Hidden in Plain Sight? Irregularities on Statutory Forms and Electoral Fraud

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

September 12, 2021

Abstract

The most common method of tabulating election results around the world is manually compiling paper forms at the local level. Recent election disputes in developing democracies, particularly in Africa, have centered on irregularities observed on these forms. However, scholars do not yet have a good understanding of the distribution of these irregularities, nor of their relationship to systematic fraud. In this paper, we theorize a catalog of irregularities that goes beyond simple vote tally editing. We use deep neural networks to identify these irregularities on forms from about 30,000 polling stations in Kenya's 2013 presidential election. We find that although irregularities manifest differently in government and opposition strongholds, they do not correlate with election outcomes, and they are unaffected by the presence of electoral observers. Taken together, our findings suggest scholars of election integrity pay greater attention to problems of benign human error and overtaxed bureaucrats.

*Assistant Professor, Purdue University. email zachwarner@purdue.edu, 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.

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

The most common method of compiling election outcomes around the world is to physically produce results forms locally, by hand, on paper. Globally, 77% of countries collate results at the polling station level before sending them on for further aggregation.[1] Recent electoral disputes, particularly in Africa, center on problems found in these forms, the official documents recording the most granular level of election results. For instance, irregularities in polling station-level election forms contributed to the nullification of presidential elections in Malawi (2019) and Kenya (2017); unsuccessful legal challenges following the Zimbabwe (2018) and Kenya (2013) elections similarly implicated form-based irregularities.

All elections produce irregularly-marked materials like incorrect tabulations (Ansolabehere et al. 2018) or rejected or miscast ballots (e.g., Friesen 2019; Power and Roberts 1995; Niemi and Herrnson 2003). In some cases, patterns of such irregularities may relate to features of the electorate (as in Stiefbold 1965) or of election administration (e.g., Challú, Seira, and Simpser 2020; Goggin, Byrne, and Gilbert 2012). While well-studied in the U.S. context, we know much less about election irregularities in low- and middle-income countries. Although form-based electoral irregularities may occur through normal human error, such errors may also be fraud. Deadlines for legal challenges create a very real verification problem: how can an impartial accounting of form-related irregularities—and their relationship to numeric election results—occur in a relatively short amount of time? Without a better understanding of what these irregularities are and how they arise, partisan actors may selectively present irregular forms as evidence of systematic fraud. In environments with low trust in democratic processes or where partisans are clustered in relatively homogenous parts of the country, form-based irregularities can pose a real risk to democratic consolidation.

In this paper, we use machine learning to catalog irregularities found on electoral forms produced at the polling station level, focusing on document quality, procedural missteps, agent participation, and vote tally editing. By automating the detection of such problems, we reduce the cost of understanding the distribution and magnitude of form-based irregularities. We then examine how

1. ACE Project (2021), question VC004: "Following the close of the voting, where are the votes first sorted and counted?"

these irregularities relate to election outcomes like turnout, vote share, and vote margins. We show that government and opposition strongholds exhibited markedly different patterns of irregularities in Kenya's 2013 election—but that these patterns do not translate to meaningful differences in election outcomes. And we use data on randomly-assigned election observers from the National Democratic Institute's (NDI's) local partner, the Election Observation Group (ELOG), to demonstrate that observation does not impact the distribution of irregularities among polling stations.

Our motivating case is Kenya's contentious 2013 presidential election. As a close election in a developing democracy with mixed success in delivering reliable elections, Kenya's 2013 contest is an ideal case to probe the distribution of election irregularities and their connection with numeric election results. The incumbent narrowly avoided a runoff, with just over 50% of the vote declared in the incumbent's favor. Moreover, local turnout at polling stations was high, with a median of 88%, and polarized, with an average margin between the top two candidates of 83%.[2] The tight margin and locally polarized results coincided with reports of electoral maladministration, leading many observers to infer that the results were evidence of fraud instead of what President Uhuru Kenyatta referred to as just "one or two clerical errors" (Gettleman 2013). Our analysis focuses on official election results forms from each polling station. Known as the "Form 34A" in Kenya, these documents provide information such as vote counts by candidate, rejected votes, and the names and signatures of presiding officers and political agents present during the count. These statutory forms provide the most granular data available on electoral process and outcomes across approximately 30,000 polling stations.

This paper makes four contributions to the study of electoral integrity (Norris, Frank, and Martnez i Coma 2014; Norris 2014; Garnett 2019). First, we describe empirically the widespread occurrence of form-based irregularities. We operationalize a series of irregularities as indicators of divergence from established norms and procedures, aside from assumptions about intentionality. This definition avoids conceptual slippage between irregularities and fraud, a problem which can have enormous consequences, including damaging trust in (however imperfect) democratic institutions

2. Shah (2015) and Shilaho (2013) provide able discussions of the broader context surrounding the election.

(e.g., Daxecker, Di Salvatore, and Ruggeri 2019; Kerr and Lührmann 2017; Erlich and Kerr 2016).[3] Moreover, by emphasizing the primary material realization of election results, our work refocuses studies of electoral integrity on the election-day production of election results by poll workers (Hall, Monson, and Patterson 2009; Burden and Milyo 2015; James 2019; Neggers 2018).

Second, we broaden scholarly attention from edited vote tallies to a number of other pieces of information that can be gleaned from statutory forms. Challú, Seira, and Simpser (2020) and Cantú (2019) examine these tallies to understand electoral procedures in Mexico, but in many African contexts, election litigation increasingly hinges on other elements of results forms. Understanding problems like missing signatures of election officials, the absence of partisan observers, or want of an official election commission stamp are equally important for scholars wishing to understand electoral administration. By focusing on such procedural requirements, our work bridges existing interrogations of electoral integrity with the basic evidentiary concerns of election litigation.

Third, we build on existing applications of deep learning methods to detect irregularities in statutory forms. Following recent advances (Torres and Cantú 2021), we implement four improvements over the current standard practice in political science. Election form irregularities are often relatively rare. This "class imbalance" presents a challenge to deep learning, as it presents few examples of irregularities on which to train the model. For both model tuning and model selection, we use metrics which are calibrated for imbalance, improving overall performance and avoiding the costly process of manually building a balanced set of training data. Further, we implement transfer learning, starting with a publicly-available model pre-trained on a massive set of images and re-training that model on our data (Caruana 1994; Bengio 2012; Razavian et al. 2014). We also use random data augmentation, permuting each input image each training epoch to synthetically increase our training sample size (Krizhevsky, Sutskever, and Hinton 2012; Wong et al. 2016). And we systematically tune our models' hyperparameters, ensuring that they are optimally learning patterns in irregularities, greatly improving performance. All four improvements increase our ability to detect

3. See Bolivia's October 2019 election and the (Organization of American States 2019) report which may have mistakenly concluded Morales won via fraud (Idrobo, Kronick, and Rodríguez 2020; Williams and Curiel 2020), leading to his resignation.

irregularities on a nationally-representative sample of election forms and enhance out-of-sample predictive power.

Last, our results complicate the notion that form-based irregularities *necessarily* correlate with electoral manipulation of numeric results. We demonstrate that, on average, no relationship exists between polling station level irregularities and election outcomes. Moving beyond observational evidence, we leverage the randomized assignment of observers involved in a parallel vote tabulation in Kenya's 2013 presidential election to show that the presence of such observers caused no change in the quality of election results forms produced at the polling station level or difference in final election results. This null contrasts with existing election observation research finding significant effects of election observation on election outcomes. Such a result is all the more surprising in the context of a closely contested election where relatively small amounts of fraud could change the outcome or induce a second-round runoff. While edited results and other form-based irregularities *can* reflect electoral malfeasance, our evidence indicates problematic forms may be symptomatic of an overtaxed electoral bureaucracy rather than a concerted attempt at electoral manipulation.[4]

Assessing election integrity using statutory forms

The 2013 Kenyan presidential election provides a compelling case in which to study form-based irregularities. Heir-apparent Uhuru Kenyatta and opposition leader Raila Odinga went into election day in a virtual tie in the polls. After a relatively calm election day, the Independent Electoral and Boundaries Commission (IEBC) announced that Kenyatta won 50.51% of the vote, clearing the threshold for a run-off by fewer than ten thousand votes out of over twelve million cast. Kenyatta's narrow victory led to a legal battle to annul the result, with opposition leaders arguing that significant irregularities had vitiated the entire election. Kenya's Supreme Court rejected this claim, and a month later Kenyatta assumed office.

Both sides' legal arguments relied, in part, on the extent and nature of widespread irregularities among polling stations' Forms 34A, which report results from each ballot box nationwide. For

4. See also Challú, Seira, and Simpser (2020) on this point.

instance, the affidavit of Janet Ong'era, executive director of the opposition party, provides a list of forms that "fail the test of integrity of electoral documents," citing reasons such as missing agent signatures or the lack of an official IEBC stamp (Republic of Kenya 2013, para. 38). Kenyatta's team replied with the argument that while examples of such problems could be found, "there were no constitutional or statutory violations" behind them (Supreme Court of Kenya 2013, para. 161). Just as partisanship drove legal interpretations of irregularities, so too did it shape public opinion of them. Because ethnicity is an important determinant of Kenyans' partisanship, 89.2% of 2014 Afrobarometer respondents from Kenyatta's ethnic group (the Kikuyu) reported believing the elections to be free and fair, with only minor problems. In contrast, 76.5% of those of opposition leader Odinga's Luo ethnicity described the election as either not free and fair, or free and fair with major problems.

Similar circumstances have driven electoral disputes in other developing democracies. Malawi's 2019 presidential election was overturned as a direct result of material alteration of forms and procedural irregularities like missing signatures or mathematical errors (Gathii and Akinkugbe 2020; Jegwa 2019). For instance, the Supreme Court of Malawi (2020, paras. 349 and 168) judgment presents numerous allegations of missing or forged signatures, including that "monitors' signatures were missing and/or presiding officers did not sign" and "some forms had forged signatures and those signatures were different from those on forms collected by monitors at the polling centres [sic]." Similarly, Zimbabwe's July 2018 presidential election was disputed by the opposition, in part due to purported irregularities with polling station-level forms, though the court decided the complaints had no merit.[5] Following Kenya's August 2017 presidential election, the Supreme Court of Kenya (2017) cataloged some of these problems in a judgment nullifying the result.[6] The Justices write in paragraph 37 that:

> Some Forms 34A and 34B lacked the names of Returning Officers; some lacked the IEBC authentication stamp; some were not signed by the candidates' agents and no reasons were given for that failure; different polling stations bore the name of the same person as the presiding officer; several Forms 34A were altered and tampered with;

5. See Chamisa vs. Mnangagwa and 24 others (Supreme Court of Zimbabwe 2018).
6. Similar complaints were also central to opposition arguments following the 2007 election (Owuor 2008).

> the number of Forms 34A handed over was not clear; several Forms 34As [sic] were signed by un-gazetted presiding officers; some forms were illegible; the handwriting and signatures on Forms 34A appeared made up; [and] some Forms 34A were filled in the same handwriting.

These legal challenges are organized around a common thread: whether irregular forms signal fraud or just human error. The courts deciding these challenges, like the voters following them, must decide which deviations from statutory procedures represent an attempt to systematically alter election results, and which are simply the result of overworked election officials doing their best with limited resources. They must process large amounts of data embedded in the physical appearance of statutory forms and relate them to micro-level data in partisanship and electoral outcomes.

For judges, voters, and scholars of electoral integrity alike, the challenge is thus to ensure that the results contained in official returns reflect, by some reasonable standard, a good faith effort by electoral administrators to communicate voters' preferences as expressed in the ballots. This "verification problem" is made all the more difficult by three factors. First, even in favorable circumstances, elections nearly always contain inevitable discrepancies and errors (e.g., Ansolabehere et al. 2018). Second, the process by which irregularities are generated is largely unknown, and partisan actors have private motives for committing fraud as well as for alleging it where none exists. Last, typically, the only legal record of an irregularity is the statutory form, which represents the primary aggregated record of political preferences expressed in a given ballot box. Rarely do lawyers or scholars have additional data that they can bring to bear on question of whether these forms accurately reflect voters' preferences.[7]

Two recent studies have used more comprehensive methods to address the verification problem in Mexican elections. Cantú (2019) applies computer vision tools—an approach we build on below—to detect vote tally alterations in the 1988 Mexican election, showing that edited forms may relate to incumbent political networks. However, as Mexican democracy has consolidated over the past few decades, the ability of partisan actors to capture the election administration process

7. For example, Justice Njoki Ndung'u dissented to the ruling nullifying Kenya's 2017 presidential election on the grounds that her physical examination of over one thousand potentially-compromised forms contained in complaints revealed little evidence of systematic irregularities (Ndungu 2017). That said, parties to a dispute can theoretically provide anecdotal and ad hoc witness-based evidence for specific polling stations when available.

has waned significantly, and yet form-based irregularities remain common. Challú, Seira, and Simpser (2020) finds that around 40% of forms in Mexico's 2009, 2012, and 2015 elections contain inconsistent tallies, but no evidence implying that the inconsistencies benefited a particular political party. Rather, drawing on a massive survey of polling station workers, the authors find that worker education, workload, and the difficulty of the math itself predict the presence and magnitude of tally problems.

While both studies deepen scholarly understanding of these errors, tally problems are just one part of the broader verification problem. Other types of irregularities observed on forms can indicate that an election's integrity has been compromised. For example, a vote tally may appear legitimate and proper, but its credibility is significantly undermined if it is reported on unofficial stationery instead of the official document required by law. As the Supreme Court of Kenya ruling quoted above indicates, important information about electoral integrity is encoded in deficiencies with the form itself. Broadening our perspective beyond just the vote tally can give us insights into the electoral process at each polling station that would otherwise be lost to history.

Besides helping us better verify that official results match voters' true intentions, identifying patterns in form-based irregularities can also shed light on the underpinnings of perceptions of electoral management bodies' legitimacy (e.g., Kerr 2014; Lundmark, Oscarsson, and Weissenbilder 2020), in turn better illuminating the relationship between electoral integrity and democratic legitimacy (e.g., Mattes 2014). Understanding this relationship among irregularities, electoral integrity, and democratic legitimacy is increasingly important in the era of social media, which aids in the diffusion of problematic forms[8]—and exacerbates the spread of misinformation around them (Allcott and Gentzkow 2017). Further, our approach to studying electoral integrity by focusing on statutory forms complements research emphasizing the institutional design of electoral management bodies (van Ham and Garnett 2019). These institutionalist studies provide an important perspective on the structural and organizational constraints to electoral integrity; our focus on micro-level outputs allows us to characterize the distribution of irregularities across polling stations that may

8. For instance, in 2020, the Ugandan electoral commission published an example form which went viral after observers discovered signs that it was likely falsified (Ahimbisibwe 2021).

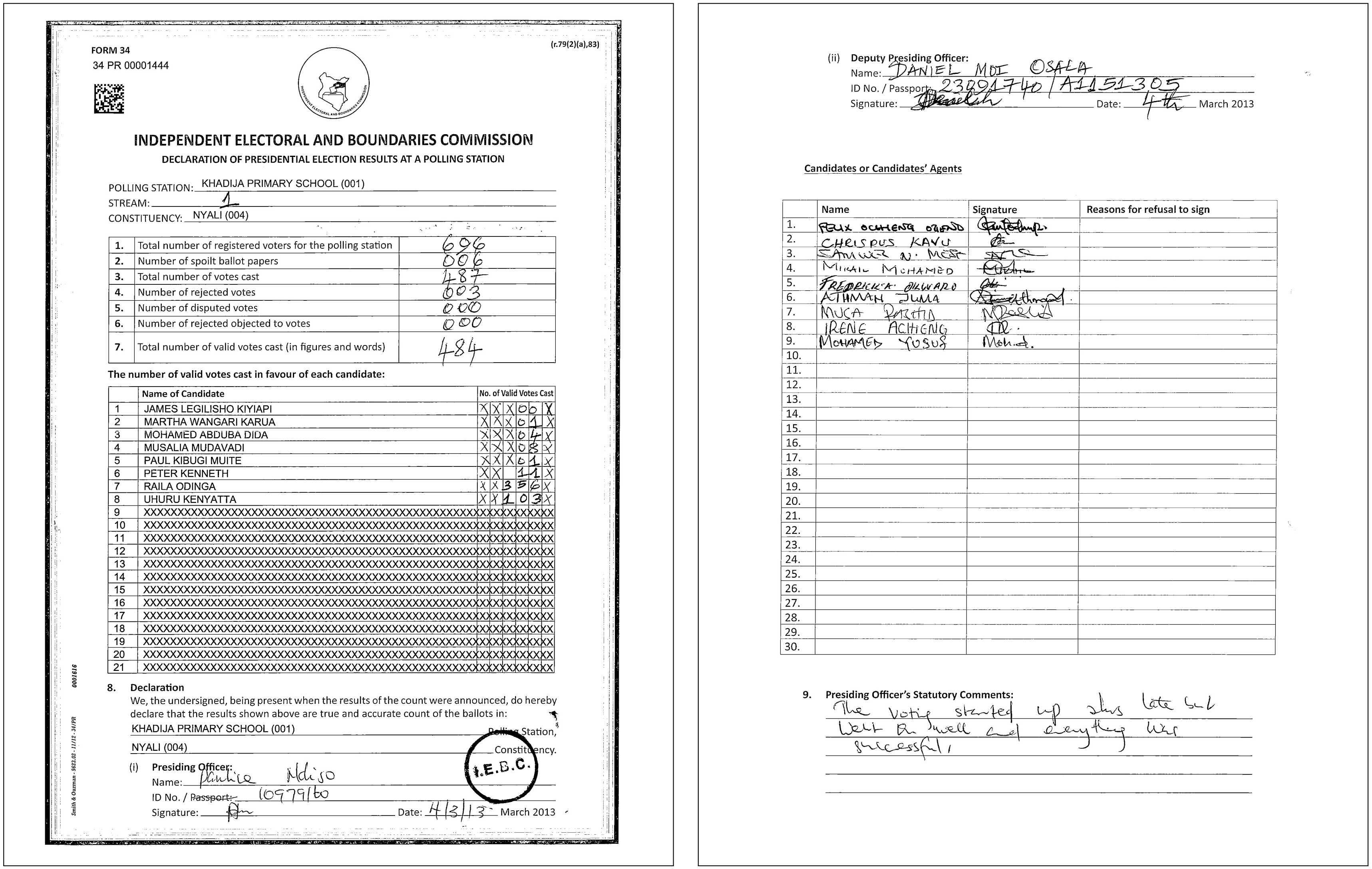
reflect those constraints. Finally, forms are produced by poll workers (James 2019; Neggers 2018). Our focus on form-based irregularities may provide some insight on the role of these workers in buttressing or eroding electoral integrity.

What do irregularities look like?

In the Kenyan model, an official form is used to report the final results at a polling station. During counting, certain individuals—usually some mix of observers, candidates' agents, security personnel, and the media—witness the process. A presiding officer records the results, stamps the document, signs it, and presents it to party agents for verification and signature. The form is then transmitted to constituency and national tallying centers where results are aggregated and announced. Figure 1 provides an example of a form 34A.

We consider four categories of irregularities that can be observed on these statutory forms. First, document problems relate to the representation of the physical paper record. Document problems can occur when the form is not properly scanned, is replaced with an improvised form, or when the document lacks basic security features such as a unique Quick Response (QR) Code. These irregularities may arise when presiding officers are confronted with election-day logistical challenges. For instance, if an official form has not been delivered in time, they may use loose-leaf paper without a QR code to report results. These irregularities could also reflect malfeasance, such as an attempt to replace an official form with a counterfeit.

Next, procedural problems include a failure to complete basic tasks prescribed by the electoral commission. We focus on two such tasks: whether the presiding officer signed the form, and whether the form was stamped with an official electoral commission stamp. Completing these tasks indicates some degree of competence on the presiding officer's part, and indirectly, that of the constituency's electoral officers' training on how to fill out the form. Problems with either procedure may suggest inattentiveness on the part of the presiding officer or bad training, or even fraud, as the actual form may have been replaced with a substitute that omits this important procedural information.


FORM 34

34 PR 00001444

(r.79(2)(a),83)

INDEPENDENT ELECTORAL AND BOUNDARIES COMMISSION

DECLARATION OF PRESIDENTIAL ELECTION RESULTS AT A POLLING STATION

POLLING STATION: KHADIJA PRIMARY SCHOOL (001)

STREAM: 1

CONSTITUENCY: NYALI (004)

1.	Total number of registered voters for the polling station	696
2.	Number of spoilt ballot papers	006
3.	Total number of votes cast	487
4.	Number of rejected votes	003
5.	Number of disputed votes	000
6.	Number of rejected objected to votes	000
7.	Total number of valid votes cast (in figures and words)	484

The number of valid votes cast in favour of each candidate:

	Name of Candidate	No. of Valid Votes Cast
1	JAMES LEGILISHO KIYIAPI	X X X 0 6 X
2	MARTHA WANGARI KARUA	X X X 0 1 X
3	MOHAMED ABDUBA DIDA	X X X 0 4 X
4	MUSALIA MUDAVADI	X X X 0 8 X
5	PAUL KIBUGI MUITE	X X X 0 1 X
6	PETER KENNETH	X X 1 1 X
7	RAILA ODINGA	X X 3 5 6 X
8	UHURU KENYATTA	X X 1 0 3 X
9	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	XXXXXX
10	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	XXXXXX
11	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	XXXXXX
12	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	XXXXXX
13	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	XXXXXX
14	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	XXXXXX
15	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	XXXXXX
16	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	XXXXXX
17	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	XXXXXX
18	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	XXXXXX
19	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	XXXXXX
20	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	XXXXXX
21	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	XXXXXX

8. **Declaration**

We, the undersigned, being present when the results of the count were announced, do hereby declare that the results shown above are true and accurate count of the ballots in:

KHADIJA PRIMARY SCHOOL (001) Polling Station,

NYALI (004) Constituency.

I.E.B.C.

(i) **Presiding Officer:**

Name: Dinhice Ndiso

ID No. / ~~Passport~~: 10979160

Signature: ______ Date: 4/3/13 March 2013

Smith & Ouzman - 5622.02 - 11/12 - 34/PR 0001616

(ii) **Deputy Presiding Officer:**

Name: DANIEL MOI OSALA

ID No. / Passport: 23091740 / A1151305

Signature: ______ Date: 4th March 2013

Candidates or Candidates' Agents

	Name	Signature	Reasons for refusal to sign
1.	FELIX OCHIENG ORENDO	[illegible]	
2.	CHRISPUS KAVU	[illegible]	
3.	SAMUEL N. MCET	[illegible]	
4.	MIKAIL MOHAMED	[illegible]	
5.	FREDRICK A. OKWARO	[illegible]	
6.	ATHMAN JUMA	[illegible]	
7.	MUGA PATTIA	[illegible]	
8.	IRENE ACHIENG	[illegible]	
9.	MOHAMED YUSUF	Moh.	
10.			
11.			
12.			
13.			
14.			
15.			
16.			
17.			
18.			
19.			
20.			
21.			
22.			
23.			
24.			
25.			
26.			
27.			
28.			
29.			
30.			

9. **Presiding Officer's Statutory Comments:**

The voting started up this late but went on well and everything was successful.

Figure 1: An example statutory result form 34A.

Third are agent-related problems. Party representatives are allowed to monitor polling stations, observing the vote count and signing off on the results. Their presence or absence may affect vote counting (e.g., Ascencio and Rueda 2019). By watching the count, an agent can ensure that ballots for her candidate are correctly counted, and that problematic ballots for other candidates are excluded from the final tally. In our analyses, we observe an agent-related problem if no agents are present; if any agent does not or refuses to sign the form; if *all* agents fail to sign; or if all of the agents' signatures appears to be identical.

Finally, we track whether or not the numeric vote tally has been altered or otherwise manually edited. Such irregularities can indicate simple arithmetic mistakes that arise during the counting and recording process (Challú, Seira, and Simpser 2020). On the other hand, they can also represent direct evidence of electoral manipulation (Callen and Long 2015; Cantú 2019), particularly when the edits make significant changes to candidates' vote shares or are concentrated in particular types of polling stations.

Identifying irregularities using deep learning

Manually identifying which forms contain irregularities is prohibitively costly. At scale, it is also unreliable, with coders interpreting rules differently or making errors (Anastasopoulos and Bertelli 2020). Following recent advances in the election fraud literature, we instead classify these documents using machine learning.

Like Cantú (2019), our modeling strategy relies on deep neural networks (more specifically, convolutional neural networks; Webb Williams, Casas, and Wilkerson 2020). Our models use information in each image of a statutory form to classify it as either containing an irregularity or not. Standard machine learning models in political science typically rely on classifiers in which the input features (i.e., covariates) are known *ex ante* (Cohen and Warner 2021; Hill and Jones 2014; Hindman 2015). These variables are usually chosen from theory or contextual knowledge—for example, democracy and shared borders are used to predict international conflict in Beck, King, and Zeng (2000). In contrast, deep neural networks attempt to classify observations using input features which are not known, but rather are learned by the model itself. This property makes deep neural networks well-suited to tasks where the model must be insensitive to irrelevant variation (LeCun, Bengio, and Hinton 2015): these models' complexity allows them to be finely tuned to identify which features are important for classification, and which are just noise (LeCun et al. 1998; Razavian et al. 2014).

Figure 2 provides a visual representation of one of the models we use. Each form is fed through a series of "filters" in the "feature extraction" step, during which the model learns the abstract shapes (e.g., lines and curves) that are useful for distinguishing irregularities. The model then predicts whether an irregularity is present on the form, using the features learned in the first step.[9]

More specifically, we first drew a stratified random sample of 3,000 forms. We paid two research assistants to independently code each form in the training set for each of the nine irregularities, and then a more experienced third research assistant to resolve all coding disputes and conduct random spot-checks to ensure data quality.[10] 2,000 of these images comprise the training sample, while 500 images were used at the end of each training iteration to measure whether the iteration improved predictive performance. Using this common sample, we fit 60 different models (separately) using different parameter values (e.g., the size of each input image) for each of our nine irregularities. After all 540 models were fit, we computed performance metrics on a true test sample composed of the remaining 500 images which were entirely held out from model training (Neunhoeffer and Sternberg 2019). We then used the best model for each irregularity to predict out on the remaining 27,000 forms (again, separately) which were not manually coded. All models were estimated using the keras module in Python via slurm arrays on our University's high-performance computing cluster (Chollet 2015).[11]

Although this setup is standard in deep learning, we build on recent scholarship (Torres and Cantú 2021) by implementing four improvements over the current standard practice in political

9. See Torres and Cantú (2021) for a recent overview of the technical details of a convolutional neural network.

10. See the online appendix for the instructions provided to coders and examples of irregularities. Disagreement among coders were found in just 3% of all coded values.

11. See the online appendix for further technical details.

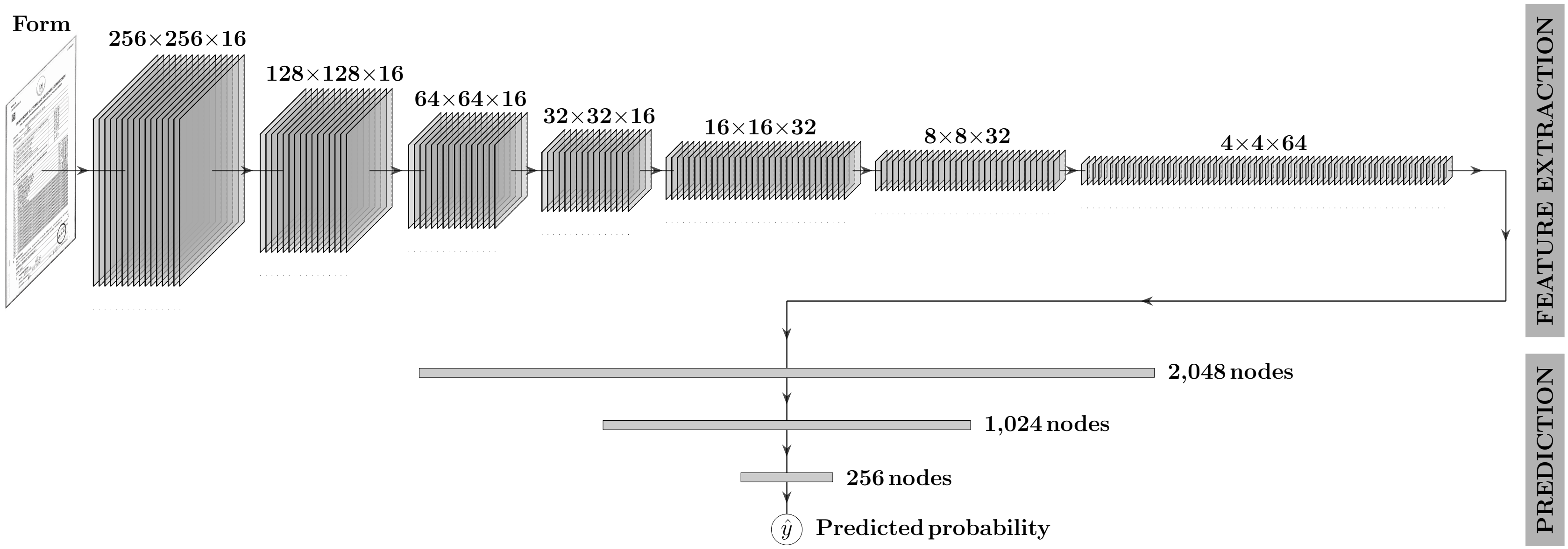


Figure 2: The structure of one of our deep neural networks. In this model, there are seven "blocks." Above each block is the dimension of the image being passed in as well as the number of filters in that block; for example, in the first block, there are 16 filters moving across a 256×256 pixel image. Together, these blocks comprise the feature extraction step of the model. The prediction step of the model begins when the output of the final block is passed to three "fully-connected layers," each with decreasing numbers of nodes (approximately equivalent to covariates in a generalized linear model). Finally, the output of these layers is passed to a sigmoid activation function which returns the predicted probability of the image containing the irregularity of interest.

science. First, we do not pre-select our training sample to achieve balance between irregularities and non-irregularities. This approach is prohibitively costly because we study nine different irregularities; to ensure balance for each, nine different training samples would have to be obtained. Instead, we tune our models using loss instead of accuracy, and select the best-performing models using an imbalance-sensitive metric (see below), again instead of accuracy. These changes de-emphasize predicting the right class (i.e., whether an irregularity is present or not) and elevates better performance over the $[0, 1]$ probability interval. With class-imbalanced data, accuracy is often maximized by just predicting 0 for every form, whereas loss heavily penalizes this behavior. Second, we introduce transfer learning, wherein we start with a model that has been pre-trained on a popular, publicly-available corpus of images (Tan et al. 2018). We need only then feed our images through this pre-trained model, allowing it to update to fit our data.[12] Although the gains to using transfer learning are most dramatic where the public corpus is most similar to the the target data (here, the scanned forms), we still see moderate performance improvements for the more difficult-to-classify irregularities, reducing erroneous predictions by as much as 40%. Third, we use data augmentation to synthetically increase our training sample size (Wong et al. 2016). Last, by tuning over a grid of 60 different parameter settings, we ensure maximum predictive performance.

Table 1 summarizes the performance of the nine best models. Because each of the irregularities we study are rare—ranging from 32% of observations to less than 1%—common metrics such as classification accuracy are inappropriate for these data. For instance, since 99.3% of forms were signed by the presiding officer, we could achieve accuracy of 99.3% by simply predicting that all forms were signed. We instead report F_1 scores, which represent the harmonic mean between precision and recall. Precision is defined as the proportion of cases predicted to have an irregularity that actually did, while recall is the proportion of irregularities that were correctly identified.[13] F_1 scores range between zero and one, with higher values reflecting better performance.

12. We select the InceptionV3 model architecture (Szegedy et al. 2016), trained on the ImageNet database (Deng et al. 2009), due to its performance with documents like our forms.

13. In other words, precision is the proportion $\frac{\text{true positives}}{\text{true positives + false positives}}$, while recall is given by $\frac{\text{true positives}}{\text{true positives + false negatives}}$.

Table 1: Image classifier performance

Irregularity	Prop. minority	F_1
Document problems		
QR code missing	0.009	0.998
Poor scan quality	0.002	1.000
Procedure problems		
Form not stamped	0.257	0.988
Presiding officer did not sign	0.007	0.997
Agent problems		
No agents listed	0.040	0.997
Any agent did not sign	0.077	0.989
No agents signed	0.049	1.000
Agent signatures appear identical	0.050	0.998
Results edited	0.316	0.928

Prop. minority indicates the proportion of observations in the held-out sample belonging to the less-common class (e.g., 34% of results were edited while 66% were not). F_1 is the harmonic mean of precision and recall among a common true test sample of 500 polling stations, ranging over [0, 1] where larger values indicate better performance.

As the Table indicates, our classification models perform exceptionally well. Two types of irregularities, scan quality and "no agents signed," are classified perfectly, with every single observation among the test sample correctly predicted. Four more models achieve F_1 scores of over 0.99, with just a handful of observations among the held-out sample misclassified. Further, as we report in the online appendix, the errors are relatively balanced, with no model routinely over- or under-predicting irregularities. Even the worst-performing model predicts extremely well, with an F_1 of 0.928 for edited results. By way of comparison, the model for edited results studied in Cantú (2019) achieves an F_1 of approximately 0.885.[14] This result corresponds to reducing the number of errors from 11 to 4.8 out of every 100 observations—a 56% reduction in the raw error rate. Taken together, these results suggest we can be very confident that the results obtained below accurately portray the true incidence of irregularities in Kenya's 2013 presidential election.[15]

14. While Cantú (2019) does not report these values, they can be approximately inferred from the confusion matrices

Irregularities, election outcomes, and observers

Our goal is to describe the empirical distribution of form-based irregularities and to shed light on the process by which these irregularities are generated. We proceed first by examining the relative frequencies of irregularities across government and opposition strongholds. To understand whether these patterns suggest a mechanism related to malfeasance or just human error, we then examine whether irregularities correlate with electoral outcomes. Last, to get further traction on the specific question of whether irregularities are produced by actors committing fraud, we use a plausibly exogenous intervention that we believe should deter fraud but not benign error: the random assignment of electoral observers to polling stations.

We define government and opposition strongholds as counties wherein at least 80% of the vote went to Kenyatta or Odinga, respectively.[16] To make the analysis more tractable, we group these irregularities into three sets of problems relating to the conceptual groups defined above; given its uniqueness and theoretical relevance, we also directly study whether results were edited. These groups are constructed as indicated in Table 1, with our three aggregate variables coded as a problem if any variable in that group is predicted to be irregular.[17]

Figure 3 provides estimates from models regressing irregularities on stronghold type, using mixed areas (counties where neither candidate has an 80% majority) as the baseline. All models include a series of controls including population density, ruggedness of terrain, ethnic fractionalization, geographic isolation, poverty, literacy rate, and night-time lights (a proxy for economic activity), as

provided. See the online appendix for details.

15. See the online appendix for a plot of the geographic incidence of each type of irregularity. Broadly, we find that there is more variation within counties than across them, with no notable patterns except more agent problems in very rural areas (likely due to agents not turning up at particularly remote polling stations).

16. Our results are also robust to using 75% or 85% thresholds. However, these vote-based thresholds are potentially endogenous to electoral manipulation itself. To guard against this risk, we also examine results from models where strongholds are defined by whether the countywide ethnic composition is 80% Kikuyu and Kalenjin (government) or Luo and Kamba (opposition). Since the ethnic composition of the electorate is fixed prior to the election via voter registration, this measure is effectively "pre-treatment." Our results are unchanged under any of these coding rules (see the online appendix).

17. Automatic clustering and dimension-reduction techniques (e.g., multiple correspondence analysis) did not reveal any systematic patterns in how these variables correlate, suggesting they are uncovering different types of statutory irregularities.

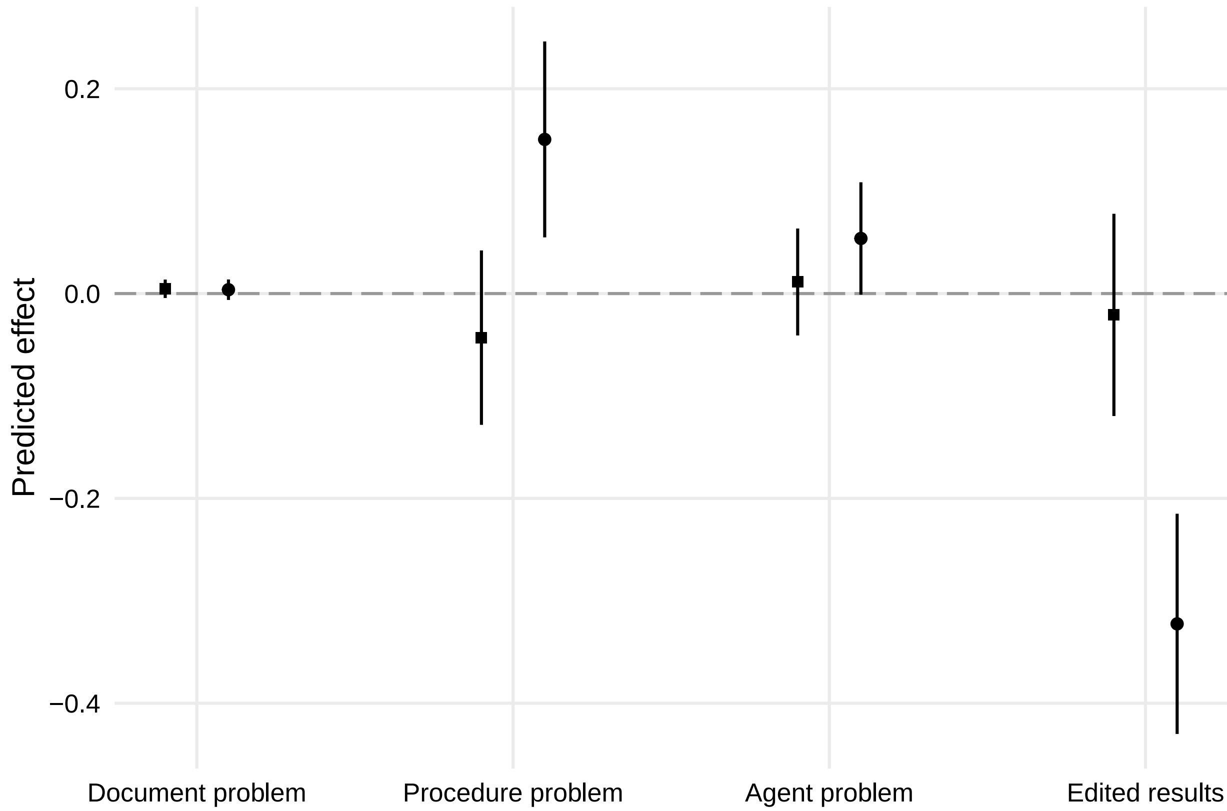


Figure 3: Frequency of electoral irregularities by polling station partisanship. Squares indicate government strongholds, with circles for opposition strongholds, relative to the baseline of competitive areas. Lines indicate 95% confidence intervals.

well as constituency fixed effects, with standard errors clustered by constituency.[18] Estimates for government strongholds are plotted as squares, with dots for opposition strongholds, and lines for 95% confidence intervals.

Polling stations in opposition strongholds are much more likely to experience procedure problems—presiding officers failing to sign or stamp forms—than those in government strongholds or in competitive counties. Since the baseline frequency of procedure problems in competitive areas is 0.20, an effect of 0.15 represents a 74% increase over the baseline likelihood of a procedure problem. Conversely, edited results are much less likely in opposition strongholds. While competitive areas are expected to have edited results at a frequency of 0.35, the expected frequency in opposition polling stations drops by 0.32—a 93% decrease. Both effects are statistically distinguishable not

18. See the online appendix for data sources, coding rules, and full regression results.

only from the baseline in competitive counties, but also from the frequency of irregularities in government strongholds. For both document and agent problems, neither government nor opposition strongholds are distinguishable from each other or from the competitive county baseline.

These results suggest that the data-generating process for procedure problems and edited results systematically differs in opposition strongholds from the rest of the country, even after parceling out local geographic, social, and economic variation. Two potential mechanisms are plausible: fraud and fumble. It is possible that opposition strongholds saw much more in-person voter fraud, such as individuals voting multiple times, opposition voters intimidating Kenyatta supporters, or party agents tampering with ballot materials. In such cases, presiding officers may not have certified the integrity of the vote, leading to more procedural problems; presumably, if these officers did not intend to certify the polling station results, then they would not care to make edits to fix errors or account for minor changes to totals induced by spoiled or rejected ballots. Alternatively, we can imagine that these patterns do not reflect fraud but rather human error. Since presiding officers are recruited and organized by parliamentary constituency, it is plausible that opposition polling stations had systematically less capable or less attentive presiding officers. Such officers would be less likely to follow basic protocols such as stamping and signing forms, and less likely to catch mistakes in vote tallies.

To help pick apart these potential mechanisms, we first regress three outcomes of the vote—turnout, Kenyatta's vote share, and the absolute margin of victory—on procedure problems and edited results (separately). All models include the same set of controls and fixed effects, and again cluster standard errors by constituency. Results from these models are in Table 2. The top panel presents results across all polling stations, while the middle and bottom panels provide estimates from models estimated just on government and opposition strongholds, respectively.

These estimates indicate that electoral outcomes do not differ meaningfully across polling stations with and without irregularities. This finding holds whether we look across the entire sample or restrict attention to either type of stronghold. Although a few estimates have statistically significant results, the effect sizes are very marginal. For instance, we find that turnout decreases by

Table 2: The relationship between irregularities and outcomes by partisanship

Irregularity	Turnout	Kenyatta vote share	Absolute margin
All polling stations			
Procedure problems	−0.00	−0.00	−0.00
Edited results	−0.00*	0.00	−0.00
Government strongholds			
Procedure problems	−0.00	−0.00	−0.01
Edited results	−0.01*	0.00	−0.00
Opposition strongholds			
Procedure problems	−0.00	−0.00	0.01
Edited results	−0.00	−0.00	0.00

$^{*}p < .05$. Each estimate is the effect of an irregularity on the dependent variable listed in each column. Stronghold definitions are provided in the text above. All models include controls and constituency fixed effects, with standard errors clustered by constituency. There are just under 30,000 observations for the "all polling station" analyses; 26% exhibit procedure problems and 32% have edited results. For government strongholds, these figures are just over 11,000, 34%, and 33% respectively; for opposition strongholds, they are over 7,000, 24%, and 29%.

0.004 across all polling stations where results are edited; since the mean turnout is 73.6 percent, this effect translates to just a 0.49% decrease from the baseline.

The absence of meaningful differences in electoral outcomes across these polling stations is suggestive evidence against the fraud hypothesis: if irregularities in opposition strongholds were produced by attempts to rig the outcome, then we would expect to see higher turnout, fewer votes for Kenyatta, and a larger absolute vote margin. In fact, among the three statistically significant estimates, two are in the wrong direction, as edited results are associated with *lower* turnout in government strongholds, which holds if we study all polling stations together. Both effect sizes are again minuscule, reinforcing our interpretation that there are simply few differences between outcomes in strongholds and competitive areas.

The evidence suggests so far that opposition strongholds have differing rates of procedure problems and edited results not because of fraud but rather fumble. However, because these data are observational, it is possible that our findings are unreliable due to unknown confounds. To get a firmer grip on the question of what produces form-based irregularities, we turn to the presence

Table 3: Electoral observers' effect on irregularities

Irregularity	Gov. strongholds	Opp. strongholds	Competitive	All
Document problems	−0.00*	−0.00	0.00	−0.00
Procedure problems	0.01	0.02	0.02	0.02
Agent problems	−0.01	0.02	−0.01	−0.00
Edited results	−0.03	0.01	−0.02	−0.01
Any problem	−0.05	0.02	−0.00	−0.01

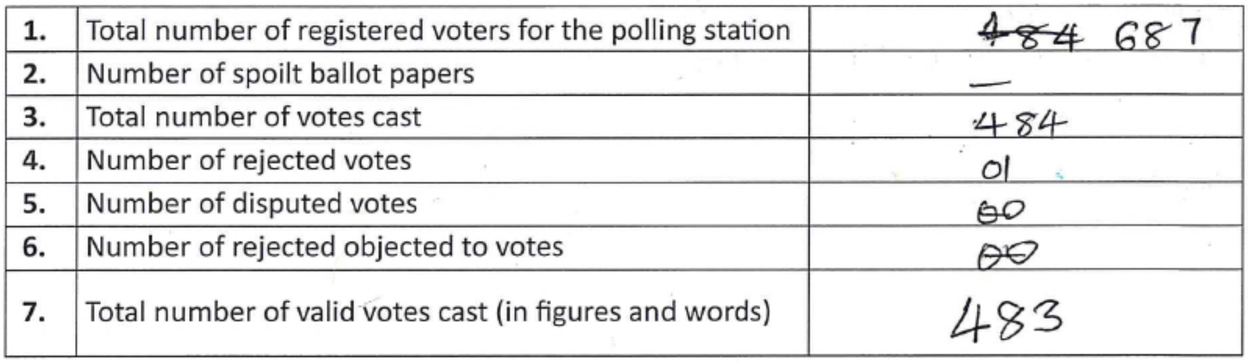
$^{*}p < .05$. Each estimate is the effect of election observation on the frequency of form-based irregularities, split by whether a county is a government or opposition stronghold. All models include controls and constituency fixed effects, with standard errors clustered by constituency. Because we cannot rule out the possibility of spillover effects of observation, these represent intention-to-treat (ITT) estimates.

of election observers. Since observers were assigned randomly to a nationally-representative sample (with oversamples of the three most populous cities), their presence provides an exogenous intervention.[19] We expect that observers should decrease fraud-based irregularities, but we do not expect that they impact irregularities that arise due to fumble. In other words, we expect that observers deter electoral malfeasance, but do little to prevent human error.

We regress each type of irregularity (as well as the presence of *any* irregularity) on the presence of an electoral observer. We include the same controls and fixed effects, continue to cluster standard errors by constituency, and add weights corresponding to the inverse probability of treatment for each polling station. Table 3 reports estimates for models run on each group of strongholds, competitive areas, and all polling stations simultaneously.

These results indicate that electoral observers do not have any effect on the occurrence of form-based irregularities. Nearly all estimates are statistically insignificant. While document problems seem to be less likely in observed polling stations in government strongholds, the effect size is small, and appears to be an artifact of the relative rareness of document problems: only 32 of the 11,517 polling stations in government strongholds had a document problem. Neither procedure

19. Observers were randomized using a randomized start procedure, selecting the 11th polling station on a master list of stations and then including every 32nd polling station after the 11th, leading to a 3% national sample of 976 polling stations. In addition, the implementing organization oversampled Mombasa, Nakuru, and Nairobi counties the main urban areas using the same process. This led to an additional 222 sampled stations in Nairobi; 274 in Mombasa; and 255 in Nakuru. In total, there were 1727 randomly sampled polling stations, across all 290 constituencies. As a result, there were no constituencies equivalent to the "pure control" we might see in a randomized saturation design (e.g., Asunka et al. 2019; Ichino and Schündeln 2012). We discuss this limitation further in the conclusion.



1.	Total number of registered voters for the polling station	~~484~~ 687
2.	Number of spoilt ballot papers	—
3.	Total number of votes cast	484
4.	Number of rejected votes	01
5.	Number of disputed votes	00
6.	Number of rejected objected to votes	00
7.	Total number of valid votes cast (in figures and words)	483

Figure 4: An example of an edited result. Here the presiding officer has incorrectly written the number of votes cast on the registered voter line before fixing the mistake.

problems nor edited results are affected by the presence of an exogenously-assigned observer.

Further evidence in favor of this "fumble" interpretation is available in the types of edits that are made to statutory forms. To study these more closely, we sampled 500 of the statutory forms manually coded as having been edited (during the construction of our training sample). We then had research assistants determine whether the original entries could be discerned, and where they could be, compared the original and final entries to understand the magnitude of these changes. The results are telling: the median change to registered voters, spoiled ballots, ballots cast, rejected ballots, disputed ballots, objections to rejected ballots, and valid votes cast are all exactly zero (with a global mean of 0.60). We then manually inspected every form where the change in any of these tallies was greater than 10, of which there were only 26. In almost every case, it was clear that the error was one of simply writing the entry on the wrong line. Figure 4 provides a characteristic example wherein the presiding officer wrote the number of votes cast in the registered voters line, crossed it out, and then filled in the information correctly. Such edits bear a much stronger resemblance to benign mistakes than they do to malfeasance.

Taken together, our results suggest that meaningful differences between government strongholds, opposition strongholds, and competitive areas do exist. And while distinguishing fraud from fumble is difficult, the absence of strong relationships between irregularities, outcomes, and observers suggests that the culprit is less likely to reflect deliberate electoral malfeasance than it is a simple lack of capacity.[20] To be sure, our results only speak to what can be detected on statutory forms, and it is possible that low administrative capacity in opposition strongholds is itself the product of larger-scale malfeasance, such as the government denying them qualified candidates for presiding officers. But the irregularities visible on the forms themselves do not directly reflect electoral fraud.

Studying irregularities in developing democracies

Around the world, election results are manually compiled with pen and paper. While irregularities on these forms are to some extent inevitable, they may also be an indicator of electoral fraud. In order to distinguish the two, scholars and observers need to know what irregularities look like and where they occur; characterizing the entire distribution of irregularities is therefore an important step in determining whether the will of the voters as expressed at the ballot box has been communicated in official election results. Yet the most prominent research on these questions has focused narrowly on vote tallies. There is too much information in the broader array of form-based irregularities to leave on the table, particularly for developing democracies where electoral integrity and legitimacy remain hotly contested.

This paper is a first attempt to characterize form-based irregularities broadly using state of the art computer vision tools. We identify irregularities relating to not just vote tally editing but also to document quality, procedural missteps, and agent participation, using data from virtually every polling station in Kenya's 2013 presidential election. We demonstrate that procedural problems are much more common, and edited results much less common, in opposition-controlled polling stations. But we also show that these patterns do not correlate to any meaningful differences in electoral outcomes such as turnout, vote share, or vote margin. Further, we show that randomly-assigned election observers do not impact the distribution of irregularities. Together, these results indicate that the widespread form-based irregularities in this election were like due to benign human error

20. In the appendix, we present additional quasi-experimental evidence of the errors-as-fumble interpretation. Using the regression discontinuity strategy in Harris (2021), we find that smaller polling stations (each with about 400 voters) within a single polling center have significantly lower rates of edited results than single polling stations with just under 800 voters. This finding suggests that decreasing the workload of presiding officer may decrease form-related problems.

rather than systematic fraud.

While our findings shed new light on form-based electoral irregularities, the limitations of our findings frame several unanswered questions for future research. We are bound by the constraints of just one election in one country. We have therefore have only identified irregularities for Kenya's 2013 presidential contest; we do not claim here to have developed an exhaustive list of all possible form-based irregularities. Indeed, given variations in form design and election type, each context will likely have its own challenges with respect to form irregularities. Future scholarship can build on this work by identifying other ways that problematic forms might threaten electoral integrity. For example, in other contexts, forms might include elements like fingerprints or bar codes; identifying irregularities with these may help recover information about the veracity of the vote tallies.

Further, the divergent patterns in form-based irregularities across government and opposition strongholds deserves more thorough explanation. Since polling station workers are recruited locally in Kenya, it may be that the quality of workers varies significantly across polling stations due to demographic characteristics that mirror partisanship—for example, Odinga-supporting areas may simply have fewer qualified or less well-educated candidates. Alternatively, these differences may arise due to other forms of partisan electioneering: it is also possible that systematic differences in polling station worker capacity arise due to structural factors imposed by the outgoing government, such as providing greater funding and staffing for their own strongholds than opposition areas.[21] Future scholarship should take a broader view of electoral administration to understand the origins of the divergent distributions of irregularities even within a single election.

Future work might also consider whether and how form-related irregularities might be interrelated across polling stations. The election observation design employed in 2013 was optimized for a parallel vote tabulation due to lingering concerns about aggregation fraud in previous Kenyan elections; this design precludes spillover analyses like those conducted in Asunka et al. (2019) and Ichino and Schündeln (2012). The direct effects we estimate could be hypothetically hiding stable unit treatment value assumption violations. We view this hypothetical as unlikely because the vast

21. For instance, see Harris (2021) for a potential explanation: policies for the creation of new polling stations may favor particular areas of the country, reducing workloads for poll workers.

majority of these estimates are null and or very small. Further, Kenyan political parties are much less well-organized than those in contexts where spillovers have been found (e.g., Ghana; Otele and Etyang 2015; Wahman 2012), making their large-scale coordination less likely. And since the process of counting votes and producing Forms 34A is comparatively quite public, systematic local fraud would require the broad complicity of a wide range of public and private observers. For all of these reasons, we view the possibility of observers generating spillover effects in form-based irregularities in Kenya as relatively remote—but nonetheless an open question.

Similarly, scholars and democracy assistance practitioners may also wish to examine the relationship between form-based irregularities and alternative manifestations of electoral fraud. This paper has focused narrowly on the question of irregularities on election results forms, to the exclusion of many different kinds of fraud that may be available to political actors' menu of manipulation (Schedler 2002; Harvey 2016). A lack of clear evidence of fraud in one part of the electoral process does not imply that the entire election writ large was free and fair. (Indeed, although our results suggest no systematic relationship between irregularities and outcomes, we cannot absolutely rule out the possibility that some individual forms were manipulated with the intent to defraud.) Just as we cannot rule out the possibility that election observation may have generate geographic spillovers, we cannot rule out the possibility that observation may have generated spillovers in *how* results were manipulated—a well-documented finding in other contexts (e.g., Callen and Long 2015; Friesen 2019). Greater scrutiny of these relationships—including through methods like those we develop here—may help improve the production of election results and, in the long run, build trust in the electoral process.

From a practical perspective, our approach can be repackaged to assist election monitors trying to evaluate electoral integrity after governments publish official form-based results. Individuals and organizations can quickly train models to identify specific classes of irregularities, perhaps even using our pretrained models as a transfer learning baseline. This approach would cut the time devoted to identifying irregularities across an entire country from months to days, vastly improving the speed by which organizations could report their assessments of the accuracy of the official results.

As computing power increases and costs decrease, these gains will become even more apparent.

Deploying large-scale machine learning methods for election monitoring is not without its challenges. Using deep learning can save organizations from manually examining tens of thousands of results forms, but hand-coding thousands of forms for a training sample may be similarly costly and time-intensive. Practitioners may not have access to the computational resources required to estimate these types of models. And perhaps most problematically, election evaluation requires broad-based public trust to be effective, yet machine learning methods such as these suffer from a "black box problem" of inscrutability (Zednik 2021). Effectively utilizing these methods to improve elections in developing democracies will require extensive efforts to communicate the intuitive nature of prediction and situate these methods in the broader movement toward such designs across the social sciences (e.g., Yarkoni and Westfall 2017; Meacham et al. 2019; Mullainathan and Spiess 2017).[22] Through deep engagement with practitioners and voters, scholars can turn our approach into a robust method for building trust in electoral integrity and legitimacy in developing democracies.

22. See also Gille, Jobin, and Ienca (2020) on how scholars can build stakeholder trust in machine learning methods.

References

ACE Project. 2021. ACE Election Database. Available at https://aceproject.org/epic-en. Last accessed 16 April 2021.

Ahimbisibwe, Patience. 2021. EC Admits Error in Poll Results Declaration Form. *Daily Monitor*. 2 March 2021. Available at https://www.monitor.co.ug.

Allcott, Hunt, and Matthew Gentzkow. 2017. "Social Media and Fake News in the 2016 Election." *Journal of Economic Perspectives* 31 (2): 211–236.

Anastasopoulos, L. Jason, and Anthony M. Bertelli. 2020. "Understanding Delegation through Machine Learning: A Method and Application to the European Union." *American Political Science Review* 114 (1): 291–301.

Ansolabehere, Stephen, Barry C. Burden, Kenneth R. Mayer, and Charles Stewart III. 2018. "Learning from Recounts." *Election Law Journal* 17 (2): 100–116.

Ascencio, Sergio J., and Miguel R. Rueda. 2019. "Partisan Poll Watchers and Electoral Manipulation." *The American Political Science Review* 113 (3): 727–742.

Asunka, Joseph, Sarah Brierley, Miriam Golden, Eric Kramon, and George Ofosu. 2019. "Electoral Fraud or Violence: The Effect of Observers on Party Manipulation Strategies." *British Journal of Political Science* 49 (1): 129–151.

Beck, Nathaniel, Gary King, and Langche Zeng. 2000. "Improving Quantitative Studies of International Conflict: A Conjecture." *American Political Science Review* 94 (1): 21–35.

Bengio, Yoshua. 2012. "Deep Learning of Representations For Unsupervised and Transfer Learning." In *Proceedings of ICML Workshop on Unsupervised and Transfer Learning,* 17–36.

Burden, Barry C., and Jeffrey Milyo. 2015. "The Quantities and Qualities of Poll Workers." *Election Law Journal* 14 (1): 38–46.

Callen, Michael, and James D. Long. 2015. "Institutional Corruption and Election Fraud: Evidence from a Field Experiment in Afghanistan." *American Economic Review* 105 (1): 354–381.

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

Caruana, Rich. 1994. "Learning Many Related Tasks at the Same Time With Backpropagation." In *Proceedings of the 7th International Conference on Neural Information Processing Systems,* 657–664.

Challú, Cristian, Enrique Seira, and Alberto Simpser. 2020. "The Quality of Vote Tallies: Causes and Consequences." *American Political Science Review* 114 (4): 1071–1085.

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

Cohen, Mollie J., and Zach Warner. 2021. "How to Get Better Survey Data More Efficiently." *Political Analysis* 29 (2): 121–138.

Daxecker, Ursula, Jessica Di Salvatore, and Andrea Ruggeri. 2019. "Fraud Is What People Make of It: Election Fraud, Perceived Fraud, and Protesting in Nigeria." *Journal of Conflict Resolution* 63 (9): 2098–2127.

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.

Erlich, Aaron, and Nicholas Kerr. 2016. "'The Local Mwananchi Has Lost Trust': Design, Transition and Legitimacy in Kenyan Election Management." *The Journal of Modern African Studies* 54 (4): 671–702.

Friesen, Paul. 2019. "Strategic Ballot Removal: An Unexplored Form of Electoral Manipulation in Hybrid Regimes." *Democratization* 26 (4): 709–729.

Garnett, Holly Ann. 2019. "Evaluating Electoral Management Body Capacity." *International Political Science Review* 40 (3): 335–353.

Gathii, James Thuo, and Olabisi D. Akinkugbe. 2020. "Judicial Nullification of Presidential Elections in Africa: Peter Mutharika v Lazarus Chakera and Saulos Chilima in Context." *SSRN Electronic Journal.*

Gettleman, Jeffrey. 2013. Voting Irregularities in Kenya Election Are Confirmed, Adding Fuel to Dispute. *New York Times*. 29 March 2013. Available at https://www.nytimes.com.

Gille, Felix, Anna Jobin, and Marcello Ienca. 2020. "What We Talk About When We Talk About Trust: Theory of Trust for AI in Healthcare." *Intelligence-Based Medicine* 1-2:100001.

Goggin, Stephen N., Michael D. Byrne, and Juan E. Gilbert. 2012. "Post-Election Auditing: Effects of Procedure and Ballot Type on Manual Counting Accuracy, Efficiency, and Auditor Satisfaction and Confidence." *Election Law Journal: Rules, Politics, and Policy* 11 (1): 36–51.

Hall, Thad E., J. Quin Monson, and Kelly D. Patterson. 2009. "The Human Dimension of Elections: How Poll Workers Shape Public Confidence in Elections." *Political Research Quarterly* 62 (3): 507–522.

Harris, J. Andrew. 2021. "Election Administration, Resource Allocation, and Turnout: Evidence from Kenya." *Comparative Political Studies* 54 (3-4): 623–651.

Harvey, Cole J. 2016. "Changes in the Menu of Manipulation: Electoral Fraud, Ballot Stuffing, and Voter Pressure in the 2011 Russian Election." *Electoral Studies* 41:105–117.

Hill, Daniel W., Jr., and Zachary M. Jones. 2014. "An Empirical Evaluation of Explanations for State Repression." *American Political Science Review* 108 (3): 661–687.

Hindman, Matthew. 2015. "Building Better Models: Prediction, Replication, and Machine Learning in the Social Sciences." *The Annals of the American Academy of Political and Social Science* 659 (1): 48–62.

Ichino, Nahomi, and Matthias Schündeln. 2012. "Deterring or Displacing Electoral Irregularities? Spillover Effects of Observers in a Randomized Field Experiment in Ghana." *The Journal of Politics* 74 (1): 292–307.

Idrobo, Nicolás, Dorothy Kronick, and Francisco Rodríguez. 2020. "Do Shifts in Late-Counted Votes Signal Fraud? Evidence from Bolivia." *SSRN Electronic Journal.*

James, Toby S. 2019. "Better Workers, Better Elections? Electoral Management Body Workforces and Electoral Integrity Worldwide." *International Political Science Review* 40 (3): 370–390.

Jegwa, Peter. 2019. Malawi Election: Court Orders New Vote after May 2019 Result Annulled. *BBC News*. 3 February 2020. Avaialble at https://www.bbc.co.uk/.

Kerr, Nicholas. 2014. "EMB Performance and Perceptions of Electoral Integrity in Africa." In *Advancing Electoral Integrity,* edited by Pippa Norris, Richard W. Frank, and Ferran Martnez i Coma, 189–210. Oxford: Oxford University Press.

Kerr, Nicholas, and Anna Lührmann. 2017. "Public Trust in Manipulated Elections: The Role of Election Administration and Media Freedom." *Electoral Studies* 50:50–67.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. 2012. "ImageNet Classification with Deep Convolutional Neural Networks." *Advances in Neural Information Processing Systems* 25 (2).

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. 2015. "Deep Learning." *Nature* 521 (7553): 436–444.

LeCun, Yann, Léon Bottou, Yoshua Bengio, and Patrick Haffner. 1998. "Gradient-Based Learning Applied to Document Recognition." *Proceedings of the IEEE* 86 (11): 2278–2324.

Lundmark, Sebastian, Henrik Oscarsson, and Marcus Weissenbilder. 2020. "Confidence in an Election Authority and Satisfaction with Democracy: Evidence from a Quasi-Natural Experiment of a Failed Election in Sweden." *Electoral Studies* 67:102216.

Mattes, Robert. 2014. "Electoral Integrity and Democratic Legitimacy in Africa." In *Advancing Electoral Integrity,* edited by Pippa Norris, Richard W. Frank, and Ferran Martnez i Coma, 211–228. Oxford: Oxford University Press.

Meacham, Sofia, Georgia Isaac, Detlef Nauck, and Botond Virginas. 2019. "Towards Explainable AI: Design and Development for Explanation of Machine Learning Predictions for a Patient Readmittance Medical Application." In *Intelligent Computing: Proceedings of the 2019 Computing Conference,* 939–955.

Mullainathan, Sendhil, and Jann Spiess. 2017. "Machine Learning: An Applied Econometric Approach." *Journal of Economic Perspectives* 31 (2): 87–106.

Ndungu, Njoki S. 2017. The Dissenting Judgement of Njoki S. Ndungu, SCJ Presidential Petition No. 1 of 2017.

Neggers, Yusuf. 2018. "Enfranchising Your Own? Experimental Evidence on Bureaucrat Diversity and Election Bias in India." *American Economic Review* 108 (6): 1288–1321.

Neunhoeffer, Marcel, and Sebastian Sternberg. 2019. "How Cross-Validation Can Go Wrong and What to Do About It." *Political Analysis* 27 (1): 101–106.

Niemi, Richard G., and Paul S. Herrnson. 2003. "Beyond the Butterfly: The Complexity of U.S. Ballots." *Perspectives on Politics* 1 (2): 317–326.

Norris, Pippa. 2014. *Why Electoral Integrity Matters.* Cambridge: Cambridge University Press.

Norris, Pippa, Richard W. Frank, and Ferran Martnez i Coma. 2014. *Advancing Electoral Integrity.* Oxford: Oxford University Press.

Organization of American States. 2019. Electoral Integrity Analysis: General Elections in the Plurinational State of Bolivia Final Report.

Otele, Oscar M., and Oita Etyang. 2015. "Party Institutionalization in Africa: Kenya's 2013 Elections in Comparative Perspective." *African Review* 42 (1): 29–57.

Owuor, Felix Odhiambo. 2008. "The 2007 General Elections in Kenya: Electoral Laws and Process." *Journal of African Elections* 7 (2): 113–123.

Power, Timothy J., and J. Timmons Roberts. 1995. "Compulsory Voting, Invalid Ballots, and Abstention in Brazil." *Political Research Quarterly* 48 (4): 795–826.

Razavian, Ali Sharif, Hossein Azizpour, Josephine Sullivan, and Stefan Carlsson. 2014. "CNN Features Off the Shelf: An Astounding Baseline for Recognition." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops,* 806–813.

Republic of Kenya. 2013. Deposition of Janet Ongera.

Schedler, Andreas. 2002. "Elections Without Democracy: The Menu of Manipulation." *Journal of Democracy* 13 (2): 36–50.

Shah, Seema. 2015. "Free and Fair? Citizens' Assessments of the 2013 General Election in Kenya." *Review of African Political Economy* 42 (143): 44–61.

Shilaho, Westen Kwatemba. 2013. "Old Wine in New Skins: Kenya's 2013 Elections and the Triumph of the *Ancien Régime*." *Journal of African Elections* 12 (3): 89–119.

Stiefbold, Rodney P. 1965. "The Significance of Void Ballots in West German Elections." *American Political Science Review* 59 (2): 391–407.

Supreme Court of Kenya. 2013. 2013 Presidential Election Judgment.

———. 2017. Presidential Petition 1 of 2017, Odinga vs. IEBC and 2 Others.

Supreme Court of Malawi. 2020. Constitutional Reference No. 1 of 2019.

Supreme Court of Zimbabwe. 2018. Chamisa vs. Mnangagwa and 24 others. CCZ 42/18.

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.

Tan, Chuanqi, Fuchun Sun, Tao Kong, Wenchang Zhang, Chao Yang, and Chunfang Liu. 2018. "A Survey on Deep Transfer Learning." In *Proceedings of the International Conference on Artificial Neural Networks,* 270–279.

Torres, Michelle, and Francisco Cantú. 2021. "Learning to See: Convolutional Neural Networks for the Analysis of Social Science Data." *Political Analysis.*

van Ham, Carolien, and Holly Ann Garnett. 2019. "Building Impartial Electoral Management? Institutional Design, Independence and Electoral Integrity." *International Political Science Review* 40 (3): 313–334.

Wahman, Michael. 2012. "Democratization and Electoral Turnovers in Sub-Saharan Africa and Beyond." *Democratization* 21 (2): 220–243.

Webb Williams, Nora, Andreu Casas, and John Wilkerson. 2020. *Images as Data for Social Science Research: An Introduction to Convolutional Neural Nets for Image Classification.* Cambridge Elements: Quantitative and Computational Methods for Social Science. Cambridge: Cambridge University Press.

Williams, Jack R., and John Curiel. 2020. Analysis of the 2019 Bolivia Election. Technical Report.

Wong, Sebastien C., Adam Gatt, Victor Stamatescu, and Mark D. McDonnell. 2016. "Understanding Data Augmentation for Classification: When to Warp?" In *Proceedings of the 2016 IEEE International Conference on Digital Image Computing: Techniques and Applications.*

Yarkoni, Tal, and Jacob Westfall. 2017. "Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning." *Perspectives on Psychological Science* 12 (6): 1100–1122.

Zednik, Carlos. 2021. "Solving the Black Box Problem: A Normative Framework for Explainable Artificial Intelligence." *Philosophy & Technology* 34 (2): 265–288.