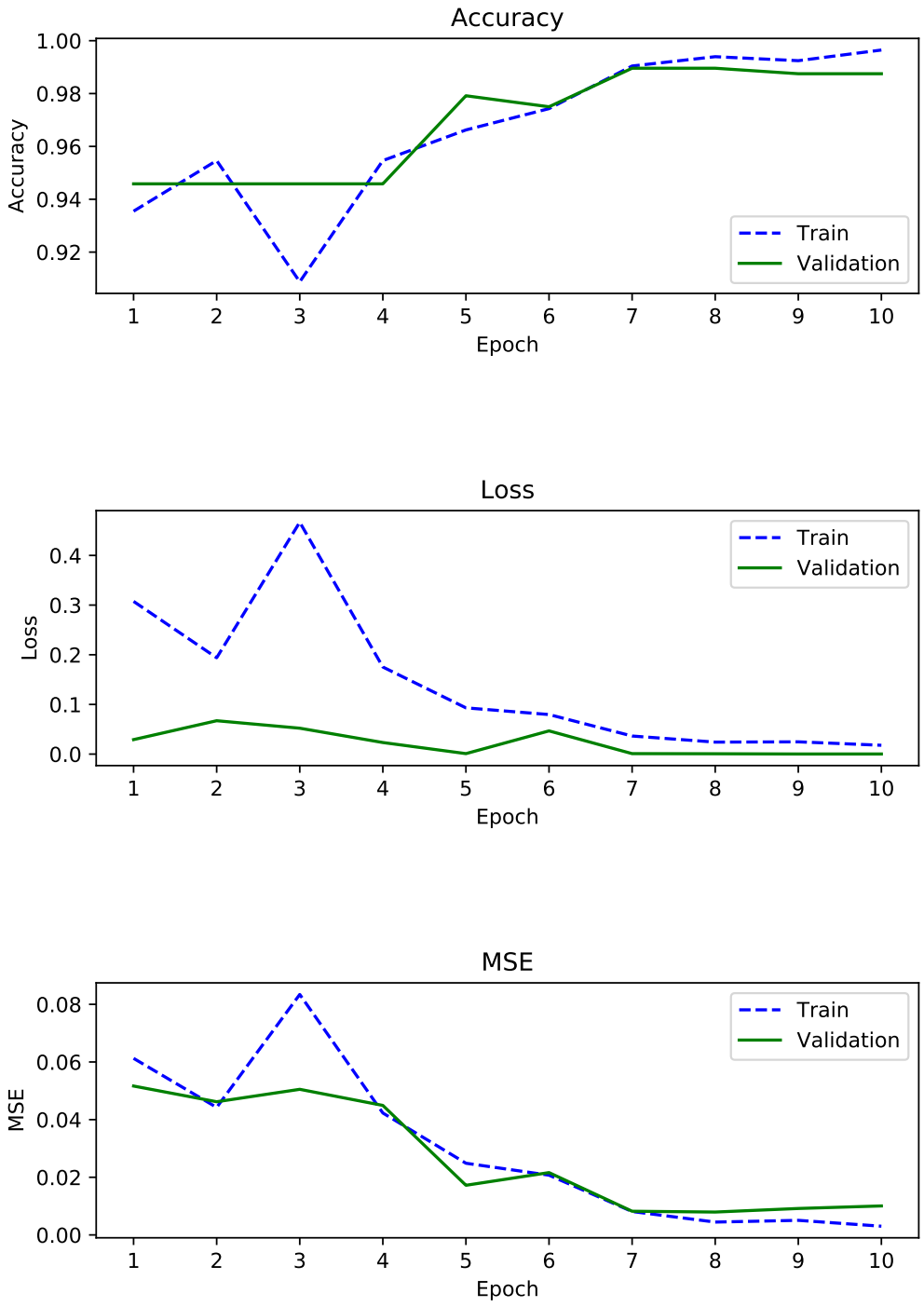


Figure A8: Training diagnostic plots, no agents signed

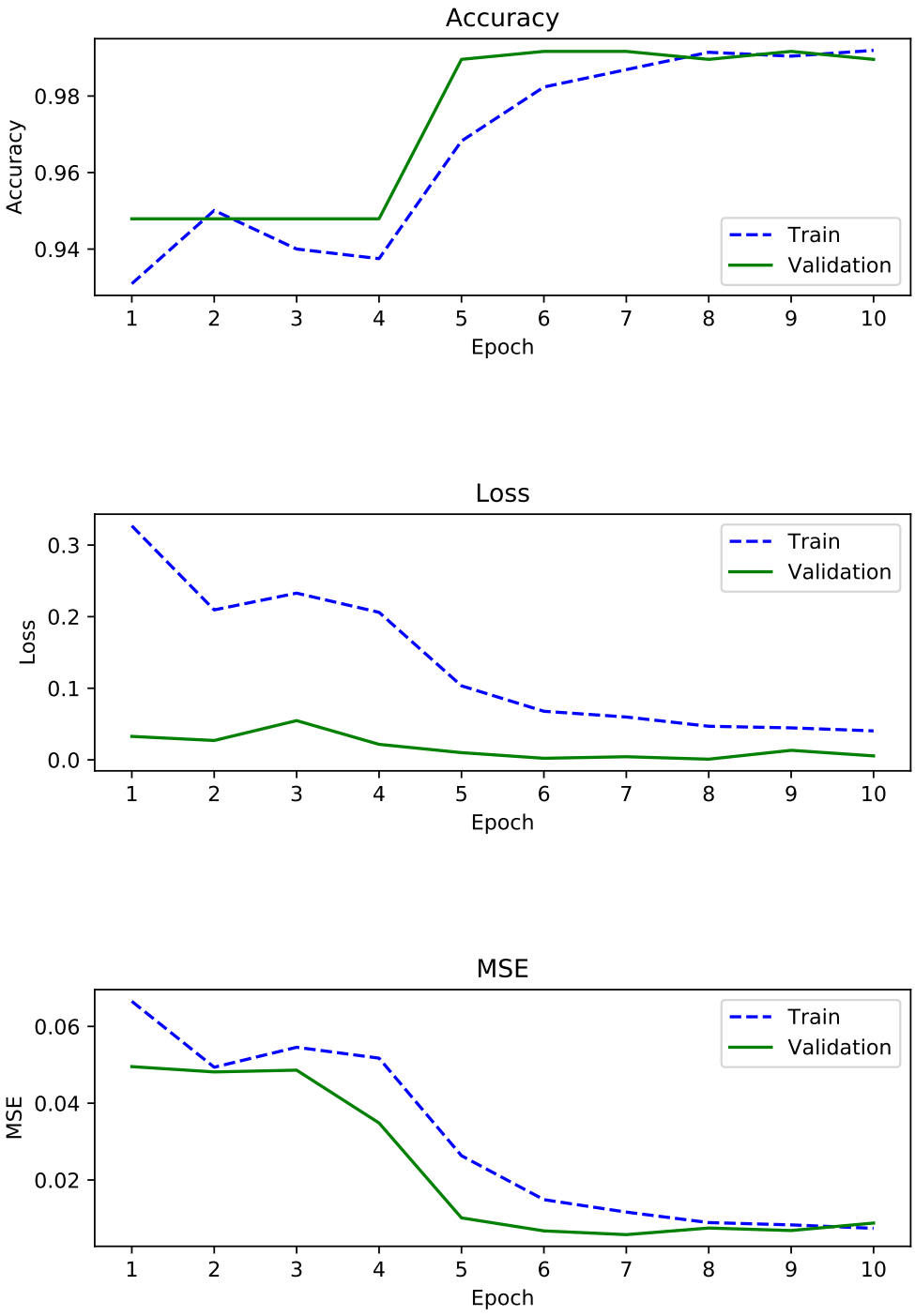


Figure A9: Training diagnostic plots, agent signatures appear identical

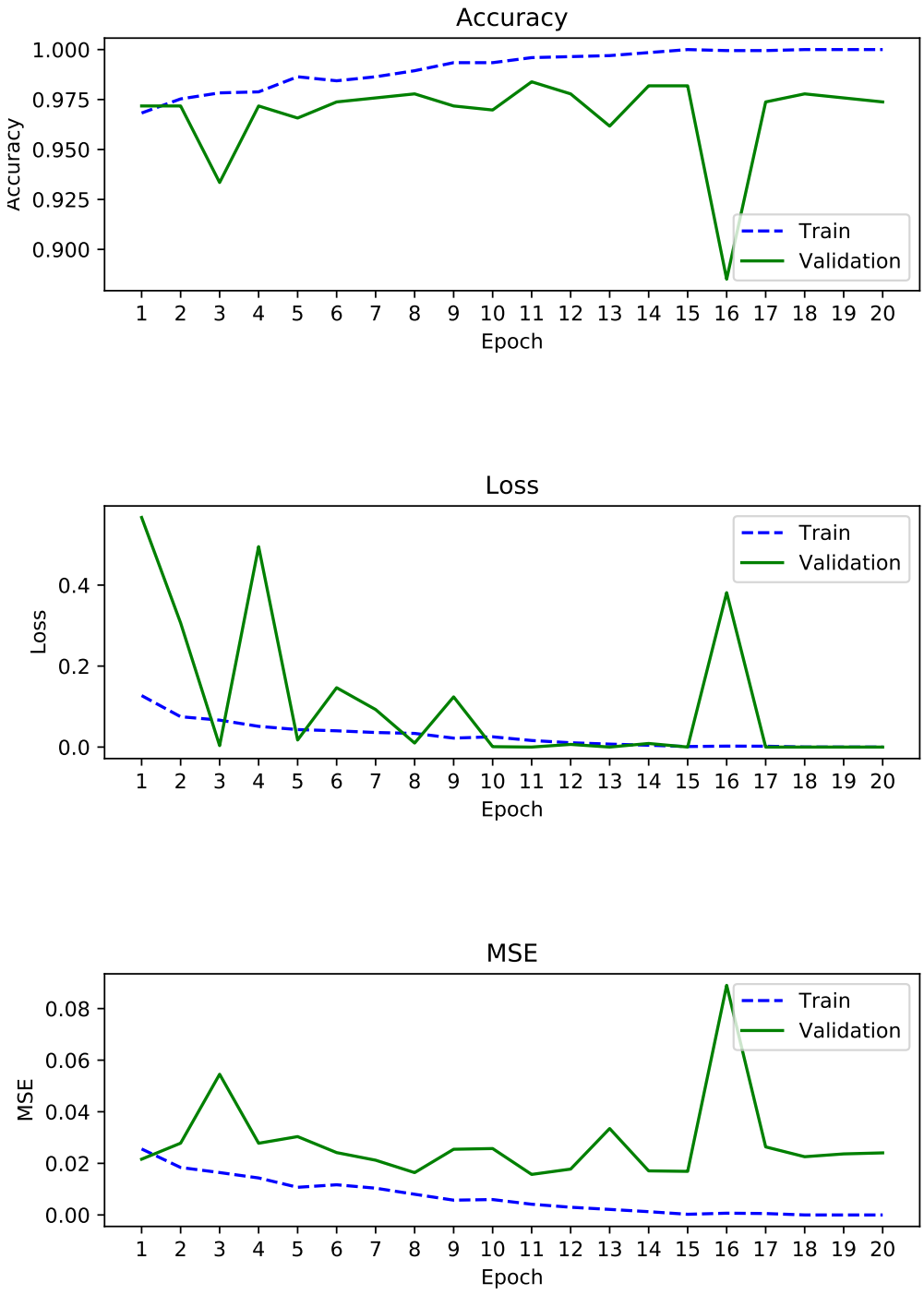


Figure A10: Training diagnostic plots, agent refusal to sign listed

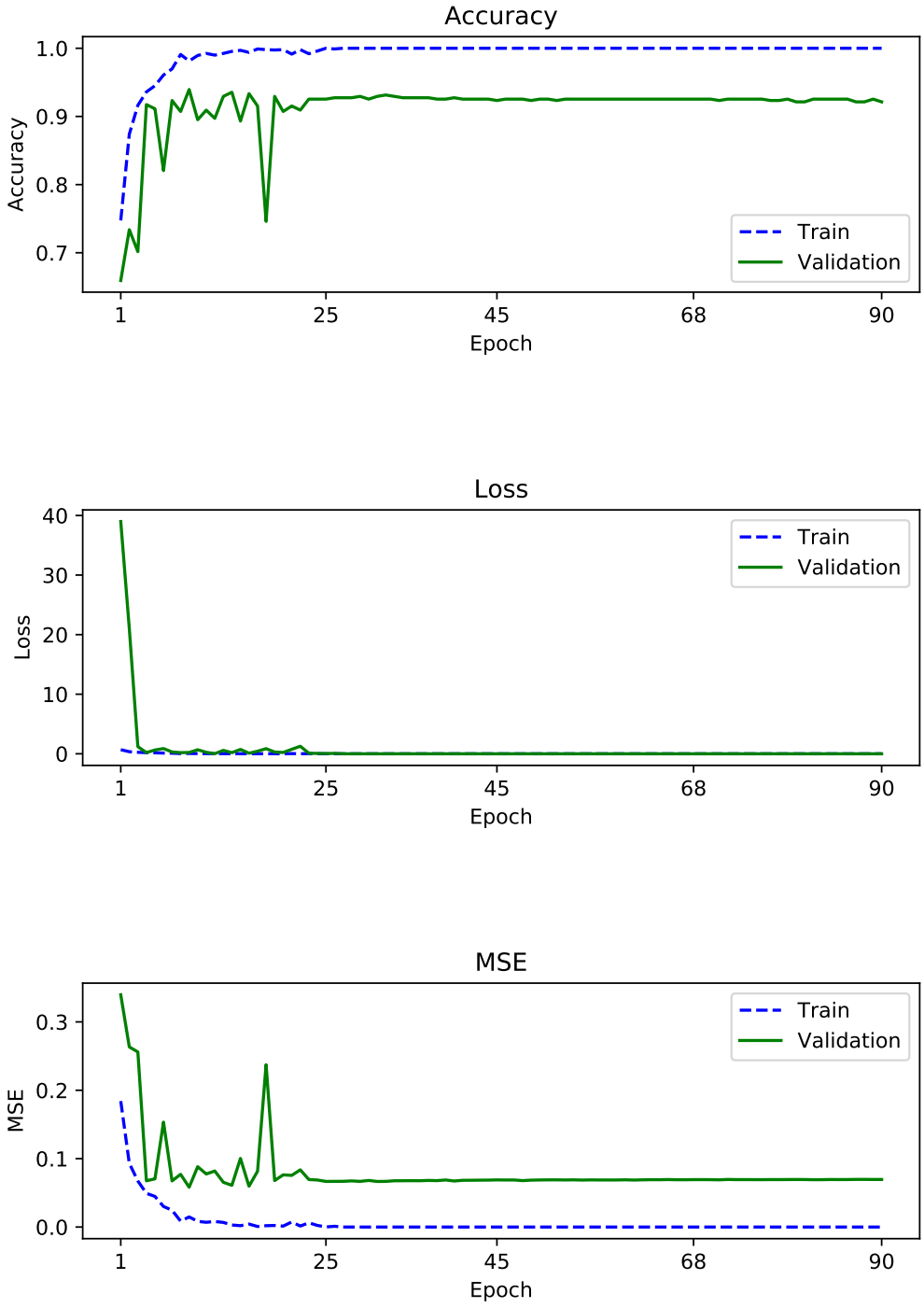


Figure A11: Training diagnostic plots, results edited

A.4 Confusion matrices, classification model training

Below are confusion matrices from our image classification models. Each is constructed using the true test sample of approximately 500 images left out from model training. Note that some samples are slightly smaller than 500 due to data missingness (i.e., missing, partially missing, corrupted, or otherwise incomplete statutory forms).

Table A1: Confusion matrix, QR code missing

	True non-irregularity	True irregularity
Predicted non-irregularity	496	0
Predicted irregularity	1	3

Table A2: Confusion matrix, poor scan quality

	True non-irregularity	True irregularity
Predicted non-irregularity	495	0
Predicted irregularity	0	2

Table A3: Confusion matrix, form not stamped

	True non-irregularity	True irregularity
Predicted non-irregularity	479	8
Predicted irregularity	4	7

Table A4: Confusion matrix, presiding officer did not sign

	True non-irregularity	True irregularity
Predicted non-irregularity	494	0
Predicted irregularity	0	4

Table A5: Confusion matrix, no agents listed

	True non-irregularity	True irregularity
Predicted non-irregularity	471	2
Predicted irregularity	2	22

Table A6: Confusion matrix, any agent did not sign

	True non-irregularity	True irregularity
Predicted non-irregularity	440	18
Predicted irregularity	6	33

Table A7: Confusion matrix, no agents signed

	True non-irregularity	True irregularity
Predicted non-irregularity	468	0
Predicted irregularity	3	26

Table A8: Confusion matrix, agent signatures appear identical

	True non-irregularity	True irregularity
Predicted non-irregularity	469	1
Predicted irregularity	2	25

Table A9: Confusion matrix, agent refusal to sign listed

	True non-irregularity	True irregularity
Predicted non-irregularity	482	4
Predicted irregularity	2	9

Table A10: Confusion matrix, results edited

	True non-irregularity	True irregularity
Predicted non-irregularity	336	25
Predicted irregularity	10	129

A.5 Computation of F_1 scores for Cantú (2019)

Table 1 in Cantú (2019) provides a confusion matrix with mean values across 20 samples, each of 150 images. We can approximate the F_1 score across this sample by multiplying each cell in the table by 3,000 (20×150). This gives 2,790 true negatives (top-left), 210 false positives (top-right), 450 false negatives (bottom-left), and 2,550 true positives (bottom-right). Applying the formula supplied in the main text (our paper), the approximate precision is 0.924, while the approximate recall is 0.850. F_1 is then given by the harmonic mean:

$$F_1 \text{ (Cantú)} = 2\left[\frac{(0.924)(0.850)}{0.924 + 0.850}\right] \approx 0.885 \tag{A1}$$

A.6 Regression results, government stronghold analysis

Table A11: Irregularities are unrelated to government party strength

	Agent problems		Edited results	
Gov. stronghold	0.00	0.01	−0.00	0.02
	(0.01)	(0.01)	(0.00)	(0.02)
Population density		−0.00		0.02*
		(0.00)		(0.01)
Terrain ruggedness		−0.00		0.00
		(0.00)		(0.00)
Ethnic fractionalization		0.02		0.07*
		(0.01)		(0.02)
Poll. station isolation		0.11*		−0.18*
		(0.05)		(0.07)
Poverty rate		−0.01		−0.01
		(0.01)		(0.01)
Literacy rate		0.01		0.01
		(0.00)		(0.01)
Night lights		−0.00		−0.02
		(0.01)		(0.01)
Constituency FE	✓	✓	✓	✓

*$p < .05$. As indicated in the main text, polling stations are coded as government strongholds if the proportion of registered voters who are Kikuyu or Kalenjin is greater than 90%. All models include constituency fixed effects and cluster standard errors by constituency.

A.7 Covariate data sources

Population density data come from Linard et al. (2012), and is computed as the log plus one of raw population density. Data on terrain ruggedness are calculated from a digital elevation model and data provided by the United States Geological Survey (2004). Ethnic fractionalization is computed as one minus the sum of squares of the size of each ethnic group expressed as a proportion of registered voters; these proportions are estimated from the 2013 voter register using the method described in Harris (2015). Polling station isolation is computed as the minimum distance to the nearest polling center, weighted by the number of registered voters, using a method described in Doogan et al. (2018). Data on the poverty rate are described in Linard et al. (2012), night time lights data are from National Oceanic and Atmospheric Administration (2013), and literacy rate data are given in Bosco et al. (2017).

A.8 Full results, Table 3 (main text)

Table A12: Turnout models, no controls

Document problem	−0.01			
	(0.02)			
Procedure problem		−0.01		
		(0.01)		
Agent problem			−0.00	
			(0.00)	
Edited results				−0.01*
				(0.00)
Constituency FE	✓	✓	✓	✓

*$p < .05$. All models include constituency fixed effects and cluster standard errors by constituency.

Table A13: Turnout models, with controls

Document problem	−0.02			
	(0.02)			
Procedure problem		−0.01		
		(0.01)		
Agent problem			−0.00	
			(0.00)	
Edited results				−0.00*
				(0.00)
Population density	−0.01*	−0.01*	−0.01*	−0.01*
	(0.00)	(0.00)	(0.00)	(0.00)
Terrain ruggedness	0.00*	0.00*	0.00*	0.00*
	(0.00)	(0.00)	(0.00)	(0.00)
Ethnic fractionalization	−0.04*	−0.04*	−0.04*	−0.04*
	(0.01)	(0.01)	(0.01)	(0.01)
Poll. station isolation	−0.10	−0.10	−0.10	−0.10
	(0.06)	(0.06)	(0.06)	(0.06)
Poverty rate	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Night lights	−0.01*	−0.01*	−0.01*	−0.01*
	(0.00)	(0.00)	(0.00)	(0.00)
Literacy rate	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Constituency FE	✓	✓	✓	✓

*$p < .05$. All models include constituency fixed effects and cluster standard errors by constituency.

Table A14: Absolute margin models, no controls

Document problem	−0.02			
	(0.03)			
Procedure problem		−0.00		
		(0.01)		
Agent problem			0.00	
			(0.00)	
Edited results				−0.01*
				(0.00)
Constituency FE	✓	✓	✓	✓

*$p < .05$. All models include constituency fixed effects and cluster standard errors by constituency.

Table A15: Absolute margin models, with controls

Document problem	0.01			
	(0.02)			
Procedure problem		−0.00		
		(0.01)		
Agent problem			0.00	
			(0.00)	
Edited results				−0.00
				(0.00)
Population density	−0.01	−0.01	−0.01	−0.01
	(0.01)	(0.01)	(0.01)	(0.01)
Terrain ruggedness	0.01*	0.01*	0.01*	0.01*
	(0.00)	(0.00)	(0.00)	(0.00)
Ethnic fractionalization	−0.38*	−0.38*	−0.38*	−0.38*
	(0.03)	(0.03)	(0.03)	(0.03)
Poll. station isolation	0.20	0.20	0.20	0.20
	(0.18)	(0.18)	(0.18)	(0.18)
Poverty rate	−0.01	−0.01	−0.01	−0.01
	(0.01)	(0.01)	(0.01)	(0.01)
Night lights	−0.02*	−0.02*	−0.02*	−0.02*
	(0.01)	(0.01)	(0.01)	(0.01)
Literacy rate	−0.03*	−0.03*	−0.03*	−0.03*
	(0.01)	(0.01)	(0.01)	(0.01)
Constituency FE	✓	✓	✓	✓

*$p < .05$. All models include constituency fixed effects and cluster standard errors by constituency.

Table A16: Kenyatta margin models, no controls

Document problem	0.06			
	(0.06)			
Procedure problem		0.00		
		(0.02)		
Agent problem			−0.00	
			(0.01)	
Edited results				0.00
				(0.00)
Constituency FE	✓	✓	✓	✓

$^{*}p < .05$. All models include constituency fixed effects and cluster standard errors by constituency.

Table A17: Kenyatta margin models, with controls

Document problem	0.05			
	(0.04)			
Procedure problem		0.01		
		(0.02)		
Agent problem			0.00	
			(0.01)	
Edited results				0.00
				(0.00)
Population density	−0.01	−0.01	−0.01	−0.01
	(0.02)	(0.02)	(0.02)	(0.02)
Terrain ruggedness	0.04*	0.04*	0.04*	0.04*
	(0.01)	(0.01)	(0.01)	(0.01)
Ethnic fractionalization	−0.23*	−0.23*	−0.23*	−0.23*
	(0.07)	(0.07)	(0.07)	(0.07)
Poll. station isolation	0.19	0.19	0.19	0.19
	(0.33)	(0.33)	(0.33)	(0.33)
Poverty rate	−0.05	−0.05	−0.05	−0.05
	(0.03)	(0.03)	(0.03)	(0.03)
Night lights	0.00	0.00	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)
Literacy rate	−0.03	−0.03	−0.03	−0.03
	(0.02)	(0.02)	(0.02)	(0.02)
Constituency FE	✓	✓	✓	✓

$^{*}p < .05$. All models include constituency fixed effects and cluster standard errors by constituency.

Table A18: Rejected vote models, no controls

Document problem	0.46			
	(0.53)			
Procedure problem		0.44		
		(0.33)		
Agent problem			0.20*	
			(0.09)	
Edited results				0.62*
				(0.10)
Constituency FE	✓	✓	✓	✓

*$p < .05$. All models include constituency fixed effects and cluster standard errors by constituency.

Table A19: Rejected vote models, with controls

Document problem	0.40			
	(0.70)			
Procedure problem		0.47		
		(0.35)		
Agent problem			0.16	
			(0.10)	
Edited results				0.55*
				(0.10)
Population density	0.47*	0.47*	0.47*	0.46*
	(0.11)	(0.11)	(0.11)	(0.10)
Terrain ruggedness	−0.03	0.03	−0.03	−0.03
	(0.05)	(0.05)	(0.05)	(0.05)
Ethnic fractionalization	1.24*	1.24*	1.24*	1.21*
	(0.27)	(0.27)	(0.27)	(0.27)
Poll. station isolation	1.68	1.63	1.67	1.63
	(1.62)	(1.62)	(1.62)	(1.61)
Poverty rate	−0.48*	−0.47*	−0.48*	−0.46*
	(0.11)	(0.11)	(0.11)	(0.11)
Night lights	0.21*	0.21*	0.21*	0.21*
	(0.10)	(0.10)	(0.10)	(0.10)
Literacy rate	−0.35*	−0.35*	−0.35*	−0.33*
	(0.16)	(0.16)	(0.16)	(0.16)
Constituency FE	✓	✓	✓	✓

*$p < .05$. All models include constituency fixed effects and cluster standard errors by constituency.

Table A20: Disputed vote models, no controls

Document problem	−0.00			
	(0.00)			
Procedure problem		0.01		
		(0.01)		
Agent problem			−0.00	
			(0.00)	
Edited results				0.00
				(0.00)
Constituency FE	✓	✓	✓	✓

$^{*}p < .05$. All models include constituency fixed effects and cluster standard errors by constituency.

Table A21: Disputed vote models, with controls

Document problem	−0.00			
	(0.00)			
Procedure problem		0.01		
		(0.01)		
Agent problem			−0.00	
			(0.10)	
Edited results				0.00
				(0.00)
Population density	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Terrain ruggedness	−0.00	−0.00	−0.00	−0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Ethnic fractionalization	−0.01	−0.01	−0.01	−0.01
	(0.00)	(0.00)	(0.00)	(0.00)
Poll. station isolation	−0.01	−0.01	−0.01	−0.01
	(0.04)	(0.04)	(0.04)	(1.04)
Poverty rate	−0.00	−0.00	−0.00	−0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Night lights	−0.00	−0.00	−0.00	−0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Literacy rate	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Constituency FE	✓	✓	✓	✓

$^{*}p < .05$. All models include constituency fixed effects and cluster standard errors by constituency.

Table A22: Spoiled vote models, no controls

Document problem	0.08			
	(0.32)			
Procedure problem		0.34*		
		(0.16)		
Agent problem			0.08	
			(0.05)	
Edited results				0.37*
				(0.03)
Constituency FE	✓	✓	✓	✓

*$p < .05$. All models include constituency fixed effects and cluster standard errors by constituency.

Table A23: Spoiled vote models, with controls

Document problem	−0.15			
	(0.24)			
Procedure problem		0.34*		
		(0.16)		
Agent problem			0.08	
			(0.05)	
Edited results				0.36*
				(0.03)
Population density	0.24*	0.23*	0.24*	0.22*
	(0.04)	(0.04)	(0.04)	(0.04)
Terrain ruggedness	−0.05*	−0.06*	−0.05*	−0.06*
	(0.02)	(0.02)	(0.02)	(0.02)
Ethnic fractionalization	0.18*	0.19*	0.18*	0.17*
	(0.08)	(0.08)	(0.08)	(0.07)
Poll. station isolation	1.38*	1.34	1.37*	1.35
	(0.69)	(0.69)	(0.69)	(0.69)
Poverty rate	−0.08*	−0.09*	−0.08*	−0.08*
	(0.04)	(0.04)	(0.04)	(0.04)
Night lights	0.02	0.02	0.02	0.01
	(0.04)	(0.04)	(0.04)	(0.04)
Literacy rate	−0.13*	−0.13*	0.13*	−0.12*
	(0.04)	(0.04)	(0.04)	(0.04)
Constituency FE	✓	✓	✓	✓

*$p < .05$. All models include constituency fixed effects and cluster standard errors by constituency.

A.9 Robustness checks, Table 3 (main text)

Table A24: Ward fixed effects

	Turnout		Abs. margin		Margin	
Document prob.	−0.01	−0.02	0.02	0.05	0.04	0.06
	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Procedure prob.	−0.01	−0.01	−0.01	−0.00	−0.03	−0.03
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
Agent prob.	−0.00	−0.00	0.00	0.00	−0.01	−0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)
Results edited	−0.00*	−0.00*	−0.01*	−0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	Bal. rejected		Bal. disputed		Bal. spoiled	
Document prob.	0.26	−0.06	−0.00	−0.00	0.24	0.03
	(0.58)	(0.69)	(0.00)	(0.00)	(0.31)	(0.25)
Procedure prob.	0.50	0.50	0.01	0.01	0.32*	0.31
	(0.36)	(0.38)	(0.01)	(0.01)	(0.16)	(0.16)
Agent prob.	0.20*	0.16	−0.00	−0.00	0.10*	0.10*
	(0.09)	(0.10)	(0.00)	(0.00)	(0.05)	(0.05)
Results edited	0.61*	0.56*	0.00	0.00	0.36*	0.36*
	(0.10)	(0.10)	(0.00)	(0.00)	(0.03)	(0.03)
Controls		✓		✓		✓
Ward FE	✓	✓	✓	✓	✓	✓

*$p < .05$. Estimates are from linear models with standard errors in parentheses. Controls are as described in Section A.7. All models include ward fixed effects and standard errors clustered by ward.

Table A25: County fixed effects

	Turnout		Abs. margin		Margin	
Document prob.	−0.01	−0.02	0.07	0.01	−0.04	0.11
	(0.01)	(0.02)	(0.05)	(0.03)	(0.06)	(0.07)
Procedure prob.	−0.00	−0.00	0.00	0.00	−0.02	−0.02
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
Agent prob.	−0.00	−0.00	0.00	0.00	−0.01	−0.01
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)
Results edited	−0.01*	−0.00*	−0.01*	−0.00	−0.00	−0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)
	Bal. rejected		Bal. disputed		Bal. spoiled	
Document prob.	−1.08	0.44	−0.00	−0.00	0.04	−0.19
	(0.91)	(0.76)	(0.00)	(0.00)	(0.16)	(0.19)
Procedure prob.	0.43	0.49*	0.01	0.01	0.31*	0.32*
	(0.21)	(0.22)	(0.01)	(0.01)	(0.14)	(0.14)
Agent prob.	0.15	0.15	−0.00	−0.00	0.06	0.06
	(0.10)	(0.10)	(0.00)	(0.00)	(0.05)	(0.05)
Results edited	0.70*	0.59*	0.00	0.00	0.37*	0.36*
	(0.09)	(0.10)	(0.00)	(0.00)	(0.03)	(0.03)
Controls		✓		✓		✓
County FE	✓	✓	✓	✓	✓	✓

*$p < .05$. Estimates are from linear models with standard errors in parentheses. Controls are as described in Section A.7. All models include county fixed effects and standard errors clustered by county.

Table A26: No fixed effects

	Turnout		Abs. margin		Margin	
Document prob.	−0.04*	−0.04*	−0.06	0.01	−0.46*	−0.11
	(0.01)	(0.01)	(0.03)	(0.05)	(0.08)	(0.14)
Procedure prob.	−0.01	−0.01	−0.01	−0.00	−0.09	−0.05
	(0.01)	(0.01)	(0.02)	(0.02)	(0.06)	(0.06)
Agent prob.	−0.00	−0.00	−0.02*	−0.01*	−0.01	0.00
	(0.00)	(0.00)	(0.01)	(0.01)	(0.02)	(0.02)
Results edited	−0.01*	−0.01*	−0.02*	−0.01*	0.06*	0.05*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)
	Bal. rejected		Bal. disputed		Bal. spoiled	
Document prob.	0.68	0.34	−0.00	−0.00	0.07	−0.27
	(0.54)	(0.99)	(0.01)	(0.01)	(0.20)	(0.36)
Procedure prob.	0.34	0.50	0.01	0.01	0.33*	0.36*
	(0.40)	(0.39)	(0.01)	(0.01)	(0.14)	(0.14)
Agent prob.	0.14	0.16	−0.00	−0.00	0.08*	0.09*
	(0.11)	(0.11)	(0.00)	(0.00)	(0.04)	(0.04)
Results edited	0.91*	0.62*	0.00	0.00	0.42*	0.36*
	(0.07)	(0.07)	(0.00)	(0.00)	(0.03)	(0.03)
Controls		✓		✓		✓

*$p < .05$. Estimates are from linear models with standard errors in parentheses. Controls are as described in Section A.7.

A.10 Regression results, Figure 3 (main text)

Table A27: Predictors of electoral irregularities

	Irregularity			
	Document	Procedure	Agent	Edited results
Population density	0.00	0.00	−0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Terrain ruggedness	0.00*	−0.00	−0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Ethnic fractionalization	0.00*	0.00	0.03*	0.05*
	(0.00)	(0.00)	(0.01)	(0.01)
Poll. station isolation	0.00	−0.01	0.07*	−0.15*
	(0.00)	(0.01)	(0.03)	(0.04)
Poverty rate	0.00	0.00	0.00	−0.02*
	(0.00)	(0.00)	(0.00)	(0.00)
Night lights	0.00	−0.00	0.01*	0.01
	(0.00)	(0.00)	(0.00)	(0.01)
Literacy rate	0.00	−0.00	−0.00	−0.01
	(0.00)	(0.00)	(0.00)	(0.00)

*$p < .05$. Data sources are described in Section A.7 above.

References

Bosco, Claudio, Victor Alegana, Tomas Bird, Carla Pezzulo, Linus Bengtsson, Alessandro Sorichetta, Jessica Steele, et al. 2017. "Exploring the High-Resolution Mapping of Gender-Disaggregated Development Indicators." *Journal of the Royal Society Interface* 14.

Cantú, Francisco. 2019. "The Fingerprints of Fraud: Evidence from Mexico's 1988 Presidential Election." *American Political Science Review* 113 (3): 710–726.

Chollet, François. 2015. Keras. https://keras.io.

Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. "ImageNet: A Large-Scale Hierarchical Image Database." In *2009 IEEE Conference on Computer Vision and Pattern Recognition,* 248–255.

Doogan, Nathan J., Megan E. Roberts, Mary Ellen Wewers, Erin R. Tanenbaum, Elizabeth A. Mumford, and Frances A. Stillman. 2018. "Validation of a New Continuous Geographic Isolation Scale: A Tool for Rural Health Disparities Research." *Social Science & Medicine* 215:123–132.

Harris, J. Andrew. 2015. "What's in a Name? A Method for Extracting Information about Ethnicity from Names." *Political Analysis* 23 (2): 212–224.

Linard, Catherine, Marius Gilbert, Robert W. Snow, Abdisalan M. Noor, and Andrew J. Tatem. 2012. "Population Distribution, Settlement Patterns, and Accessibility across Africa in 2010." *PLoS ONE* 7 (2).

National Oceanic and Atmospheric Administration. 2013. DMSP-OLS Nighttime Lights Time Series Version 4, Stable Lights Product. Image and data processing by NOAA's National Geophysical Data Center. DMSP data collected by US Air Force Weather Agency.

Szegedy, Christian, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. 2016. "Rethinking the Inception Architecture for Computer Vision." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* 2818–2826.

United States Geological Survey. 2004. USGS SRTM 30M Digital Elevation Model. Available at https://www.usgs.gov. Last accessed 11 July 2020.