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

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

July 14, 2020

Abstract

Confidence in election results is a central pillar of democracy, but in many developing countries, elections are plagued by a number of irregularities. Such problems can include incredible vote margins, sky-high turnout, and statutory forms that indicate manual edits. Recent scholarship has sought to identify these problems and use them to quantify the magnitude of election fraud in countries around the world. In this paper, we argue that this literature suffers from its reliance on an ideal election as a baseline case, and delineate an alternative data-generating process for the irregularities often seen in developing democracies: benign human error. Using computer vision and deep learning tools, we identify statutory irregularities for each of 30,000 polling stations in Kenya's 2013 presidential election. We show that these irregularities are uncorrelated with election outcomes and do not reflect systematic fraud. Our findings suggest that scholars of electoral integrity should take care to ensure that their methods are sensitive to context and account for the substantial challenges of administering elections in developing democracies.

*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.

The median turnout among all polling stations in Kenya's 2013 presidential election was 88%. The average margin between the top two candidates was 83%. When such figures coincided with credible reports of electoral maladministration, many observers inferred that these results were evidence of fraud instead of what President Uhuru Kenyatta referred to as just "one or two clerical errors" (Gettleman 2013). Sky-high turnout, extreme vote margins, and allegations of malfeasance are common in developing democracies like Kenya, leading many scholars to develop methods for detecting and quantifying election fraud (Alvarez, Hall, and Hyde 2008; Asunka et al. 2019; Beber and Scacco 2012; Cantú 2019; Enikolopov et al. 2013; Klimek et al. 2012; Myagkov, Ordeshook, and Shakin 2009; Rozenas 2017). Typically, these scholars' goal is to distinguish reliable election returns from fraudulent ones by analyzing anomalous results. But as Hanlon's Razor dictates,[1] there is a third possibility: results which are unreliable not because of fraud, but because of simple human error.

In this paper, we consider *electoral irregularities*, a broader and more common problem than electoral fraud. We define an irregularity as a deviation from a procedural rule or norm during an election. While electoral fraud is defined by an inferred intent to affect an election outcome, irregularities are the unavoidable result of humans working under pressure to produce a precise vote count.[2] They may arise for many reasons, including a poor work environment, insufficient training, or inadequate or missing election materials. They may or may not influence election results. And they are increasingly important. While motivated political agents frequently present irregularities as evidence suggestive of fraud, recent legal judgments have suggested that the incidence of irregularities alone may be sufficient grounds on which to nullify an election (Weinberg 2008). If common electoral irregularities can make or break democratic elections, characterizing these irregularities and examining their relationship to election outcomes takes on particular importance.

1 . Often cited as "do not attribute to malice what can be explained by incompetence," the idea was originally formulated by von Goethe (1774 [1957], 14) as "misunderstandings and lethargy perhaps produce more wrong in the world than deceit and malice do."

2 . Strictly speaking, fraud is a type of electoral irregularity that arises from an attempt to manipulate election results. However, for clarity, we use "irregularities" to refer to those irregularities that do not arise from fraud throughout the paper.

We study these irregularities in the context of Kenya's 2013 presidential election. Six years following an election marred by violence, the 2013 election saw 86% turnout among the electorate of over 14 million voters. Uhuru Kenyatta won 50.15% of the vote, just enough to clear the 50% threshold required to avoid a runoff. While the election was consistent with a parallel vote tabulation conducted by the Elections Observation Group (2013), opposition leaders alleged fraud. 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 more broadly.

We begin by gathering official election results as collated on paper at each polling station. Known as "Form 34A" in Kenya, these documents provide information such as vote counts by candidate, rejected votes, and the names and signatures of agents observing the count on behalf of the candidates. The presiding officer at each polling station completes each form manually before scanning it and digitally transmitting it to a national tallying center. These statutory forms provide the most granular data available on electoral process and outcomes across Kenya's approximately 30,000 polling stations.

These forms comprise a corpus that is too large for human coding, so we turn to computer vision and machine learning tools to identify irregularities on these forms. Specifically, we draw a stratified sample of 3,000 forms (nationally representative at the county level), which research assistants independently coded for irregularities. Building on Cantú (2019), we train deep neural networks to recognize irregularities among this training sample, which we then use to identify irregularities among the remaining forms. We examine the relationship between these irregularities and polling station-level outcomes such as turnout and vote share.

We show that irregularities among these statutory forms do not relate to election outcomes in a way that might reflect fraud. Irregularities are not concentrated in areas where one party or ethnic group is dominant. Nor is there any meaningful relationship between problems on statutory forms and turnout, the vote margin, or ballots which are rejected, disputed, or spoiled. We also examine the geographic and socioeconomic correlates of irregularities, finding that they occur in more urban, ethnically fractionalized constituencies, in contrast with expectations about fraud being conducted in ethnic enclaves or party strongholds (Gutiérrez-Romero 2014). Last, we demonstrate that the typical manual edit of a statutory form does not significantly impact final vote tallies. Taken together, our evidence paints a clear picture of an election rife with irregularities—but ones which appear to be unimportant for the eventual outcome. Though our design does not allow us to rule out fraud entirely, our findings indicate that election irregularities do not in themselves amount to fraud.

This paper contributes to the study of electoral integrity in four main ways. First, we demonstrate the widespread occurrence of irregularities in statutory forms as a phenomenon unto itself, and not as merely an artifact of systematic fraud. We conceptualize a series of irregularities as indicators merely of divergence from established norms and procedures, divorced from an underlying assumption of intentionality. While the gap between widespread irregularities and systematic fraud might seem slight, conceptual slippage between the two can have enormous consequences. For example, after Evo Morales narrowly won re-election in Bolivia's October 2019 election, the Organization of American States (2019, 9) issued a report that concluded that "a series of willful actions were taken to alter the results expressed at the poll," leading to Morales's resignation and exile. Recent research has determined that a substantial portion of the evidence for this determination resulted from a flawed statistical analysis of anomalous patterns in the election returns (Idrobo, Kronick, and Rodríguez 2020; Williams and Curiel 2020). In this case, mistaking irregularities for malfeasance may have removed a duly-elected head of state. Rather than protecting Bolivians' democratic rights, the search for evidence of electoral fraud may have undermined them.

Second, we broaden scholarly attention from a few key irregularities—abnormal patterns in vote tallies, vote shares, turnout, and edited results—to a number of other pieces of information that can be gleaned from statutory forms. We show how the presence or absence of features such as candidate agents' signatures or an official stamp can inform an analysis of an election's reliability.

Our third contribution is methodological: we build on Cantú (2019) by using deep learning to detect irregularities in statutory forms, introducing at least three notable improvements over the current state-of-the-art. To avoid overfitting, our models use random data augmentation in lieu of a pre-balanced sample, which is costly to obtain. Further, we implement transfer learning, a process in which we start with a publicly-available model which has been pre-trained on a comprehensive corpus of images before re-training it for our purposes (Caruana 1994; Bengio 2012; Razavian et al. 2014). Last, we tune our models' hyperparameters. All three innovations enhance out-of-sample predictive power. As a result, we are able to significantly improve our ability to detect irregularities, cutting the error rate reported in Cantú (2019) by more than one-third.

Finally, we show that, despite their prevalence, edited forms appear to have far less impact on election outcomes than previously assumed. This finding contrasts with assumptions discussed in the Organization of American States (2019) report on Bolivia, Kenyan opposition leaders' legal argument following their loss in the 2013 election (Supreme Court of Kenya 2013), and the evidence in Cantú (2019). While it is clear that edited results can reflect electoral malfeasance, our results indicate that they can also just be indicative of overtaxed and undersupported electoral administrators. Scholars should therefore take care when citing edited returns—even when they are widespread—as *prima facie* evidence of systemic fraud.

Separating fraud from irregularities

Electoral integrity is a growing concern globally (Norris 2014). Unsurprisingly, the majority of this literature in comparative politics focuses on "bad behavior"—election fraud, voter intimidation, and the like.[3] Improving electoral integrity depends on our ability to measure this bad behavior, so scholars have developed a range of methods for quantifying electoral fraud. Most prominent are tests of the distributions of vote shares, vote counts, and turnout rates to infer whether anomalous patterns amount to credible evidence of fraud (Alvarez, Hall, and Hyde 2008; Beber and Scacco 2012; Cantú 2014; Cantú and Saiegh 2011; Ferrari, McAlister, and Mebane 2018; Klimek et al. 2012; Mebane 2006; Mebane and Sekhon 2004; Medzihorsky 2015; Montgomery et al. 2015; Myagkov, Ordeshook, and Shakin 2009; Rozenas 2017). More recently, Cantú (2019) examines statutory forms directly, shifting focus from identifying fraud

3 . See Birch (2011) for a comprehensive introduction.

across a distribution of results to identifying it in the form of edited results at the polling station level.

However, there is a clear conceptual distinction between intentional, systematic fraud and unintentional, non-systematic fumble.[4] The expertise, training, logistics, and coordination required to successfully implement elections—particularly national elections—are anything but trivial. As a result, weaknesses at any point in the chain may lead to irregularities that are neither malicious nor substantively important with respect to the outcome. When fraud does occur, it does so against a backdrop of these inevitable irregularities. Of course, differentiating between fraud and relatively innocuous irregularities is difficult. This challenge is reflected in the lack of research seeking to describe and explain the prevalence and distribution of irregularities within which fraud might be nested.[5]

This blurring between error and fraud is pernicious for several reasons. First, without greater knowledge of when, where, and in what forms irregularities occur, our understanding of fraud is inherently limited. If we do not know what irregularities look like, then everything irregular will look like fraud, and we will conflate the two. Second, because irregularities are a function of the quality of election administration, and election administration is typically conducted by subnational governments, irregularities are likely to be geographically concentrated. For instance, states, counties, and other local governments are usually responsible for supplying election day presiding officers.[6] Variation in the quality of officials recruited, as well as the training provided to them, may shape when and where irregularities occur. This geographic concentration makes irregularities even easier to misjudge as evidence of fraud. Greater attention to basic contours of irregularities can help overcome these problems, improving our understanding of irregularities *per se* as well as of fraud.

Drawing a clear conceptual distinction between irregularities and fraud comes with another important benefit: it creates space for realistic electoral processes to

4 . Hyde (2008) provides an excellent discussion of the relationship between irregularities, fraud, and intent.

5 . One exception is Ansolabehere et al. (2018), which examines disparities between election night tallies and later recounts in two Wisconsin elections.

6 . In Kenya, parliamentary constituencies are responsible for providing these officers.

play out. In particular, decoupling irregularities from fraud blunts the ability of politically motivated parties to cast doubt on legitimate election results by focusing on irregularities which may or may not indicate manipulation. Such claims have been increasingly powerful in shaping countries' democratic trajectories, particularly in contexts with low trust and easy dissemination of evidence of irregularities via social media. Like in the Bolivian case discussed above, Kenya's 2017 election was annulled because it suffered from widespread irregularities, with only tenuous evidence to link these problems to systematic fraud (Cheeseman et al. 2019). More recently, Malawi's 2019 presidential election was annulled on the grounds that some statutory results had been edited using Tipp-Ex (a brand of white-out), lacked the signature of a presiding officer, or had small mathematical errors (Jegwa 2019). Given how common these problems are even when fraud is absent, treating them as fraud leaves little scope for building trust in democratic processes in most of the developing world.

Finally, because allegations of fraud correlate with increased post-election violence (Daxecker 2012), scholars have noted the need to tune their methods for studying fraud to be sensitive to false positives (Rozenas 2017). Yet incorrectly associating a series of irregularities to systematic fraud can similarly contribute to voters' insecurity. Shifting scholarly attention to the many processes through which benign irregularities can arise may help temper this risk of generating false positives that exacerbate electoral violence.

What do irregularities look like?

Enumerating every potential irregularity is far too broad a task for any one paper. Given the complexity of electoral administration, there are a great many mechanisms by which irregularities might arise. Further, because they are (by definition) unintentional, irregularities are at least to some extent a function of randomness. And because electoral processes vary so much across countries, irregularities which may be common in one context may be impossible in another and vice versa.

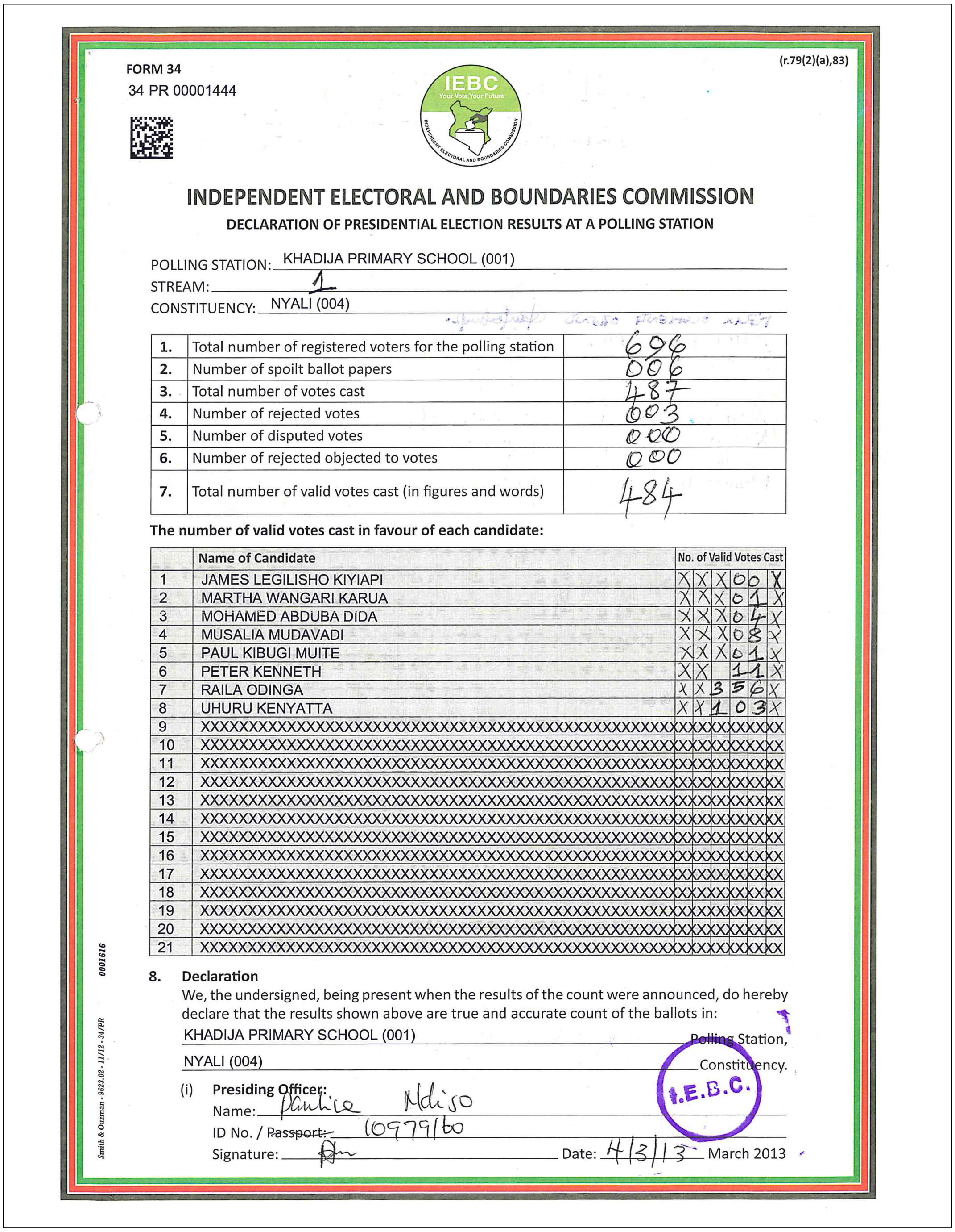
We confine ourselves to a much narrower task here: we examine the irregularities which can be observed on the basic statutory forms that constitute legal election results. Although our data are still limited to Kenya's 2013 presidential election, we identify problems which can be observed in many countries that follow Kenya's basic electoral process of producing election results.[7] 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 using a nationally-prescribed methodology, then stamps the document, signs it, and presents it to present agents for verification and signature. The form is then transmitted to a national tallying center where results are aggregated and announced. Figure 1 provides a complete example of one such form in our data.

We consider four categories of irregularities that can be observed on these statutory forms.[8] First are document problems, which relate to the representation of the physical paper record. Document problems can occur when the form is not properly scanned and transmitted for aggregation or when the document lacks security features such as a unique Quick Response (QR) Code. These irregularities may arise when presiding officers solve election-day difficulties. 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 may also reflect fraud, such as an attempt to replace an official form with a counterfeit.

Second are procedural problems, which indicate that basic tasks prescribed by the electoral commission may not have been completed. We focus on two such tasks due to their commonness among developing democracies: 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, bad training, or even fraud, as the actual form may have been replaced with a substitute that omits this important

7 . For instance, Bolivia, Malawi, and the Mexican election studied in Cantú (2019) all follow this broad outline.

8 . See the online appendix for more details and examples of these elements.



FORM 34 (r.79(2)(a),83)

34 PR 00001444

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	0	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	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	X	X	X	X	X	X
10	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	X	X	X	X	X	X
11	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	X	X	X	X	X	X
12	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	X	X	X	X	X	X
13	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	X	X	X	X	X	X
14	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	X	X	X	X	X	X
15	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	X	X	X	X	X	X
16	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	X	X	X	X	X	X
17	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	X	X	X	X	X	X
18	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	X	X	X	X	X	X
19	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	X	X	X	X	X	X
20	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	X	X	X	X	X	X
21	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	X	X	X	X	X	X

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) **Presiding Officer:**

Name: Pauline Ndiso

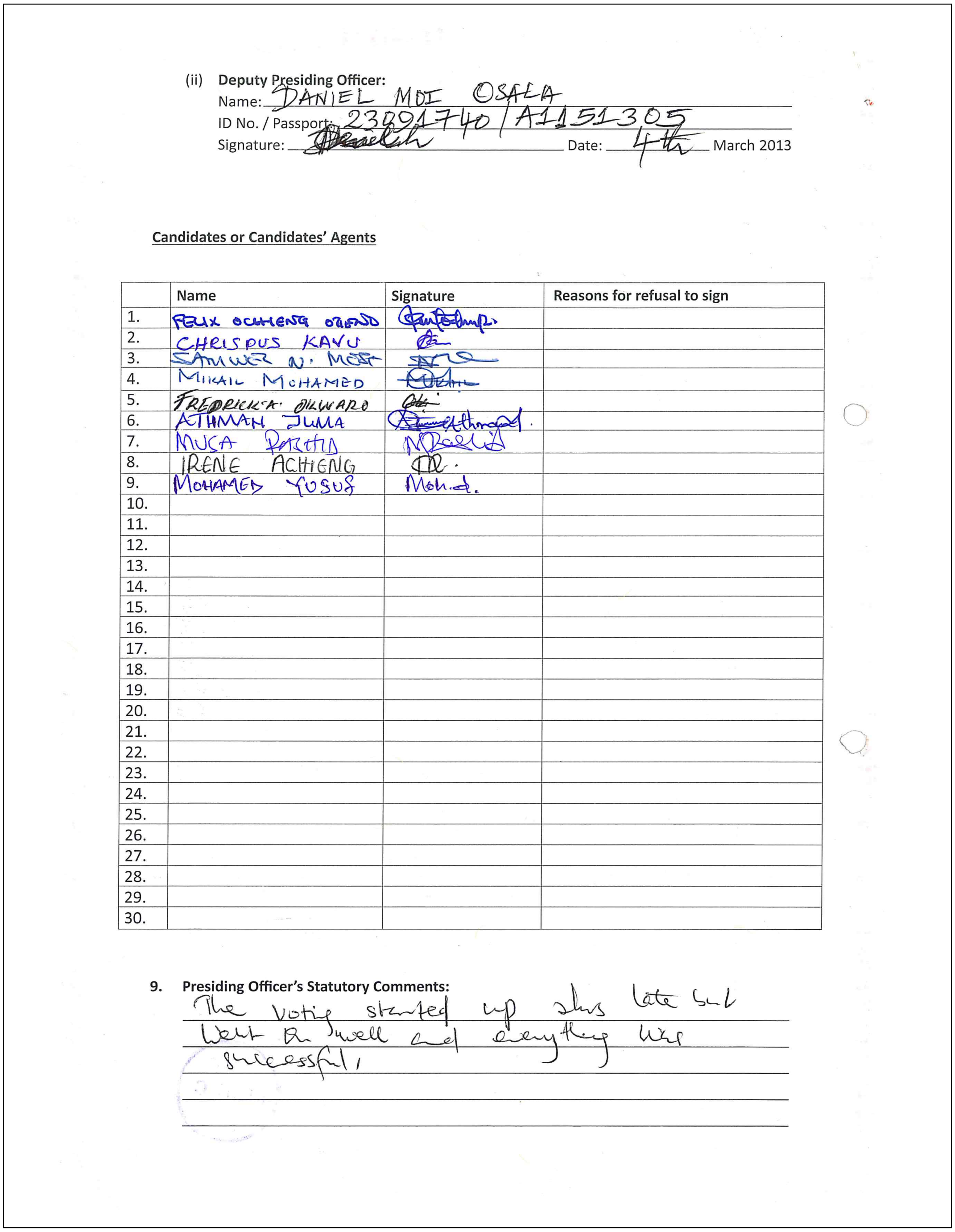
ID No. / ~~Passport:~~ 10979160

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

I.E.B.C.

Smith & Ouzman - 9823.02 - 11/12 - 34/PR 0001616

Figure 1: The first page of an example statutory result.



(ii) **Deputy Presiding Officer:**
Name: DANIEL MOI OSALA
ID No. / Passport: 23891740 / A1151305
Signature: _______________ Date: 4th March 2013

Candidates or Candidates' Agents

	Name	Signature	Reasons for refusal to sign
1.	FELIX OCHIENG ORENDO		
2.	CHRISPUS KAVU		
3.	SAMWEL N. MOSE		
4.	MIKAIL MOHAMED		
5.	FREDRICK A. OLWARO		
6.	ATHMAN JUMA		
7.	MUSA RASHID		
8.	IRENE ACHIENG		
9.	MOHAMED YUSUF		
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 slow late but went on well and everything was successful.

Figure 1 (continued): The second page of an example statutory result.

procedural information.

Third are agent-related problems. Ideally, individuals representing political candidates attend each polling station, observe the vote count, and sign off on the results. These agents' presence or absence can have an enormous impact on vote counting. 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. On the other hand, when only one agent is present, that agent may be able to induce the presiding officer to adjust totals to favor a given candidate; when no agents are present, a presiding officer may be more prone to adjust vote totals to benefit his preferred candidate. We examine five different variables related to agents: whether agents are present at the polling station, as indicated by any names listed; whether all agents listed signed the form; whether any agents refused to sign the form, as indicated in the space reserved for explaining refusals; whether any agent at all signed, capturing the minimal situation of at least one agent present; and whether the agent signatures appear different from one another, guarding against a presiding officer falsifying agents' presence.

The last type of irregularity we study refers to the integrity of the vote tally, focusing on whether these tallies have been manually edited. Such irregularities can indicate simple arithmetic mistakes that arise during the counting and recording process. On the other hand, they can also represent direct evidence of aggregation fraud (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.

Our conceptual goal in defining these types of irregularities is to more clearly distinguish the presence of an irregularity from the presence of fraud. This permits us to characterize the distribution of form-based irregularities independent from numeric election results. Separating irregularities in this way allows us to understand whether irregularities align with results in a way that suggests an intent to defraud.

Kenya's presidential election of 2013

The 2013 Kenyan presidential election provides a compelling case in which to study these irregularities. Among the eight candidates, opposition leader Raila Odinga and heir-apparent Uhuru Kenyatta went into election day in a virtual tie in the polls. To avoid a second-round runoff, one candidate would need an outright majority of votes nationally, as well as 25% of the vote in at least half of Kenya's 47 counties. After a relatively calm election day, the Independent Electoral and Boundaries Commission (IEBC) announced that Kenyatta had satisfied the majority requirement with 50.51% of the vote—less than ten thousand votes over the threshold, out of over twelve million cast. The announcement of such a narrow victory led to an immediate legal battle to annul the result, with opposition leaders arguing that significant irregularities had vitiated the entire election. Kenya's Supreme Court rejected this argument, and a month later Kenyatta assumed office.

The arguments and evidence presented in the case focused on a number of issues, including voter registration, election materials procurement, transmission of results, and vote tally aggregation. But central to both sides' arguments were the extent and meaning of widespread irregularities among the polling station forms. For 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" (Supreme Court of Kenya 2013, para. 161). For both sides, the fundamental focus was the legal import of irregularities with Form 34A—the focus of our study.

While the issues and arguments in this case may seem focused on byzantine regulations regarding obscure electoral procedures in Kenya, they speak to questions of much broader interest: whether and when an irregularity on the election results in their most basic state indicates fraud. These questions follow elections in many developing democracies. As cited above, elections in Kenya in 2017, Bolivia in 2019, and Malawi in 2019 all saw similar disputes over polling station returns, as did Zimbabwe's

2018 election. As the legal strategies of political aspirants place increasing scrutiny on the basic materials of elections, more such disputes will likely arise.

This case is also ideal because it allows us to compare our results against estimates of fraud using existing distributional methods. For instance, using ward-level data,[9] Ferrari, McAlister, and Mebane (2018) estimate that 29% of returns in this election were the result of "incremental fraud" wherein vote tallies are adjusted (as opposed to being wholly fabricated). Similarly, Mebane (2016) estimates that 35% of wards experienced incremental fraud, translating to approximately 750,000 tainted votes (or around 6% of the electorate). To develop a more complete picture of scholarly expectations around fraud in this election, we generate resampled kernel density, Bayesian finite mixture model, Benford's second-digit, and last-digit estimates of fraud using our polling station-level data (Beber and Scacco 2012; Ferrari, McAlister, and Mebane 2018; Mebane 2006; Rozenas 2017). Table 1 presents these results. Taken as a whole, these distribution-based approaches conclude that Kenya's 2013 election results were significantly fraudulent.

Identifying irregularities using deep learning

To study statutory forms directly, we must identify which forms contain irregularities. With some 30,000 forms, manual coding is prohibitively costly, so we instead classify these documents using computer vision and machine learning tools.

Supervised machine learning techniques require a "training set" of data in which the irregularities are manually coded. We built this training set using a stratified random sample of 3,000 documents. Two research assistants independently coded each form in the training set for each of the ten irregularities, marking them as `1` if an irregularity was present, `0` if no irregularity was present, and `NA` if the information could not be discerned—for instance, if the form has scanned incorrectly and it was impossible to

9 . There are 290 single-member district parliamentary constituencies in Kenya, grouped into 47 counties, which are devolved political units, each with a governor. Within each constituency are a number of county assembly wards, each electing a county-level representative. See the online appendix for a visualization of Kenya's internal political boundaries.

Table 1: Estimates of fraud in Kenya's 2013 presidential election

Method	Quantity	Statistic	Result
Resampled Kernel Density (RKD)	Vote shares	F	0.27*
Bayesian finite mixture model	Vote counts	F ("incremental")	0.05*
	Vote counts	F ("extreme")	0.14*
Benford second-digit test	Vote counts	χ^2	5,069.70*
	Turnout	χ^2	185.36*
Last-digit test	Vote counts	χ^2	414.66*
	Turnout	χ^2	6.44

$^*p < .05$. The "F" statistic refers to the proportion of polling station returns identified as fraudulent. All methods are implemented with respect to votes for Uhuru Kenyatta, the candidate for the incumbent party. Resampled kernel density estimation is implemented as described in Rozenas (2017), and produces a 95% credible interval of $[0.13, 0.45]$. The Bayesian finite mixture model is implemented in R using the eforensics package, as described in Ferrari, McAlister, and Mebane (2018) and Mebane (2016). This approach produces credible intervals of $[0.00, 0.10]$ and $[0.08, 0.22]$ for "incremental" and "extreme" fraud, respectively. For details on the second-digit and last-digit tests, see Mebane (2006) and Beber and Scacco (2012).

identify whether agents refused to sign.[10] We then hired a more experienced research assistant to conduct random spot-checks to ensure data quality, as well as to resolve all coding disputes.[11]

Like Cantú (2019), our modeling strategy relies on deep neural networks. These models use information in each image of a statutory form to classify it as either containing the irregularity of interest 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, forthcoming; Hill and Jones 2014; Hindman 2015). These variables are usually chosen from theory or contextual knowledge—for example, a series of geopolitical features 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

10 . See the online appendix for the instructions provided to coders and examples of irregularities.

11 . Irregularities were well-defined and identifiable, with disagreement among coders just 3% of all coded values.

model must be insensitive to irrelevant variation (LeCun, Bengio, and Hinton 2015): by stacking multiple layers of learning, such models can be finely tuned to identify which features are important for classification, and which are just noise.

In particular, we estimate convolutional neural networks (CNNs). CNNs are designed for data that are structured as multiple arrays bound together, as our data are, since each image of a statutory form is comprised of three color channels. In the first step, features such as lines and curves are extracted from the raw image by passing the data through convolutional filters, each of size 3×3 pixels, and then through pooling filters (of size 2×2 pixels) to merge similar features into one. This combination of a convolutional layer and a pooling layer forms the main building block of a CNN through which abstract features are learned, and can be repeated many times to increase these features' complexity. In the second step, the output of the final pooling filter is flattened into a one-dimensional vector (not unlike a vector of covariates) and fed into a fully-connected neural network consisting of multiple layers of nodes. This step uses the features extracted in the first step to predict whether the image contains the irregularity of interest. As each image is fed through the model, weights are updated to improve predictive performance through the feedback mechanism known as backpropagation (Rumelhart, Hinton, and Williams 1986). Because of the complexity of the model, images are loaded in batches; one complete iteration over all images in all batches is referred to as a learning epoch. CNNs' flexibility and performance have made them the gold standard in supervised classification of visual data and especially document analysis (LeCun et al. 1998; Razavian et al. 2014).

We introduce three improvements to this general framework used in Cantú (2019). First, we do not pre-select our training sample to achieve balance between irregularities and non-irregularities. This approach is prohibitively costly not just due to the relative rareness of some irregularities (see Table 2 below), but also because a different training sample would have to be obtained for each of our 10 irregularities. Instead, we build data augmentation into our models. As each batch is loaded, each image is randomly permuted via rotation, cropping, flipping, and other small adjustments.[12] This process synthetically increases our training sample size, as no image is exactly

12 . See the online appendix for a complete description of the data augmentation process.

repeated across epochs, in turn decreasing the propensity for the model to overfit the training data—resulting in better overall prediction (Krizhevsky, Sutskever, and Hinton 2012; Wong et al. 2016).

Second, we introduce transfer learning. Increasing the complexity of a CNN by adding more convolutional layer-pooling layer combinations can yield better predictions by allowing it to learn more complex features in an image. However, this complexity makes models extremely computationally costly to train, effectively setting an upper bound on how many layers we can include in our model.[13] Transfer learning overcomes this problem by allowing us to start from a baseline of 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. We select the InceptionV3 model architecture (Szegedy et al. 2016), trained on the ImageNet database (Deng et al. 2009), due to its performance with these statutory forms.

Our last improvement over the Cantú (2019) approach is to tune our hyperparameters. CNNs' performance can vary widely depending on a number of factors, including the extent of image preprocessing, the number of images in a batch, the number of convolutional and pooling layers, and the size of the fully-connected layers in the prediction step, to name a few (Li et al. 2020). To maximize predictive accuracy and ensure we are properly identifying irregularities, we tune over five model architectures, three data augmentation settings, two image sizes, and two batch sizes.[14] For each of our 10 irregularities, we tune over the 60 models in this $\{5 \times 3 \times 2 \times 2\}$ grid, resulting in 600 fitted models.

Each model is fit on only 2,000 images from the training set, while a validation set of 500 images is used at the end of each epoch to minimize cross cross-entropy loss (i.e., maximize predictive accuracy). After each model is done fitting, we compute performance metrics on a true test sample composed of the remaining 500 images

13 . The models we study take approximately 6-48 hours to estimate on a cluster with 50Gb of allocated memory; as discussed below, we estimate over 600 of these models.

14 . The five model architectures vary the number of convolutional layer-pooling layer blocks and the number of filters within them, as well as the size and number of fully connected layers in the prediction step. See the online appendix for details.

which were entirely held out from model training. The hyperparameter values that maximize predictive performance on this test set for a given irregularity are then saved and used to identify that irregularity among the remaining 27,000 forms. All models are estimated using the keras module in Python via slurm arrays on Cardiff University's high-performance computing cluster (Chollet 2015).

Predicting irregularities

Table 2 summarizes the performance of the tuned models. Because each of the irregularities we study are rare (ranging from 31% of observations to less than 1%) common metrics such as classification accuracy are inappropriate for these data. For instance, since 99.6% of statutory forms scanned correctly, we could achieve accuracy of 99.6% by simply predicting that every form scanned correctly. 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.[15] F_1 scores range between zero and one, with higher values reflecting better performance.

As the Table indicates, our classification models perform exceptionally well. Two types of irregularities, scan quality and presiding officer signature, are classified perfectly, with every single observation among the test sample correctly predicted.[16] Five more variables achieve F_1 scores of over 0.99, with just 1-7 observations among the held-out sample of 500 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 0.957 for edited results. By way of comparison, the model for edited results studied in Cantú

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

16 . The 495 forms that scanned correctly were predicted to be good scans, while the two forms that did not scan correctly were predicted to be bad scans; the 494 forms with presiding officer signatures were all predicted to have one, while the 4 without signatures were predicted to not have one. (Each sample is slightly smaller than 500 observations due to missingness among the statutory forms.)

Table 2: Image classifier performance

Irregularity	Prop. minority	F_1
Document problems		
QR code missing	0.006	0.999
Poor scan quality	0.004	1.000
Procedure problems		
Form not stamped	0.030	0.988
Presiding officer did not sign	0.008	1.000
Agent problems		
No agents listed	0.048	0.996
Any agent did not sign	0.103	0.978
No agents signed	0.052	0.997
Agent signatures appear identical	0.052	0.996
Agent refusal to sign listed	0.026	0.992
Results edited	0.308	0.957

Prop. minority indicates the proportion of observations in the held-out sample belonging to the less-common class (e.g., 31% of results were edited while 69% 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.

(2019) achieves an F_1 of approximately 0.885.[17] Our results correspond to reducing the number of errors from 11 to just 7 out of every 100 observations—a 36% reduction in the raw error rate. Across all irregularities, we expect an error rate of just 1.58 cases per 100, or 475 misclassified cases (per irregularity) in our sample of approximately 30,000 polling stations.

To make the analysis more tractable, we group these irregularities into three sets of problems relating to the conceptual groups defined above: the production and transmission of the statutory form itself; basic procedural requirements of certifying

17 . While Cantú (2019) does not report these values, they can be approximately inferred from the confusion matrices provided. Details can be found in our online appendix.

the form; and observation of the electoral process by candidates or their agents. Given its uniqueness and theoretical relevance, we also directly study whether results were edited. These groups are constructed as indicated in Table 2, with our three aggregate variables coded as a problem if any variable in that group is predicted to be irregular.[18]

Irregularities in the 2013 presidential election

Figure 2 summarizes the average occurrence of each irregularity category by constituency, grouped by county. Each dot indicates the mean proportion of polling stations within a constituency exhibiting a given problem, with lines for 95% confidence intervals, colored by county.[19]

Two notable patterns emerge. First, agent and editing irregularities are much more frequent than either document- or procedure-related problems. Second, both agent and editing problems appear nation-wide, and are not concentrated in any particular stronghold. For instance, neither agent problems nor edited results are more likely in incumbent ethnic strongholds, defined as polling stations for which the electorate is more than 90% Kikuyu or Kalenjin.[20]

In addition, Figure 2 indicates that most document problems are concentrated in two constituencies (one in each of Kitui and Vihiga counties), wherein all polling stations exhibited irregularities. Since election administration is organized at the constituency level, the most plausible interpretation of this result is that administrators in these two constituencies may have failed to provide adequate training to presiding officers on how to scan results. In Vihiga, Emuhaya constituency shows 55 forms with document problems, with only one unproblematic form. Upon closer inspection, the reason for this finding is indeed a training problem: forms were accidentally scanned in landscape mode, cutting off a significant part of the page.[21] Procedural problems

18 . 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.

19 . These estimates are generated by estimating a linear probability model with constituency fixed effects, clustering standard errors by constituency.

20 . See the online appendix for regression results.

21 . The outlier in Kitui is the result of Kitui West constituency only having one polling station form

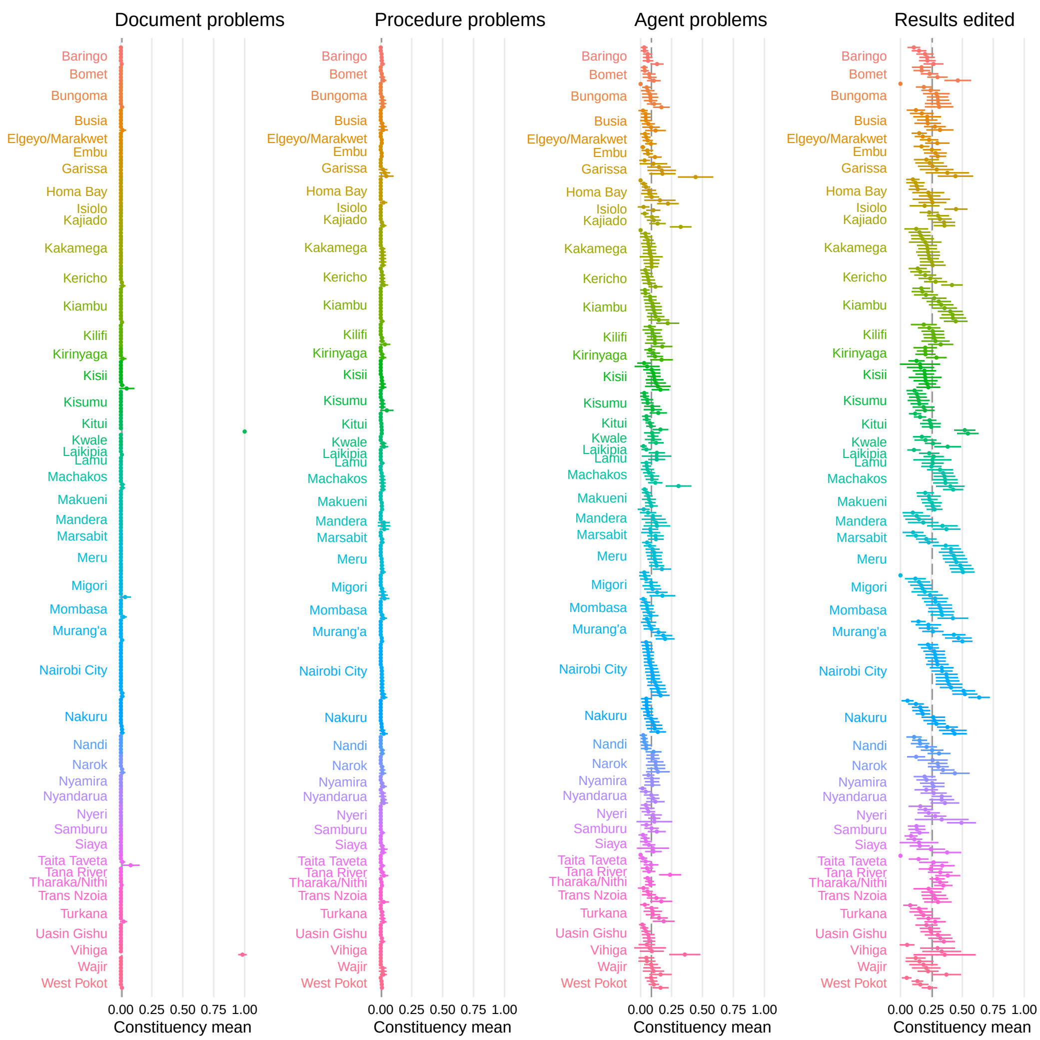


Figure 2: Proportion of polling stations with statutory irregularities. Dots indicate constituency fixed effects estimated using a linear probability model, with lines for 95% confidence intervals. Colors correspond to counties.

are similarly rare.

Greater variation is apparent in agent problems and edited results, but again no clear patterns emerge. Agent problems appear to vary within counties instead of across them. Such problems occur less frequently in pro-Kenyatta counties such as Bomet and Nandi, but more frequently in Meru and Murang'a, both of which were also government strongholds. Similarly, agent problems were less frequent in Siaya and Kakamega but more frequent in Kilifi and Kwale, all opposition counties.

Nor do these results suggest more irregularities in hotly contested counties. For example, Nairobi City county is one of the places with the most edited results and ended up supporting Odinga by a margin of 53% to Kenyatta's 47%. However, the next-most competitive county, Narok, experienced an average number of edited results; Samburu county, the next-most competitive after Narok, had fewer than average edited results. Overall, Figure 2 suggests little to no relationship between partisanship or polarization and statutory irregularities.

Thus far, we have examined election irregularities as an outcome. In Table 3, we examine how these irregularities relate to election results. We regress polling station-level voting outcomes on each of these problems, using linear probability models with constituency fixed effects and standard errors clustered by constituency. For each set of results, we estimate simple models where the only independent variable is the presence of an irregularity, as well as models with a series of controls. These controls include population density, ruggedness of terrain, ethnic fractionalization, geographic isolation, poverty, literacy rate, and night-time lights (a proxy for economic activity).[22]

In the top half of the Table, we study three dependent variables: turnout, the absolute vote margin, and Kenyatta's vote margin, all expressed as proportions (the first two ranging over $[0, 1]$ and the latter over $[-1, 1]$). Only problems relating to edited results have any relationship to turnout or the absolute vote margin. In both cases, however, the relationship runs in the opposite direction from what an expectation of

available, which happened to have a document irregularity.

22 . See the online appendix for data sources, coding rules, and full regression results. Our estimates are robust to using ward fixed effects, county fixed effects, or no fixed effects.

Table 3: Irregularities and electoral outcomes

	Turnout		Abs. margin		Margin	
Document prob.	-0.01	-0.02	-0.02	0.01	0.06	0.05
	(0.02)	(0.02)	(0.03)	(0.02)	(0.06)	(0.04)
Procedure prob.	-0.01	-0.01	-0.00	-0.00	0.00	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
Agent prob.	-0.00	-0.00	0.00	0.00	-0.00	0.00
	(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)	(0.00)	(0.00)
	Bal. rejected		Bal. disputed		Bal. spoiled	
Document prob.	0.45	0.40	$-0.00^{\dagger}$	$-0.00^{\dagger}$	0.08	-0.15
	(0.53)	(0.70)	(0.00)	(0.00)	(0.32)	(0.24)
Procedure prob.	0.44	0.47	0.01	0.01	0.34^{*}	0.34^{*}
	(0.33)	(0.35)	(0.01)	(0.01)	(0.15)	(0.16)
Agent prob.	0.20^{*}	$0.16^{\dagger}$	-0.00	-0.00	0.08	0.08
	(0.09)	(0.10)	(0.00)	(0.00)	(0.05)	(0.05)
Results edited	0.62^{*}	0.55^{*}	0.00	0.00	0.37^{*}	0.36^{*}
	(0.10)	(0.10)	(0.00)	(0.00)	(0.03)	(0.03)
Controls		✓		✓		✓
Constituency FE	✓	✓	✓	✓	✓	✓

$^{*}p < .05,^{\dagger}p < .10$. Estimates are from linear models with standard errors in parentheses. Controls are as described in the text. All models include constituency fixed effects and standard errors clustered by constituency. Full results are reported in the online appendix.

fraud would predict, as polling stations with statutory forms showing edited results have lower turnout and a smaller absolute vote margin. Moreover, the miniscule effect sizes suggest that vote tallies are largely unchanged even when results are edited, and the edits do not appear to favor either Kenyatta or Odinga, as there is no relationship to the vote margin at all. Taken together, the most plausible interpretation of these results is that edited results reflect not deliberate manipulation of results but rather more mundane corrections to human error in counting ballots.

One alternative explanation is that these results reflect electoral manipulation not by ballot-stuffing, but rather by ballot-shaving. In this scenario, partisans systematically discard valid votes to favor particular candidates. The negative relationship between edited results and turnout suggests this is a possible mechanism, though the null result for Kenyatta margin suggests it would have to be carried out by partisans for both sides. At the same time, the negative coefficient for absolute margin indicates that partisans would have to be shaving votes in polling stations their opponents won, i.e., shaving votes for the winner—an implausible mechanism given the stark polarization at the polling station level.[23]

Nevertheless, the bottom half of Table 3 investigates this alternative explanation by studying similar models where the dependent variables are instead the number of ballots that were rejected, disputed, or spoiled at each polling station. Broadly, these estimates suggest that statutory irregularities correspond to more of these ballots. Yet they also indicate that the vote-shaving hypothesis is implausible: the presence of any such irregularity has such a small effect size that, at most, only one ballot per polling station is likely to be affected. Aggregated, this suggests that perhaps as many as 30,000 votes could have been shaved—not enough to sway the outcome of the election, but perhaps enough to get Kenyatta past the 50% runoff threshold. While we cannot definitively rule this possibility out, it seems unlikely that the incumbent party would embark on a nationwide plan to violate electoral law by having its officers shave one vote here and there, rather than invest in altering tallies in a few key constituencies.[24] A much more plausible interpretation is just that the presence of problematic ballots makes the process of counting and aggregating ballots more difficult, leading to more errors by electoral administrators, resulting in more edited results.

If irregularities in statutory forms do not reflect electoral manipulation, but rather simple errors, then we would expect to observe them relatively evenly throughout Kenya, as hinted in Figure 2. To examine this question more closely, we regress each of the four problem categories on the geographic and socioeconomic variables used as

23 . Only 4.18% of polling stations had an absolute margin of less than 10 percentage points, and the average absolute margin was 71.30 percentage points.

24 . Indeed, the theory that rigging occurred in the tallying of results in Nairobi, rather than at individual polling stations, featured centrally in the legal battle over the election results.

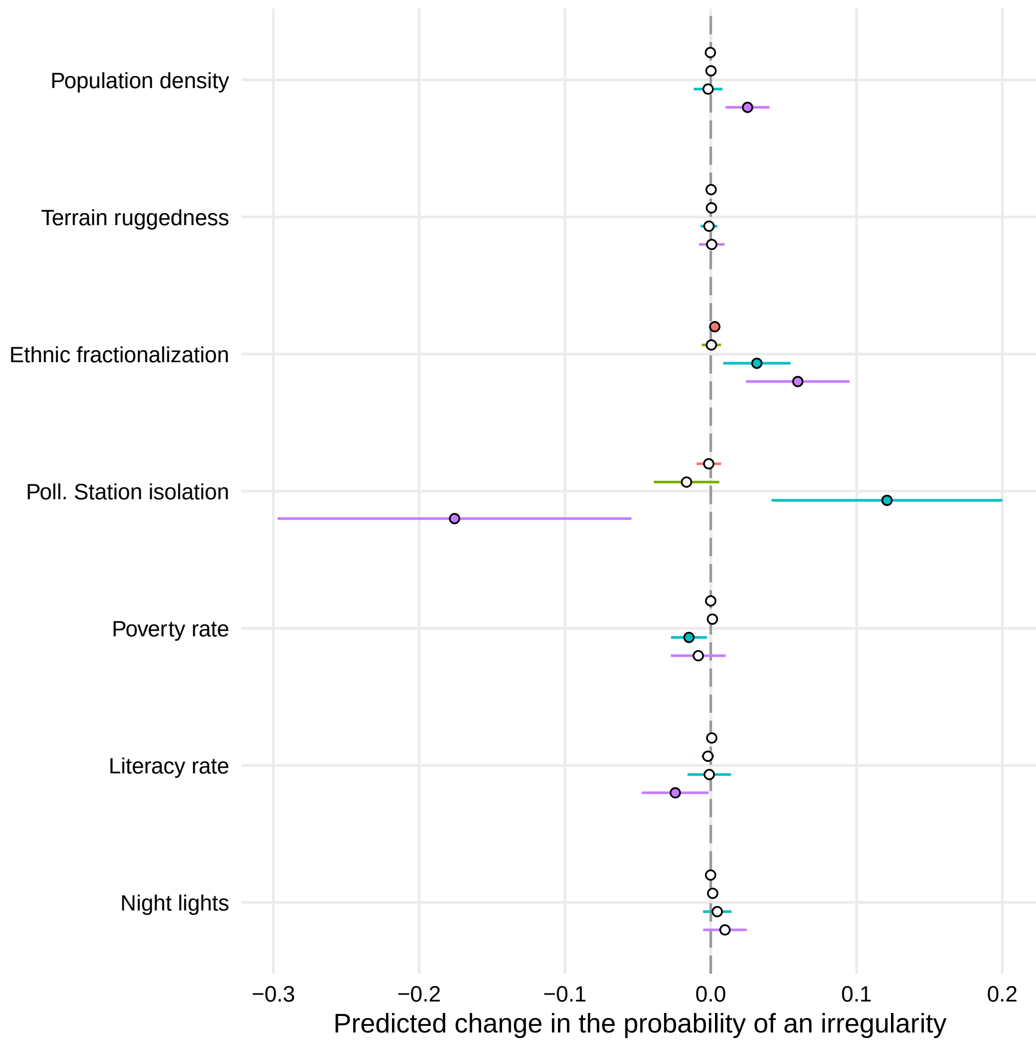


Figure 3: Predicted relationship between polling station characteristics and the probability of statutory irregularities. Dots indicate estimates from linear probability models, with lines for 95% confidence intervals. Hollow (filled) dots indicate estimates that cannot (can) be statistically distinguished from zero at $p < .05$. Colors refer to irregularity types, with orange for document problems, green for procedure problems, blue for agent problems, and purple for edited results.

controls in the analysis above. Figure 3 plots the coefficients from these four models.[25]

None of our geographic or socioeconomic variables is meaningfully related to document or procedural problems (in orange and green, respectively). Agent problems (blue) and edited results (purple), however, appear to be driven by at least three separate mechanisms. First, the strongest correlate of agent problems and edited results is polling station isolation. This is an intuitive result: more isolated polling stations are much more likely to be costly for candidate agents to travel to, resulting in irregularities that arise from agents not being present to observe the process (and verify that they have done so by signing the statutory form). And without agents present during vote tabulation, presiding officers at isolated polling stations are less likely to catch mathematical errors, resulting in fewer edits to results. A second mechanism explains the link between more urban and diverse areas and higher rates of statutory irregularities. We think it is likely that this result tells a story about contestation—that such polling stations are more hotly contested, leading to more disputed ballots, more edits to results, and more refusals to sign by candidates' agents. Third, the small negative effect of literacy on the probability of edited results may reflect a competence mechanism: better-educated electoral administrators may be less likely to make errors counting ballots, requiring fewer changes to the statutory form.

Finally, to study more closely the types of edits that are made, we sampled 500 of the statutory forms manually coded (during the construction of our training sample) as having been edited. 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.

25 . See the online appendix for complete regression results. Note that unlike the foregoing analysis, these models do not include constituency fixed effects because much of the variation we are interested in does not vary significantly by constituency, so including them would be an "easy test" of our theory. Nevertheless, the results are effectively unchanged if such fixed effects are included.

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.

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 systematic fraud.

Taken together, these results present a clear picture. In Kenya's 2013 presidential election, statutory irregularities are not concentrated in particular constituencies or counties, do not correlate with electoral outcomes, do not produce meaningful changes in vote tallies, appear to be produced not by partisan processes but rather logistical and competency-based mechanisms, and simply look like fixes to human mistakes. While we cannot definitively rule out electoral manipulation, we find very little evidence in support of systematic fraud. Our analysis suggests instead that irregularities are just minor deviations from an ideal election, entirely consistent with electoral administration in a developing democracy.

Conclusion

Administering elections is costly and difficult even for rich democracies with long histories of voting. For developing democracies like Kenya, the challenges of conducting an election are even greater, resulting in a number of common irregularities that often look like fraud. In this paper, we demonstrate that these irregularities do not always indicate an intent to manipulate electoral results. Statutory forms may frequently show document or procedural problems, may indicate no observers were present to verify the count, or may reflect edited tallies. But when analyzed closely using cutting-edge deep learning tools and extremely granular data, these issues appear to be the result not of malfeasance but of simple human error.

Our work has several limitations. Most importantly, we study only irregularities related to the production of statutory results. Yet the absence of clear patterns linking irregularities to fraud in this case does not prove that Kenya's electoral process in 2013 was entirely free from fraud—only that form-based irregularities do not reflect significant changes to the election results. Malfeasance can occur at many other times in an electoral cycle (Norris 2014), and there are many points of entry for fraudulent behavior. For instance, fraud can occur during the aggregation stages of an election, when polling station results are transferred from one form to another, or from a statutory form to a centralized database of results (Callen and Long 2015; Cantú 2019). The approach we explore here should therefore be seen as a complement to, and not a substitute for, existing methods for examining electoral processes.

A more pessimistic critique of our approach is that our data are produced by Kenya's Independent Electoral and Boundaries Commission, which may itself might have been party to fraudulent production of these forms. While we acknowledge the fraught electoral history in Kenya, we think this theory implausible. If the forms had been manufactured wholesale, we would expect the organization of the original files to be more uniform, rather than haphazardly bundled in *ad hoc* documents concatenated by constituency. We would expect little variation in scan or document quality. We would expect more uniformity in terms of writing implements, ink, signatures, and other idiosyncratic details arising on election day—such as the random commentary and tea stains marking the normal passage of business. And most importantly, we would expect to see many fewer normal irregularities related to agent problems and edited results. Nevertheless, this critique does raise an important point regarding the provenance and chain of custody for election data. Future efforts might seek to standardize the production, storage, and preservation of election materials of the type we study here.[26]

(Future work to be written...)

26 . See https://www.openelectiondata.net/ for one first step in that direction.

References

Alvarez, R. Michael, Thad E. Hall, and Susan D. Hyde, eds. 2008. *Election Fraud: Detecting and Deterring Electoral Manipulation.* Washington, DC: Brookings Institution Press.

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

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.

Beber, Bernd, and Alexandra Scacco. 2012. "What the Numbers Say: A Digit-Based Test for Election Fraud." *Political Analysis* 20 (2): 211–234.

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.

Birch, Sarah. 2011. *Electoral Malpractice.* Oxford: Oxford University Press.

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. 2014. "Identifying Irregularities in Mexican Local Elections." *American Journal of Political Science* 58 (4): 936–951.

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

Cantú, Francisco, and Sebastián Saiegh. 2011. "Fraudulent Democracy? An Analysis of Argentina's *Infamous Decade* Using Supervised Machine Learning." *Political Analysis* 19 (4): 409–433.

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.

Cheeseman, Nic, Karuti Kanyinga, Gabrielle Lynch, Mutuma Ruteere, and Justin Willis. 2019. "Kenya's 2017 Elections: Winner-Takes-All Politics as Usual?" *Journal of Eastern African Studies* 2 (215-234).

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

Cohen, Mollie J., and Zach Warner. Forthcoming. "How to Get Better Survey Data More Efficiently." *Political Analysis.*

Daxecker, Ursula E. 2012. "The Cost of Exposing Cheating: International Election Monitoring, Fraud, and Post-Election Violence in Africa." *Journal of Peace Research* 49 (4): 503–516.

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.

Elections Observation Group. 2013. The Official Results are Consistent with Elections Observation Group's Parallel Vote Tabulation (PVT). Available at https://www.ndi.org/.

Enikolopov, Ruben, Vasily Korovkin, Maria Petrova, Konstantin Sonin, and Alexei Zakharov. 2013. "Field Experiment Estimate of Electoral Fraud in Russian Parliamentary Elections." *Proceedings of the National Academy of Sciences* 110 (2): 448–452.

Ferrari, Diogo, Kevin McAlister, and Walter R. Mebane Jr. 2018. Developments in Positive Empirical Models of Election Frauds: Dimensions and Decisions. Paper presened at the 2018 Annual Meeting of the Society for Political Methodology, Provo, UT, July 16-18, 2018.

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.

Gutiérrez-Romero, Roxana. 2014. "An Inquiry into the Use of Illegal Electoral Practices and Effects of Political Violence and Vote-Buying." *Journal of Conflict Resolution* 58 (8): 1500–1527.

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.

Hyde, Susan D. 2008. "How International Election Observers Detect and Deter Fraud." In *Election Fraud: Detecting and Deterring Electoral Manipulation,* edited by R. Michael Alvarez, Thad E. Hall, and Susan D. Hyde, 201–215. Washington, DC: Brookings Institution Press.

Idrobo, Nicolás, Dorothy Kronick, and Francisco Rodríguez. 2020. Do Shifts in Late-Counted Votes Signal Fraud? Evidence from Bolivia. Working Paper.

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/.

Klimek, Peter, Yuri Yegorov, Rudolf Hanel, and Stefan Thurner. 2012. "Statistical Detection of Systematic Election Irregularities." *Proceedings of the National Academy of Sciences* 109 (4): enikolopovetal13.

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.

Li, Liam, Kevin Jamieson, Afshin Rostamizadeh, Ekaterina Gonina, Moritz Hardt, Benjamin Recht, and Ameet Talwalkar. 2020. A System for Massively Parallel Hyperparameter Tuning. Conference on Machine Learning and Systems 2020.

Mebane, Walter R., Jr. 2006. Election Forensics: The Second-Digit Benford's Law Test and Recent American Presidential Elections. Paper presented at the Election Fraud Conference, Salt Lake City, Utah, September 29-30, 2006.

———. 2016. Election Forensics: Frauds Tests and Observation-level Frauds Probabilities. Paper presented at the 2016 Annual Meeting of the Midwest Political Science Association, Chicago, IL, April 7-10, 2016.

Mebane, Walter R., Jr., and Jasjeet S. Sekhon. 2004. "Robust Estimation and Outlier Detection for Overdispersed Multinomial Models of Count Data." *American Journal of Political Science* 48 (2): 392–411.

Medzihorsky, Juraj. 2015. "Election Fraud: A Latent Class Framework for Digit-Based Tests." *Political Analysis* 23 (4): 506–517.

Montgomery, Jacob M., Santiago Olivella, Joshua D. Potter, and Brian F. Crisp. 2015. "An Informed Forensics Approach to Detecting Vote Irregularities." *Political Analysis* 23 (4): 488–505.

Myagkov, Mikhail, Peter C. Ordeshook, and Dimitri Shakin. 2009. *The Forensics of Election Fraud: Russia and Ukraine.* New York: Cambridge University Press.

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

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

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.

Rozenas, Arturas. 2017. "Detecting Election Fraud from Irregularities in Vote-Share Distributions." *Political Analysis* 25 (1): 41–56.

Rumelhart, David E., Geoffrey E. Hinton, and Ronald J. Williams. 1986. "Learning Representations by Back-Propagating Errors." *Nature* 323:533–536.

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

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. Proceedings of the International Conference on Artificial Neural Networks.

von Goethe, Johann Wolfgang. 1774 [1957]. *The Sufferings of Young Werther.* Translated by Bayard Quincy Morgan. New York: Frederick Ungar.

Weinberg, Barry H. 2008. *The Resolution of Election Disputes: Legal Principles That Control Election Challenges.* Washington, DC: International Foundation for Electoral Systems.

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? Proceedings of the 2016 IEEE international conference on digital image computing: techniques and applications.