Using Machine Learning to Map Concentrated Animal Feeding Operations in New Mexico

Report prepared for the McGovern Foundation
January 2021

Authors: Vanessa Ehrenpreis, Marshall Worsham, Nick Clarke, and Adam Buchholz

Acknowledgements

This study would not have been possible without the generous support of the McGovern Foundation. Special thanks to Nikita Japra, AB Srinivasan, Hazem Mahmoud, and Claudia Juech for their help throughout this project. The authors would also like to thank Margaret McCall for her work on the grant application, Matthew Grimm for his heroic effort coding images, and Nick Boyd for his modeling expertise. Additional thanks to Dan Lorimer of the Rio Grande Sierra Club and Kathy Martin for their invaluable insights on New Mexico's regulatory context. Finally, the authors extend their thanks to the New Mexico Center for Civic Policy.

About Mapping for Environmental Justice

Mapping for Environmental Justice, a project of Earth Island Institute, is a team of policy, data-science, and community-outreach experts working with communities to develop environmental justice maps for advocacy and policy making. Their easy-to-use, publicly available maps paint a holistic picture of intersecting environmental, social, and health impacts experienced by communities across the U.S. MEJ envisions a society where all people live in healthy and safe communities regardless of race or socioeconomic status. Find out more at **mappingforej.berkeley.edu** or email at **hello@mappingforej.org**

Table of Contents

Executive Summary

Currently, no government agency has comprehensive information of Concentrated Animal Feeding Operations in the United States. This study helped fill that data gap by developing the first dataset of CAFOs in New Mexico. We identified 160 CAFO facilities accross the state using an open-source machine learning algorithm.

Small family farms are rapidly disappearing in the wake of "factory farms" that specialize in producing one crop or animal on a massive scale. Regulatory agencies call these factory farms Concentrated Animal Feeding Operations [CAFOs]. CAFOs raise animals in confinement, usually at high densities, to produce meat, dairy, or eggs. While CAFOs have made U.S. agriculture more productive, there are serious environmental and health effects of concentrating livestock and animal waste in small areas.

CAFOs can contaminate air quality and water sources, which have direct effects on workers' and surrounding communities' health. CAFOs have been connected to serious ground and surface water quality problems including nutrient build-up and contamination from pathogens, chemicals, or common agricultural substances. Studies have also connected CAFOs to air pollutants such as ammonia, hydrogen sulfide, methane, and particulate matter, all of which have associated human health risks. As a result, communities surrounding CAFOs are at risk of respiratory illnesses, declining mental health, and high blood pressure, among other ailments. CAFOs pose an even greater health risk for agricultural employees who are exposed to pollutants and dangerous working conditions daily

Despite CAFO's health and environmental impacts, little public information is available about the locations and characteristics of these facilities. **Even at the federal level, no agency has comprehensive knowledge of the size, location, and number of CAFOs that exist in the United States.**[1] Data gaps remain at both the federal and state levels, in part because identifying and monitoring CAFOs is traditionally very labor intensive. Previous efforts employed teams to physically drive or fly over agricultural areas, or manually pore over millions of images to visually identify potential CAFOs. However, recent studies have pointed to machine learning as a methodology for efficiently and accurately identifying CAFOs.[2]

This study fills the current data gap by creating the first dataset of CAFOs in the American Southwest. We used open-source machine learning algorithms to identify and geolocate CAFOs in New Mexico. We applied a convolutional neural network (CNN) architecture for image recognition to data from the USDA National Agricultural Imagery Program (NAIP). We processed imagery covering the state at 1 m resolution and took a quasi-random stratified sample of images to create training and testing datasets. We then trained a CNN to identify CAFO components from image features and localized the positive-classified images to pinpoint the locations of unique CAFO facilities.

Our effort identified 160 distinct cattle CAFO facilities across New Mexico, which accords with prior best-guess estimates by state-level regulators and local advocacy partners. The model was 87.5% accurate, with 14 false positives that were generally center-pivot irrigated fields, warehouses, highway intersec-

tions, and airport facilities. We also cross-referenced the CAFO locations with census data to understand the environmental justice impact of these facilities. We found that communities with CAFOs are lower-income, have lower rates of high school graduation, and have higher exposure to PM 2.5 than the state average.

This study successfully developed a model to identify CAFOs in New Mexico which could have wide applicability for other semi-arid western states in which CAFOs go under-monitored and underregulated. We aim to expand this project and close the CAFO data gap for the entire United States.

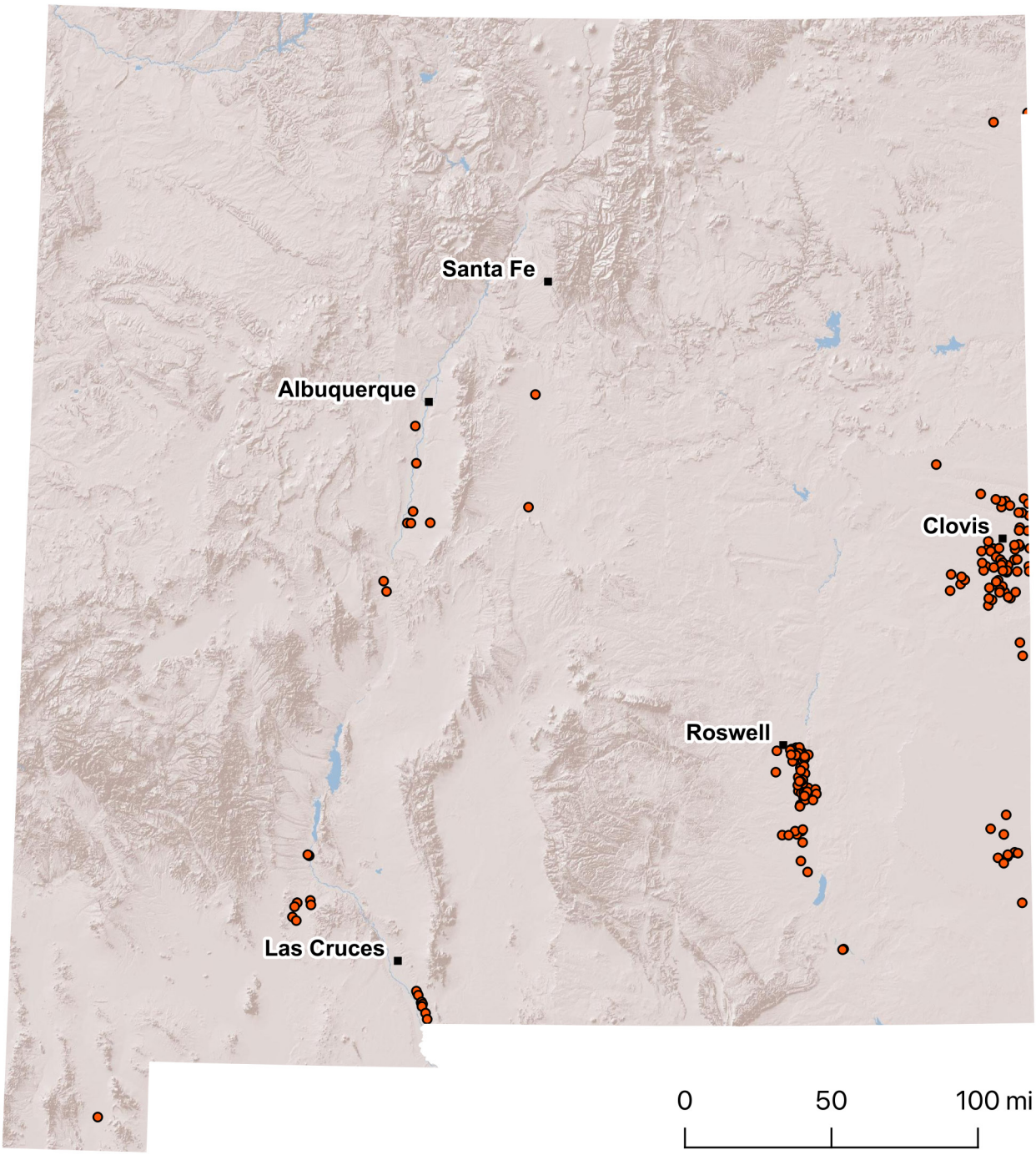


Figure 1. *Our study located 160 concentrated cattle feeding operations in New Mexico. Each CAFO facility is indicated by a red dot, with the highest density of facilities around Clovis and Roswell.*

Introduction

What is a CAFO?

U.S. livestock farming has industrialized over the past 50 years. Financial pressure to be more productive and profitable has fueled the rise of large corporately contracted farms across the poultry, dairy, beef, and pork sectors. Small family farms are rapidly disappearing in the wake of "factory farms" that specialize in producing one crop or animal on a massive scale. These operations are more efficient because they typically confine a single species of livestock in a limited area such as a feedlot, barn, or fenced in pen.[3] One of these facilities can raise as many as 2 million chickens or 800,000 hogs at one time.[4] As of 2020, there are an estimated 1.6 billion animals in 25,000 factory farms across the U.S.[5]

Regulatory agencies call these factory farms Concentrated Animal Feeding Operations [CAFOs]. CAFOs raise animals in confinement, usually at high densities, to produce meat, dairy, or eggs.[i] The EPA classifies CAFOs according to herd type and size, and the manner they discharge waste into the water supply. The number of CAFOs in the U.S. increased by about 230 percent between 1982 and 2002.[6] Current estimates put the number of CAFOs in the U.S. around 20,000, although advocacy organizations believe this could be an underestimate.[7]

CAFOs are ubiquitous across the U.S. These facilities have many environmental and health impacts as a result of their intense manure production. Above, a tractor is seen spreading manure.

Environmental & Health Effects of CAFOs

While CAFOs have made U.S. agriculture more productive, there are serious environmental and health effects of concentrating livestock and animal waste in small areas. Most of these impacts come from the large amount of manure CAFOs produce. Depending on the type and number of animals in a CAFO, manure production can range between 2,800 tons and 1.6 million tons a year.[8] In addition to nutrients like nitrogen and phosphorous, manure can also contain human pathogens, growth hormones, antibiotics, heavy metals, and chemicals used in agricultural operations.[9] Because there are no treatment requirements for animal waste, most CAFO manure is redistributed on, or discharged from, the facility's land in solid, slurry, or liquid form.[10] The USDA found a clear association between farm size and the concentration of manure, where large operations are more likely to apply manure to their

i CAFOs are a subset of Animal Feeding Operations [AFOs]. AFOs are agricultural operations where animals are kept and raised in confined situations. The EPA considers an agricultural operation an AFO if it confines animals for 45 or more days and does not have vegetation growth over the facility.

facilities more intensively.[11] While CAFOs are regulated on the amount and manner in which they can store and discharge manure, runoff, improper storage, and accidental releases continue to impact the surrounding environment.[12] As the following sections discuss, CAFOs can contaminate air quality and water sources, which have direct effects on workers' and surrounding communities' health.

Runoff from CAFOs can have disatrous effects on surrounding areas' water quality, including eutrophication and contamination from pathogens.

Water Pollution

Runoff from CAFOs is associated with groundwater and surface water contamination. As CAFOs apply more manure to their facilities, contaminants are more likely to leach into surrounding waterways. As a result, the agricultural sector is the leading contributor of pollutants to lakes, rivers, and reservoirs in the U.S.[13] The EPA found that states with high concentrations of CAFOs experience on average 20-30 serious water quality problems a year from manure mismanagement.[14] These "serious water quality problems" include nutrient build-up and contamination from pathogens, chemicals, or common agricultural substances.

CAFO runoff can contaminate surrounding surface waters with nutrient build-ups. High levels of nitrates and phosphates cause algal blooms that decimate aquatic ecosystems. These manure discharges have been linked to fish kills throughout the U.S.[15]

Furthermore, CAFO runoff can eventually leach into groundwater, polluting it with nitrates, chemicals, and pathogens. Groundwater contamination is even more difficult to trace and treat than surface water pollution because it is invisible. Undetected contamination means that the estimated 150 pathogens in animal manure, most commonly E. coli, Salmonella and Giardia, can spread and infect humans.[16]

Many Americans rely on groundwater as their water supply therefore, any contamination poses a serious health risk. One study estimated that millions of Americans living in or near farming communities have drinking water that is contaminated by dangerous amounts of nitrates and coliform bacteria from manure.[17] Drinking contaminated water can lead to diarrhea, dehydration, and even death for immuno-compromised populations.

Air Pollution

Manure from agricultural operations also emits gases and particulate matter that harm surrounding air quality, especially when concentrated at high levels. Studies have connected CAFOs to pollutants such as ammonia, hydrogen sulfide, methane, and particulate matter, all of which have associated human health risks.[18] Communities surrounding CAFOs are at risk of respiratory illnesses, declining mental health, and high blood pressure, among other ailments.
Several studies have found that CAFOs increase rates of asthma in surrounding communities, because of their toxic hydrogen sulfide, ammonia, and particulate matter emissions. CAFOs account for an estimated 64% to 86% of total global anthropogenic ammonia emissions alone.[19] Long-term exposure to these substances can cause asthma to develop into more serious ailments such as heart and lung disease.[20]

In addition to declining bodily health, odors from CAFOs also have deleterious impacts on mental health. Odor plumes from CAFOs are often overpowering and expansive, covering a large radius that extends far beyond the single CAFO facility. This pervasive stench causes significant lifestyle changes for residents that can lead to anxiety, depression, and even PTSD.[21] For example, one neighbor of a CAFO in New Mexico described how residents can no longer spend time outside their homes because "there are too many flies."[22]

CAFO workers face a variety of workplace hazards including harmful pollutants and bodily injury.

Worker Impacts

CAFOs pose an even greater health risk for agricultural employees who are exposed to pollutants and dangerous working conditions daily. The confined nature of livestock farming "exposes workers to manure dust, bacteria, and other particulates that can damage respiratory passages and lead to airway obstruction."[23] This was evident in the COVID-19 pandemic, when agricultural workers were at greater risk of contracting coronavirus due to cramped, unventilated facilities.[24] Additionally, agricultural workers are more likely to be injured because of interactions with livestock. Dairy farming has the second-highest prevalence of injuries among all US agriculture groups, most of which happen during milking activities.[25] One study found that 76% of dairy workers surveyed had at least one body part affected by an occupationally related musculoskeletal injury.[26]

Furthermore, the worker demographics of CAFOs mean immigrant workers disproportionately bear these health risks. Immigrant laborers comprise most dairy workers (51%) in the U.S., and "dairies that employ immigrant labor produce 79% of the U.S. milk supply."[27] However, estimating the exact proportion of immigrant agricultural laborers in the U.S. is difficult because data from the U.S. Department of Agriculture likely does not include a complete count of all migrant, undocumented, or seasonal workers.[28] In the absence of accurate data, the health impacts of CAFOs on immigrant workers remain underestimated.

Regulating CAFOs

As the number of CAFOs grew in the 1970s and 80s, legislators recognized the need to regulate them as a source of pollution. In 1972, the Clean Water Act delegated power to the EPA to regulate CAFOs through the National Pollution Discharge Elimination System [NPDES]. NPDES sets effluent limitation guidelines and standards for all point sources of pollution. While the EPA formally regulates CAFOs, it generally delegates its authority to state and local environmental agencies that administer NPDES permits. Under this system, a CAFO would apply for an NPDES permit through their appropriate state agency.[ii] These permits limit what and how much a facility can discharge and set stipulations for monitoring and enforcement.

However, this regulatory system has several flaws. First, NPDES exempts any CAFO that is not "actively" discharging waste. Rather than requiring all CAFOs to be permitted prior to discharging, the burden of proof is on regulatory agencies to demonstrate a CAFO is discharging and requires a permit. This rarely occurs because federal, state, and local agencies do not have the resources to proactively monitor and enforce facilities for violations. Despite being the watchdog for CAFO pollution, the EPA inspected a mere 0.6% of all CAFOs in 2017.[29] In the absence of strong federal oversight, state regulatory agencies have little impetus or resources to conduct their own monitoring activities. Inconsistent enforcement means CAFOs can discharge toxic pollutants with few, if any, consequences.

New Mexico is one of the nation's softest regulatory environments for CAFOs. It is one of only four states that does not have delegation authority from the EPA to write NPDES permits. Instead, New Mexico relies on the regional EPA office to issue permits, which

ii This is the case for all but four states that do not have NPDES delegation authority, one of which is New Mexico. The nearest regional EPA office handles NPDES permitting for these states.

results in permits that are not specific to local contexts and more difficult to enforce. After an intense push from advocacy organizations, New Mexico enacted "the Dairy Rule" in 2015 to fill this regulatory gap and protect groundwater. Under the rule, all actively discharging dairies are required to apply for a permit through the NM Environment Department, monitor their water quality, and install liners in all waste lagoons.[30] While the regulation was a landmark agreement between the dairy industry and environmental advocates, the NM State Legislature has yet to assess the Dairy Rule's efficacy in protecting groundwater.

Furthermore, despite CAFO's health and environmental impacts, little public information is available about these facilities. In 2008, the Government Accountability Office found that the "EPA does not have comprehensive, accurate information on the number of permitted CAFOs nationwide. As a result, EPA does not have the information it needs to effectively regulate these CAFOs."[31] The EPA has made some strides in collecting information on CAFOs, but data gaps remain at both the federal and state levels. Advocacy organizations' efforts to unearth CAFO data have had limited success. A 2019 study from the Natural Resources Defense Council found data on 7,595 CAFOs in 40 states, leaving more than half of the 17,000+ CAFOs the EPA estimated to exist unaccounted for in the agency's own data.[32] New Mexico regulatory agencies are particularly opaque with their state-level CAFO data; the state scored low across all six of Natural Resources Defense Council's measures of CAFO regulatory transparency.[33]

CAFOs in New Mexico: An Environmental Justice Issue

New Mexico is home to some of the biggest farms in America. Most of New Mexico's CAFOs are dairy farms with herds averaging around 2,300 cows per farm, which is the largest average herd size in the U.S.[34] These "mega" farms make New Mexico the ninth-largest milk-producing and fourth-largest cheese producing state in the U.S. with approximately 329,000 dairy cows.[35] Dairy is the state's largest agricultural sector and a major contributor to New Mexico's economy. The dairy industry contributes an estimated $105 billion dollars annually to New Mexico's economy, accounting for over 4% of the state's total GDP.[36]

The economic benefits from New Mexico's concentrated dairy industry are not without environmental and social costs. Most of the state's dairy farms are concentrated along a stretch of I-10 in southern New Mexico, which has been termed "Dairy Row" due to its overpowering odor.[37] These farms also pose a serious threat to the surrounding watershed. A 2009 analysis from the New Mexico Environment Department found that two-thirds of the state's dairies were contaminating groundwater.[38] Because 78% of New Mexicans rely on groundwater for their drinking water, any contamination has dire implications on human health.[39]

Furthermore, people of color and low income are disproportionately exposed to the negative health and environmental effects of CAFOs, making New Mexico's dairies an environmental justice issue. CAFOs tend to be located in areas where residents do not have the political power to resist the siting or expansion of such agricultural operations.[40] A study of pork CAFOs in North Carolina found "nine times more CAFOs in areas where there was more poverty and higher percentages of nonwhite people even after adjusting for population density as a measure of rural location and cheaper land."[41] New Mexico is no exception. Anthony, a town located along Dairy Row, is more than 98 percent Hispanic or Latino, and nearly half of its residents are living in poverty.[42] Many of these residents also work in the dairy industry, which means they're constantly exposed to pollutants. Mapping the health and environmental impact of CAFOs is a key component of understanding environmental justice in New Mexico because of these exposure risks.

What Will a CAFO Map Accomplish?

Accurate CAFO location data is crucial for regulators and the public. Regulators need to be able to identify CAFOs for proper permitting, monitoring, and enforcement. State agencies lack the resources and technical expertise to develop a machine-learning methodology, so this dataset is valuable with many potential applications such as identifying unpermitted or illegal CAFOs and geographically targeting regulatory enforcement resources.

For residents, knowing the locations of CAFOs will strengthen advocacy efforts and elucidate whether communities of color and low income are disproportionately exposed to pollutants. This dataset will provide advocates with an objective, scientifically supported tool to argue for enhanced environmental protection in their communities.

Methodology

From an analytical perspective, identifying CAFO facilities amid all of the other possible land surface features in a state is something of a needle-in-the-haystack challenge. Previously, mapping CAFOs required employing teams to physically drive or fly over agricultural areas or to manually pore over millions of images to visually identify potential CAFOs. Both approaches are time-intensive: on the order of three to five years for a full statewide census.[43] In the former case, the approach is not feasible over large domains, including many states in the western U.S.

One characteristic of CAFO operations makes them readily identifiable in aerial imagery: they tend to exhibit distinctive, relatively consistent geometrical features and coloration patterns when viewed from above. This characteristic, along with the availability of spatially extensive, high-resolution aerial data, makes the analytical problem of identifying CAFOs a promising application case for deep learning--based computer vision techniques.

We derived the locations of cattle CAFOs in New Mexico by applying a convolutional neural network (CNN) architecture for image recognition to data from the USDA National Agricultural Imagery Program (NAIP). We processed imagery covering the state of New Mexico at 1 m resolution and took a quasi-random stratified sample of images to create training and testing datasets. We then trained a CNN to identify CAFO components from image features and localized the positive-classified images to pinpoint the locations of unique CAFO facilities. Our methods relied substantially on the analytical architecture reported by Handan-Nader and Ho in 2019[44] , although we introduced several significant modifications to their methods in our data pipeline, modeling strategies, and image-localization procedures. The analysis yielded high classification accuracy on both repeated stratified k-fold testing, and on a 20% reserved test set (N=9236).

Data

Data source

NAIP produces high-resolution aerial imagery of the land surface in the contiguous U.S. acquired during the growing season. The program operates annually, imaging a third of U.S. states each year; a new dataset is produced for any given state, therefore, every three years. The data portray spectral reflectance observations from the Earth surface in four bands (Red, Green, Blue, and Near-Infrared). The images are acquired at a 1 m ground sample distance with a horizontal accuracy tolerance of 6 m to ground control points. The data are orthorectified using the National Elevation Dataset at 1 m resolution and color balanced by normalizing each image's radiometric range to that of adjacent images.

While the USDA Natural Resource Conservation Service makes the data available to the public, acquiring images at the scale of the state is infeasible with current federal image download servers and interfaces. To improve scalable access to such nominally publicly available data, Amazon Web Services (AWS) has developed a Registry of Open Data[45], a set of cloud-native open data repositories that third parties, including researchers and analysts, can access and query through AWS-based APIs. At the time of our analysis, the registry hosted the full set of NAIP imagery covering the continental U.S from 2012 to 2018, and was managed by the private GIS-software firm Esri.

Using the Python libraries boto3 and botocore, which comprise the Python wrappers for the AWS Software Development Kit (SDK), we sourced NAIP imagery acquired in 2018 over the entire state of New Mexico from the Registry of Open Data's naip-analytic bucket on AWS S3. The data were stored on S3 as digital ortho quarter quad tiles (DOQQs), each covering a 3.75 x 3.75 minute quarter quadrangle, with a 300 m buffer on all sides. These were stored on AWS as cloud-optimized GeoTiffs and compressed using the Lerc1 lossless compression algorithm.

Once accessed and decompressed the dataset comprised 8024 images. These provided complete coverage of the state with 450 m overlap between adjacent images, as Fig. 2 details. Along borders, images included data for bordering states up to the maximum image dimensions. All images were projected in the UTM coordinate reference system (CRS), under which New Mexico is divided into two zones, 12 and 13. Approxi-

mately 75 percent of the images were in UTM Zone 13, and the remainder in Zone 12. To maintain fidelity in on-the-ground distance measurements, we preserved the images in their native CRS and developed code to accommodate multiple zones.

The native resolution of the images was either 60cm or 1 m, and they were of approximately equal dimension in pixels: mean width 10255 pixels (σ=197), mean height 12207 pixels (σ=83). For those images at 1 m resolution, the surface distance depicted was equivalent to the pixel dimension. For those at 0.6 m resolution, they represented on average 6153 m (σ=118) in the x-dimension by 7324 m in the y-dimension (σ=50). Each image was approximately 550 MB in size, for a total data burden of approximately 4.4 TB.

Our neural network architecture required input imagery with 299x299 pixel dimensions. Conveniently, these dimensions produced an appropriate scale for detecting whole CAFO facilities and the constituent elements of larger facilities when the source images were at 1 m resolution. We therefore upscaled 0.6 m images to 1 m to achieve a consistent spatial resolution across the dataset. The small-scale input requirements, however, required us to further tile the NAIP images. We used the API from Descartes Labs[46], a private geospatial analytics firm, to create a micro-scale tile grid over the state. The grid included ~5.8 million tiles, each 259x259 m with a 20 m buffer on each side, for a total dimension of 299x299 m. Adjacent tiles overlapped 40 m. As with the NAIP imagery, the tiles were projected into a zonal UTM CRS, with Zone 12 and Zone 13 tiles rendered separately. On average, ~730 tiles intersected each NAIP image. The rationale for using a static grid, rather than tiling the images all at once, or building a tiling step into the modeling pipeline, has to do with our data and modeling architecture, as we describe in the next section. This architecture required each micro-image to have a permanent identifier, so that we could track it through training and testing, where applicable, and through final classification.

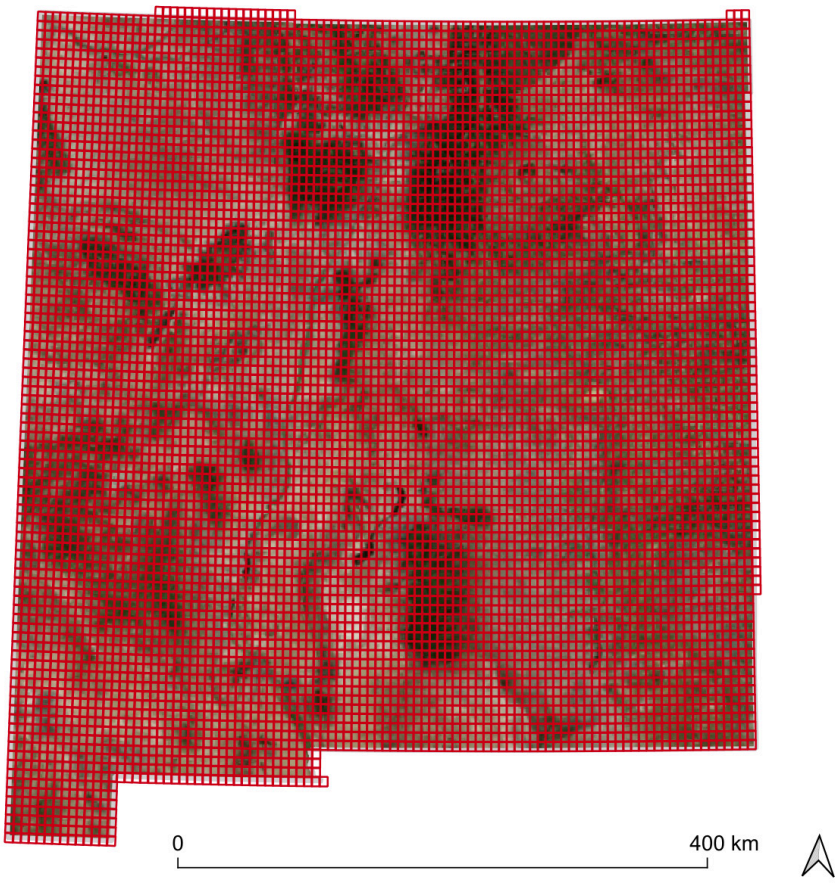


Figure 2. *Boundaries of NAIP image tiles overlying New Mexico in the AWS Registry of Open Data naip-analytic bucket. Each red-outlined polygon indicates the extent of one image in the registry. Tiles overlap by 450 m.*

Data pipeline

As the foregoing paragraph hints, one novel contribution of this project was the development of a data pipeline that relies almost exclusively on calls to a public data repository on AWS S3 servers, without our needing to maintain a separate, independent data store either locally or in the cloud. Such an approach has several advantages. For one, it saves costs---on the order of hundreds of dollars per month for a project centered on one large U.S. state. Increasing the scale of the project, say, to cover multiple states or the entire U.S. would entail less than linear cost increases, incurred primarily in the scaling up of computing power, with minimal growth in data storage costs. As long as the public repository remains available, this approach offers a flexible and rapidly scalable framework for ingesting new remote sensing data and performing analyses using different kinds of modeling tools or evaluating different types of land-surface features.

The essence of this approach depended on a relational database that connected the tile grid to the images in the S3 store and that allowed us to quickly read image data

from S3 and then temporarily store both data and geospatial metadata in memory while we performed manipulations and classification operations.

Once we had identified the relevant images in the S3 store and generated the micro-scale grid, we executed the remainder of the analysis in a high-performance Linux computing environment on the Cloudera Data Platform, provided to us through the Cloudera Foundation (now McGovern Foundation) Data4Change Accelerator Program. The platform simplified many of the data-engineering and environment-management requirements that would have otherwise required manual configuration. The computing environment sat in a virtual private server based on an AWS P3 Elastic Compute instance with 64 vCPUs, 488 GB memory and 8 GPUs, 128 GB GPU memory.

Our first step was to run a geospatial overlay procedure using the Python library geopandas. This operation yielded a dataframe that related each micro-tile to any of the S3 images that it intersected (Fig. 3). We stored these relationships along with the unique identifier of each tile (uuid), the bounding coordinates of the tiles and the larger images, the images' URL handles on S3, and the images' source CRS in a single table in a Spark database. The computational time at this step was somewhat intensive: running in series with an r-tree spatial indexing method on all cores in the high-performance environment took approximately three hours.

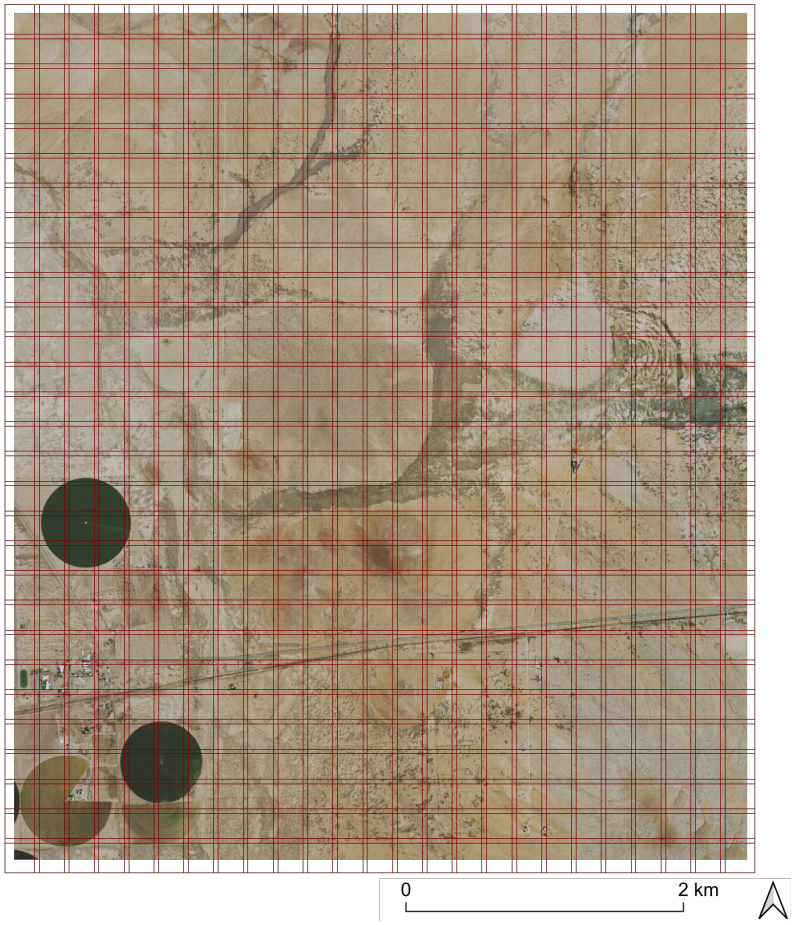


Figure 3. *One NAIP image with micro-scale grid tiles overlaid.*

In addition to the table of image-tile associations, we created a master table to store the uuids of each tile, as well as tables to store metadata attributes created during manual image tagging, assignments for manual image tagging, credentials for users of the application we developed for manual tagging, model parameters, model performance metrics, and model classification decisions (Table 1).

Table A1. Database schema is available in the Appendix.

Training set development

The next step was to create an image set for model training and testing. In Python, we developed a program that could download NAIP image data for any given tile, and we deployed the program, first, to generate the training/testing set using a quasi-stratified sampling procedure. Based on prior documentation efforts by Project Counterglow[47] we knew the coordinate locations of 56 confirmed or likely CAFO facilities in the state. At each of these points, we queried the NAIP imagery and downloaded imagery for the tile containing the point and its 8 adjacent tiles. Then, we used Google Places to identify the coordinates of land-surface features that could be easily confused for CAFOs, based on their surface expressions and nearby land cover. The search terms included "store", "campground", "RV park", "airport", "warehouse", "cemetery", and "shopping mall", among others. For each of these coordinate points, we again downloaded the tile containing the point and the 8 adjacent tiles. Finally, we took a random sample of 20000 image tiles that had not been sampled in either of the prior procedures and downloaded their corresponding image data. Concatenating these three sampling sets yielded a total of 24998 images.

Using the R programming language and RShiny, an API for developing interactive web applications, we developed an application for manually labeling images

in the training/testing set. The application called on the database and the sampled images to sequentially serve randomly selected images from the training-testing set, which human interpreters could evaluate.

We hired and trained two image coders to use the app to tag images. As an image was served, the coders inspected the image and compared it against Google Maps to determine whether the image depicted a CAFO facility. They registered their decision in the app interface, indicating whether the image depicted a "CAFO" or "non-CAFO," along with information, in the former case, about the likely animal operation type (cattle, swine, poultry, other), the proportion of the facility depicted in the image, the proportion of the image covered by the facility, and qualitative comments. A stratified random subsample of images was served to both coders, such that 10 percent of the total training-testing image set was double-coded, enabling us to evaluate cross-coder decision reliability. The coders met weekly with the research team to review the double-coded decisions and to resolve any classification conflicts and to make a final determination for any images with an "unsure" tag. Inter-coder reliability was high (92.5% agreement in the first round of conflict-resolution and improving to 97.0% in the final round).

The coders evaluated the full set of 24998 images. On inspection, approximately 4200 of the images (16%) were either duplicates or were missing substantial data (i.e., the images had 0 values for more than 25% of pixels in one or more bands). In most cases, the missing pixels were the result of sampling at the edge of a NAIP image. We removed the problematic samples and concluded that their proportion was not large enough to justify resampling and retagging; however, we modified the image-tiling code to avoid this problem in subsequent steps. We downloaded and stored the tagged image set temporarily in a private S3 bucket until model training and testing were complete.

In total, the coders evaluated 24998 images. After the initial cleaning to remove the problematic samples, the final set included 20760 images, of which 740 substantially depicted CAFOs and 20020 did not.

Modeling

Big-picture, we used a deep-learning image processing procedure to address our classification prob-

Table 1. *Metadata fields populated during manual image tagging.*

Item	Variable	User prompt	Selections	Stored Data Values
1	CAFO classification	Choose one classification for the image to the right.	Non-CAFO	0
			CAFO	1
2	Animal operation type	If you chose "CAFO" in (1), what type(s) of animal operation is depicted in the image? Otherwise, leave NA	Cattle	Cattle
			Swine	Swine
			Poultry	Poultry
			Other	Other
			Unsure	Unsure
			NA	NA
3	CAFO components	If you chose "CAFO" in (1), what component(s) of the CAFO are visible? Otherwise, leave blank.	Structure	Structure
			Feedlot	Feedlot
			Lagoon	Lagoon
			Plumbing	Plumbing
			NA	NA
4	Proportion of image occupied by CAFO	If you chose "CAFO" in (1), what proportion of the image is occupied by the CAFO? Otherwise, leave blank.	none of the image (0%)	0
			very little of the image (less than 10%)	1
			some of the image (between 10% and 25%)	2
			much of the image (between 25% and 50%)	3
			most of the image(between 50% and 75%	4
			almost all of the image (more than 75%)	5
5	Proportion of CAFO depicted in image	If you chose "CAFO" in (1), what proportion of the CAFO is depicted in the image? Otherwise, leave blank.	none of the CAFO (0%)	0
			some of the CAFO (between 33% and 66%)	1
			much of the CAFO (between 33% and 67%)	2
			most of the CAFO (between 67% and 100%)	3
6	Notes	Notes (250-character limit)		<Text string>

lem. Specifically, we applied transfer learning to retrain the last layer of a convolutional neural network (CNN) to differentiate CAFO and non-CAFO images. We initially anticipated that we would need to develop multi-class models or altogether separate models for different types of animal operations (e.g., cattle, swine, poultry). However, our sampling procedure for the training-testing set yielded only a handful of images depicting non-cattle CAFO operations. The sample sizes for these other facility types were therefore too small for multi-class or multi-model approaches. Based on conversations with local environmental organization partners, moreover, it appears that cattle operations are by far the dominant CAFO type in New Mexico and the highest priority for monitoring. Accordingly, we trained a single CNN to obtain class probability scores for a binary classification problem, differentiating only between CAFOs and non-CAFOs.

Using subsets of the training-testing dataset, we ran pilot tests exploiting two convolutional neural networks (CNNs): Inception V3 and ResNet18. The CNNs were previously trained on larger datasets, such as ImageNet, which comprises millions of images. We accessed pretrained versions of the models through the *PyTorch* framework. In both cases, we preserved the pretrained weights and froze all but the final layer to minimize the computational expense of retraining the models. We then removed the final layer from the models. The final layer is typically a user-specified regression or classification function that takes inputs from prior layers to and fits them to a given functional shape using a user-specified loss function, in order to make a probabilistic decision. Here, we dropped this layer and defined a separate logistic function, which took in pooled features extracted from the second-to-final layer of the CNN and fit a sigmoid curve using the pooled values and their associated class labels over a maximum of 2000 iterations. This approach afforded two advantages: first, it dispensed with forward-feeding and back-propagation steps in a fully-connected CNN layer, which were overkill for our classification problem. This saved milliseconds of computational time per image with negligible loss in classification accuracy (<-0.001), which amounted to hours of savings when the model was applied over the full image set; second, it gave us flexibility to find and set an optimal classification threshold when we turned to applying the model to new data.

We applied several common techniques for reducing overfit on the training set. These included: (A) oversampling to address class imbalance; (B) image augmentation and dropout; (C) k-fold cross-validation; and (D): lambda-optimization of the logistic classifier.

(A) Our data naturally have a large class imbalance (28:1 CAFO:non-CAFO in training, and likely two to three orders of magnitude higher in the full image set) To mitigate the negative impacts of this imbalance on classification performance, we oversampled CAFO-tagged images in the training-testing sets. We used a simple random minority-class oversampler to balance the CAFO/non-CAFO proportion evenly. In prototype testing, the random oversampling technique outperformed both the Synthetic Minority Oversampling Technique (SMOTE) and Adaptive Synthetic Sampling (ADASYN), which synthesize new images from various stochastic combinations of values from within the original dataset's feature space. By contrast, the random technique simply replicates existing images in the minority class. We randomly divided the oversampled data into two sets---80% for training, 20% for testing.

(B) We also deployed image augmentations, including random horizontal and vertical flips, z-score color enhancement, and random addition of Gaussian noise, to reduce overfit to the training data. And we allowed inputs to randomly drop out of the neural network during training, in essence reducing model complexity.

(C) With the 80% training set, we trained the model within a k-fold cross-validation framework, using 10 folds and 3 repeats. This process began by making 10 further random non-overlapping splits of the training set, each also in an 80%-20% proportion (1). For each split, the hybrid CNN-logistic model was trained using the new 80% subset (2) and then applied to predict class membership in the 20% subset (3). Accuracy statistics were generated for each split (4), then averaged across the ten splits (5). Steps 1-5 were then repeated three times with new random 10-fold splits to decrease the dependence of model performance on any given division of the data.

(D) At each fold, i, of the 10-fold cross validation procedure, we optimized the logistic function using a regularization parameter, λ. We specified a linear space of 10 possible λ values between 0.0001 and 0.01 and, at each λ, fit the logistic function using the training features from fold. We then selected the λ that maximized mean classification accuracy when the model was

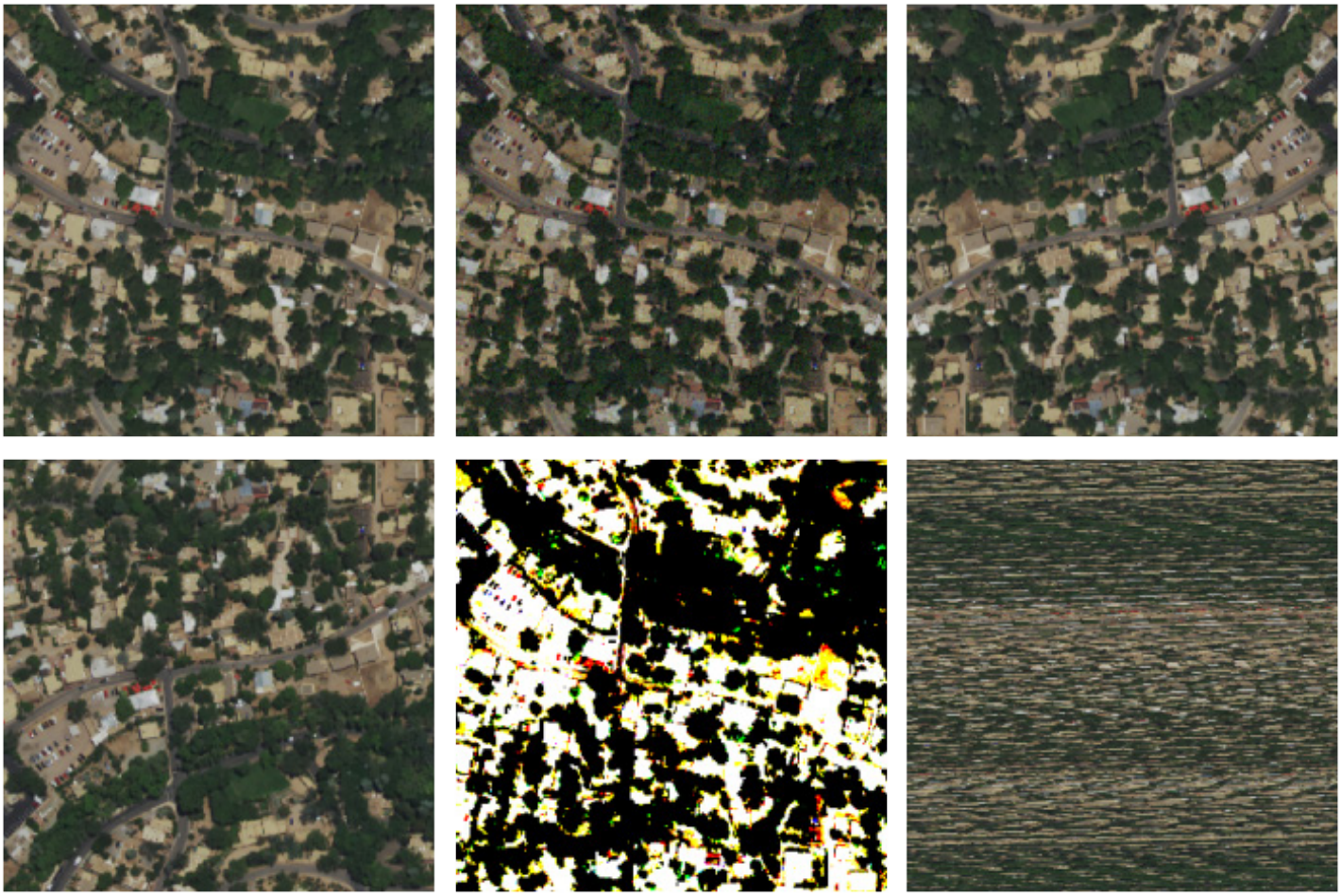


Figure 4. *Examples of image augmentations applied in the data engineering pipeline, shown here applied to a single image. Any given image in the dataset randomly received exactly one of these augmentations or none before classification.*

applied to the ith test set and used this to compute the accuracy statistics for fold i. This approach allowed us to achieve a minimally biased estimate of model skill.

We performed the steps described thus far on progressively larger subsets of the training data (N=500, N=1000, N=5000) using both ResNet18 and Inception V3. We generated five standard accuracy statistics at a 0.50 logistic classification threshold. Equations for these statistics is available in the Appendix.

Table 2 shows the performance of each model for each pilot size. We selected ResNet18 for final modeling based on its superior performance in the pilots.

Once we had landed on a suitable modeling strategy, we retrained the model using the k-fold procedure to estimate mean accuracy statistics. We then trained on the entire training set (N=29158 with oversampling) and tested on the 20% reserved for final testing (N=7290 with oversampling). After using the CNN to extract features from the training and testing sets, we produced precision and recall curves and receiver-operator characteristic curves, which report performance at all possible logistic-classification thresholds between 0 and 1. We then fit the logistic classifier a further 100 times for thresholds spanning the linear space between 0.01 and 0.99, exclusive, and produced final performance statistics at the threshold that optimized the F1 score, which we report below.

This step was one of the more computationally intensive in our workflow. Running against 60 CPUs and 1 GPU took approximately 7 hours.

Model deployment

The final step was to apply the top-performing model to the remainder of the data from New Mexico. Our ability to do this rested on the statistically reasonable and methodologically standard assumption that training-testing performance can be generalized to overall performance, as long as one can assume that the training-testing set is representative of the full population of data.

We retrained the model on the full training-testing dataset (N=20670). We then used the image-tile association table and tiling procedure detailed in the "Data pipeline" section to ingest NAIP data from S3 and to extract image arrays from all tiles not used in training and testing. We deployed the CNN to extract features from each image and predicted the image's class with the logistic classifier at the optimal threshold from the final round of testing. We ran this procedure in parallel with the pool function in the Python multiprocessing package, using batches of 12 NAIP images. Running against 60 CPU cores and one GPU required 2.8 days of computing time. The bulk of the computing demand came from tiling the NAIP images (~64%) and extracting features with the CNN (~28%).

When an image was classified positive (i.e., had a sufficiently high likelihood of depicting at CAFO), we stored its uuid and the identifier of the NAIP image from which it was tiled in a new database table.

Deduplication and localization

From advocacy and regulatory perspectives, class predictions on image arrays are less interesting than the locations of CAFO facilities. We therefore needed to translate the raw modeling results into more useful information. We developed a method for consolidating image-level results into point coordinates centered on unique CAFO facilities. The approach involved two steps: localization and deduplication.

To localize facilities, we took the centroid of the tile corresponding to each positive-classified image and stored its coordinates in its respective UTM zone CRS. We then computed the great curve distance from each point to all other points in the positive set. We considered any pair or cluster of points within 500 m of each other to comprise part of the same facility. For any such pair or cluster, we defined a convex-hull polygon around the related points and stored the centroid coordinates, and then removed the points that had made up the cluster from the result set. We repeated this procedure until the minimum distance between all points exceeded 500 m. We translated the final set of points into the WGS84 geographic CRS and stored them in GeoJSON format.

Reiteration

We manually validated the model-identified CAFO locations by referring back to the NAIP imagery and Google Maps imagery underlying the point coordinates. On doing so, we identified an unexpectedly high false-positive rate and, in visually inspecting the surrounding areas near the model-identified points, a substantial number of CAFO facilities that the model had missed (reported in "Results"). We deduced that the unexpectedly weak performance of the model on new data likely resulted from a violation of the assumption of representativeness. A high false-positive rate suggested that there was insufficient training data depicting land-surface features that the model could mistake for CAFO features, while a high false-negative rate suggested insufficient representation of the various shapes and forms that different kinds of CAFO facilities can take. As a result, we undertook a second run of the entire modeling procedure with an expanded training dataset.

From the result set, we retrieved the uuids of tiled images containing each point, re-classified the false-positives and added the false negatives. We then took the 8 adjacent images surrounding each point and manually tagged those. We added the new images to the original training and testing set and their corresponding labels to the image label set. One further issue was that the original training set included some images in which a small proportion of the pixels actually depicted features attributable to CAFO facilities. The remainder of the pixels included fields, highways, and bare earth, which likely introduced confusion into training. We therefore filtered out any samples in which <10% of the image depicted a facility. The resulting dataset size was 24204: 1114 depicted CAFOs and 23090 depicted non-CAFO land-cover types. We again oversampled the minority class and split the resulting dataset (N=46180) into 80%–20% training and testing sets. We then re-trained the model using the 10-fold cross-validation procedure with three repeats, generating accuracy metrics at each fold, and re-trained on the full training set (N=36944) to generate final accuracy statistics on the reserved 20% test set (N=9236). Finally, we trained the model on the full set of labeled images and applied the model to the remaining 4.5 million images, following the steps detailed above.

Having retrieved the positive-classification results at the image level, we localized and deduplicated to derive a new set of points, which we added to the first set of validated positive coordinates and stored in GeoJSON format.

Results

Here we present results from the pilot tests and from the second iteration of the modeling procedure. Results from the first run are included as an appendix.

Pilot tests

First, we report model performance from the initial pilot tests of ResNet18 and Inception V3 CNN frameworks, on the 20% of images reserved for testing in each test (table 2).

Image-level model decisions

In a classification procedure, the optimal classification threshold may vary as a function of the accuracy statistic selected for optimization or of the analyst's tolerance for false-positives or false-negatives. Both of these, in turn, depend on the context and objectives of the operation. We therefore present the final model's performance along a continuum of classification thresholds, and at the specific classification threshold we selected for this problem.

The logistic classifier assigned a score of 1 to any image with a CAFO probability score above a given threshold and a score of 0 to any image with a CAFO probability score below the threshold. The model decisions were then compared to the manual class labels to estimate recall, precision, and specificity. Figures 5 and 6 present these metrics at all classification thresholds between 0 and 1 for the second model run, after expanding the training set and reparameterizing the model. Fig. 5 shows a receiver-operator characteristic (ROC) curve, which plots specificity against recall and provides an estimate of the tradeoff between true-positive and true-negative rates at each threshold. Fig. 6 shows a precision-recall curve, which plots precision against recall and offers an estimate of the true-positive and false-positive tradeoff at each threshold. A curve approximating a 90° angle indicates strong performance. For all metrics, performance improves as values approach 1, and values greater than 0.95 indicate very high model skill.

In the 10-fold CV process, we used a static classification threshold of 0.5. In the final run, we used an optimizer to select the classification threshold that maximized the F1 statistic on the training set. F1 can be interpreted as the harmonic mean between precision and recall. Optimizing on this metric achieved a suitable balance between the preference for true-positives and aversion to false-negatives. On the 20% reserved test set, the optimal class threshold was 0.544.

The first panel of table 3 reports overall accuracy, precision, recall, and F1 as mean values and their standard deviations aggregated across the 10 folds of the CV procedure at threshold=0.500. The second panel reports the final metrics on the single run against the 20% reserved testing set at threshold=0.544. Fig. 6. shows the classification decisions for 36 randomly selected images from training and testing.

Facility-level results

The first model iteration produced 894 positive-classified images. Localizing and deduplicating the raw results yielded 565 point coordinates. On manual inspection, we found that only 86 points corresponded to unique true-positive CAFOs, and we identified an additional 64 facilities on the ground, which the model had missed. These high false-positive and false-negative rates suggested both that the model was overfit to training data and that training data were insufficiently

Table 2. *Mean and standard deviation of accuracy, precision, recall, and F1 statistics for pilot tests of classification strategies based on ResNet18 and Inception V3 CNNs.*

		Pilot 1 N = 500 Train N = 400 Test N = 100				**Pilot 2** N = 1000 Train N = 800 Test N = 200				**Pilot 3** N = 5000 Train N = 4000 Test N = 1000			
Model		Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1
ResNet 18	mean	0.949	0.927	0.976	0.950	0.911	0.889	0.941	0.913	0.935	0.928	0.945	0.936
	σ	0.023	0.039	0.027	0.022	0.035	0.049	0.041	0.034	0.011	0.015	0.022	0.011
Inception V3	mean	0.946	0.912	0.988	0.947	0.928	0.907	0.953	0.928	0.932	0.922	0.941	0.931
	σ	0.031	0.054	0.029	0.030	0.031	0.048	0.028	0.031	0.010	0.009	0.015	0.010

representative of the universe of land-surface features that appeared in imagery.

After enhancing the training set, reparameterizing the model, and running a second iteration, we found 27 additional positive-classified images beyond those already found in the first run. These depicted 13 unique potential CAFO facilities, although 9 appear to be defunct or non-operational. Manual inspection confirmed that these comprised 11 new facilities not previously identified, along with 14 false-positives, which all represented either smaller animal enclosures, rodeo facilities, or equestrian facilities. An additional 20 undetected sites were found in review.

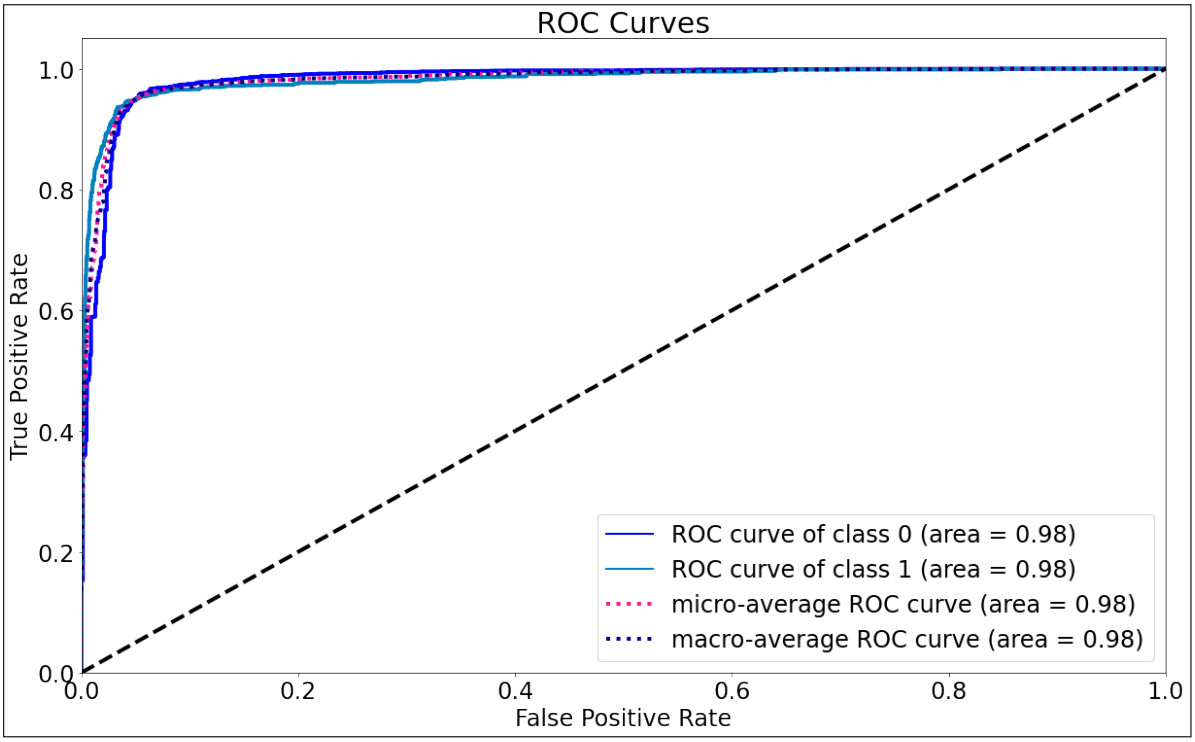


Figure 5. *Receiver-operator characteristic (ROC) curve plotting specificity and recall at all possible classification thresholds. The area under the curve is 0.98.*

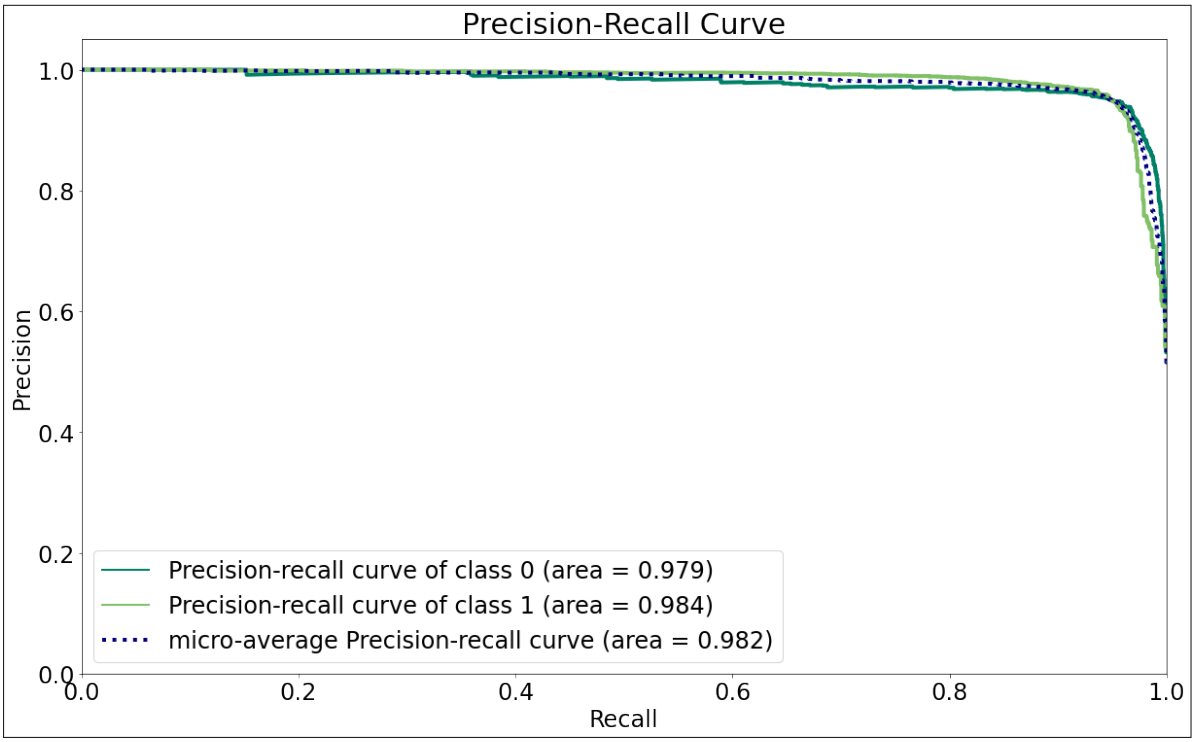


Figure 6. *Precision-recall curve showing true-positive/false-positive tradeoffs at all possible classification thresholds. The area under the curve is 0.98.*

Table 3. *Accuracy, precision, recall, and F1 statistics for final classification on 20% test set.*

			Statistic			
	Class Threshold		Accuracy	Precision	Recall	F1
10-fold CV	0.500	mean	0.950	0.949	0.952	0.951
		σ	0.003	0.005	0.004	0.003
Test set	0.544		0.952	0.956	0.947	0.952
10-fold CV	0.500	mean	0.950	0.949	0.952	0.951
		σ	0.003	0.005	0.004	0.003
Test set	0.544		0.952	0.956	0.947	0.952

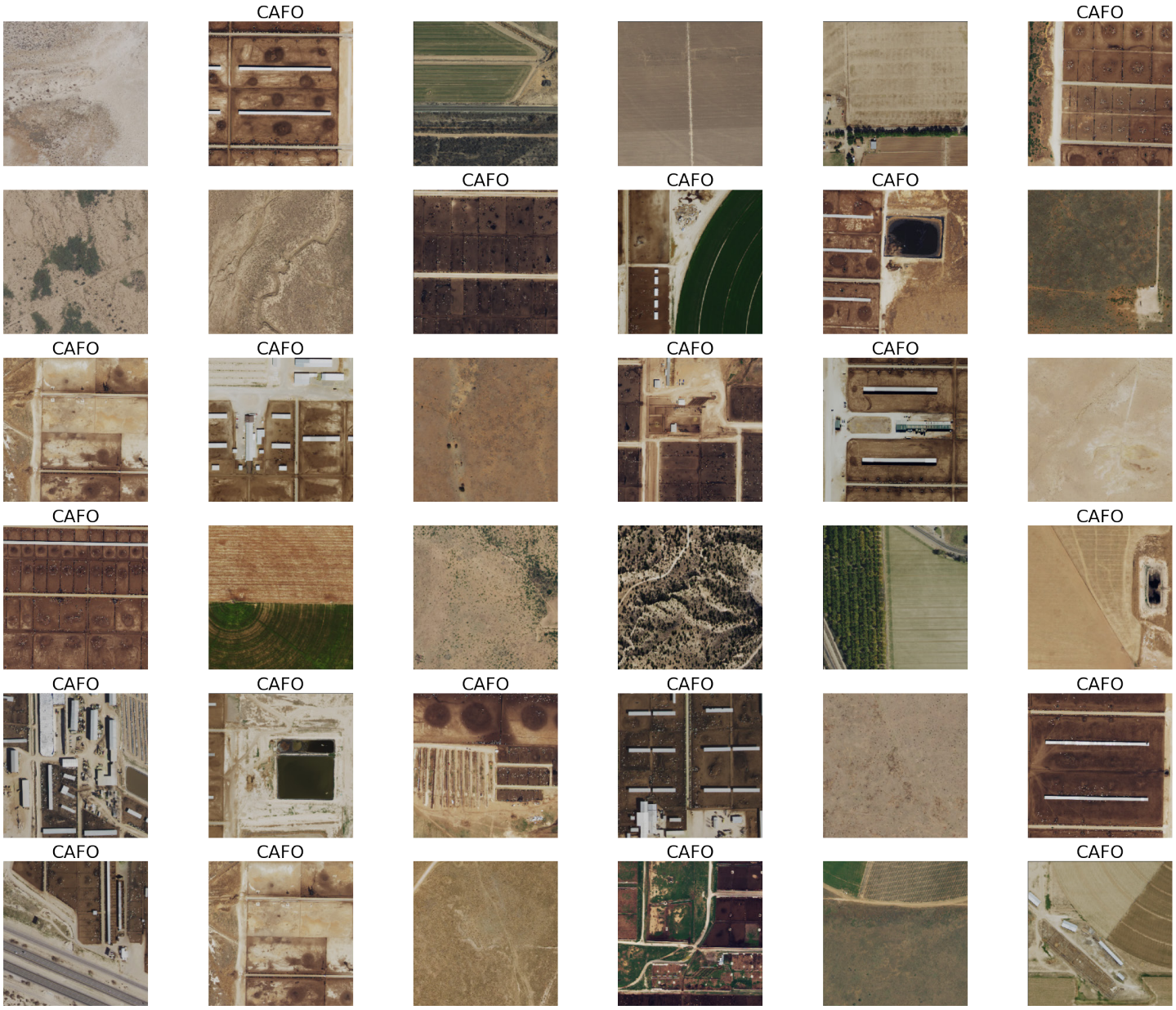


Figure 7. *Example class decisions for 36 randomly selected images during testing.*

Discussion & Applications

In sum, our effort identified 160 distinct cattle CAFO facilities across New Mexico, which, by order of magnitude, accords with prior best-guess estimates by state-level regulators and local advocacy partners. The algorithm accurately identified 87.5% of the CAFOs in our final dataset. This project built on previous efforts to demonstrate the technical feasibility of identifying CAFOs from satellite imagery, and created the first comprehensive such dataset in the American Southwest. We aim to expand this project and close the data gap for the entire United States.

False-positive images generally included center-pivot irrigated fields (which co-occurred in some CAFO-positive training images), warehouses, highway intersections, and airport facilities. Further effort could be spent improving the representation of these confusable land-surface features in the training data.

We also examined the environmental justice impact of these CAFOs by comparing CAFO locations to New Mexico's census data from the 2014-2018 American Communities Survey. We found that communities with CAFOs are lower income, have lower rates of high school graduation, and have higher exposure to PM 2.5 than the state average. These communities are also highly linguistically isolated: the average percentage of households without a member who speaks English fluently in communities with CAFOs is 9.58%, compared to the state average of 5.68%. These communities are more linguistically isolated than 79% of the state.

A major limitation of this analysis is its computational burden. The full training pipeline required close to 12 hours, while applying the model to address the final classification problem required nearly three days on high-performance computing infrastructure. Far more the computational intensivity lay in pilot testing and iterating. These requirements could make this kind of analysis out of reach for many organizations with small budgets or limited technical capacity.

Once the pipeline and modeling framework was established, the actual runtime was tolerable. Equally, because of our low-storage approach to data ingestion, the modeling is relatively low cost and easily scalable.

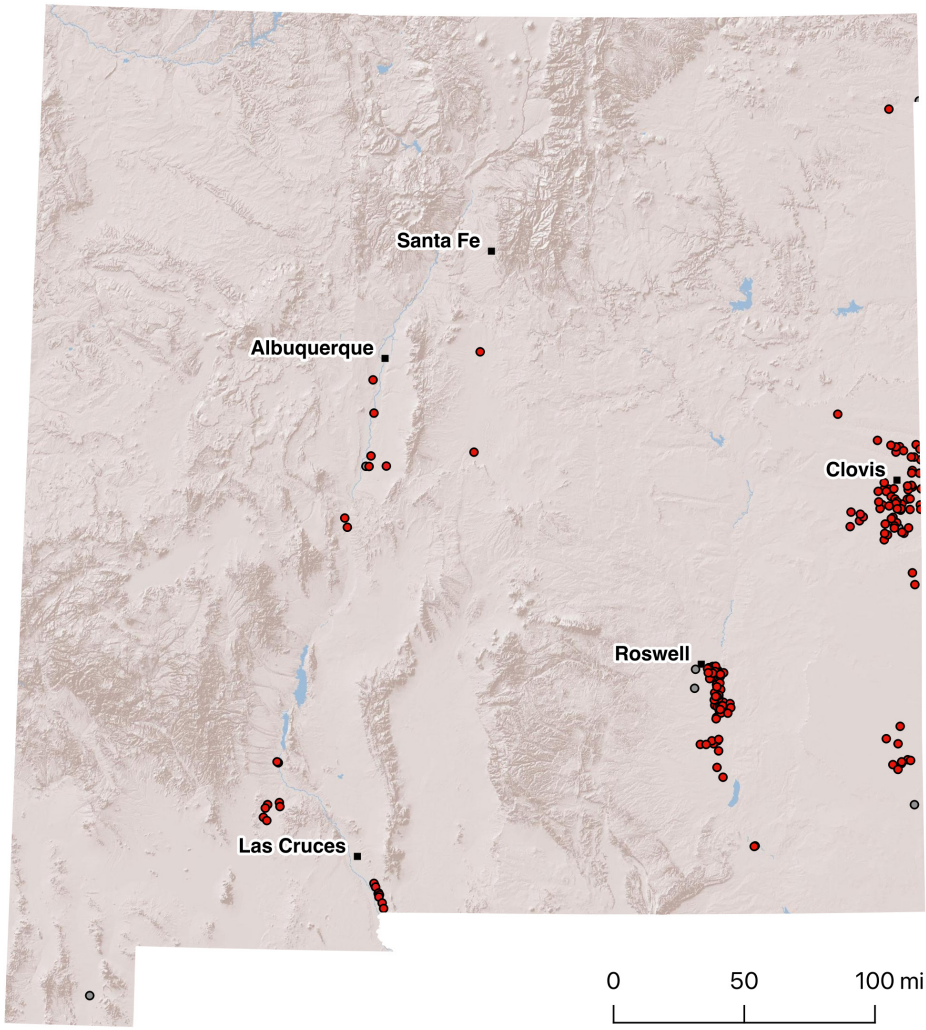


Figure 8. *Our study found 160 CAFOs total in New Mexico. Approximately 9 of these facilities are non-operational, as indicated by gray dots on the map.*

Appendix

Figure A1. *Equations for model accuracy, precision, recall, specificity, and F1.*

$$Accuracy = \frac{N\ true{-}positive\ predictions}{N\ images}$$

$$Precision = \frac{N\ true{-}positive\ predictions}{N\ true{-}positive\ predictions + N\ false{-}positive\ predictions}$$

$$Recall = \frac{N\ true{-}positive\ predictions}{N\ positive\ images}$$

$$Specificity = \frac{N\ true{-}negative\ predictions}{N\ negative\ images}$$

$$F1 = \frac{2 \times precision \times recall}{(precision + recall)}$$

Table A1. *Database schema including fields and data types.*

Table Name	Fields	Data Types
all_images		
	image_id	INTEGER PRIMARY KEY
	uu_id	TEXT NOT NULL UNIQUE
	tag_count	INTEGER DEFAULT 0
	url	TEXT NOT NULL UNIQUE
	manu_tag	INTEGER DEFAULT 0
final_labels		
	label_id	INTEGER PRIMARY KEY
	image_id	INTEGER NOT NULL UNIQUE
	uu_id	TEXT NOT NULL
	cafo_class	INTEGER NOT NULL
	animal_type	TEXT
	cafo_components	TEXT
	prop_img	INTEGER
	prop_cafo	INTEGER
	notes	TEXT
	user_id	TEXT
raw_labels		
	label_id	INTEGER PRIMARY KEY
	image_id	INTEGER NOT NULL
	uu_id	TEXT NOT NULL
	cafo_class	INTEGER NOT NULL
	animal_type	TEXT
	cafo_components	TEXT
	prop_img	INTEGER
	prop_cafo	INTEGER
	notes	TEXT
	user_id	TEXT
assignments		
	assignment_id	INTEGER PRIMARY KEY
	image_id	INTEGER NOT NULL
	uu_id	TEXT NOT NULL
	user_id	TEXT NOT NULL
	tag_count	INTEGER DEFAULT 0

Table A1 continued.

Table Name	Fields	Data Types
users		
	user_id	INTEGER PRIMARY KEY
	username	TEXT NOT NULL
	pwd	VARCHAR(255
s3_img_tile_associations		
	index	INTEGER
	uuid	TEXT
	geometry	TEXT
	s3_img_url	TEXT
	s3_img_geometry	TEXT
	src_crs	TEXT
model_params		
	model_id	INTEGER PRIMARY KEY
	cnn	TEXT NOT NULL
	class_function	TEXT NOT NULL
	class_model	TEXT NOT NULL
	oversampling_strategy	TEXT NOT NULL
	tts_seed	INTEGER
	training_augmentation	INTEGER NOT NULL
	augmentations	TEXT NOT NULL
	test_p	REAL NOT NULL
	train_size	INTEGER NOT NULL
	test_size	INTEGER NOT NULL
	kf_nsplits	INTEGER
	kf_repeats	INTEGER
	kf_seed	INTEGER
	testing_augmentation	INTEGER NOT NULL
	class_thresholds	TEXT NOT NULL
	optimal_class_threshold	REAL
	test_class_threshold	REAL NOT NULL
	n_cpus	INTEGER
model_performance		
	model_id	INTEGER PRIMARY KEY
	mean_kf_accuracy	REAL
	mean_kf_precision	REAL
	mean_kf_recall	REAL
	optimal_lambda	REAL
	optimal_class_threshold	REAL
	test_accuracy	REAL
	test_precision	REAL
	test_recall	REAL
	test_f1	REAL
	test_classreport	TEXT
	train_time	REAL
img_geodata		
	index	INTEGER
	uuid	TEXT
	src_crs	TEXT
	geometry_wkt	TEXT
model_decisions		
	index	INTEGER
	model_id	INTEGER
	image_id	INTEGER
	uu_id	TEXT
	cafo_class	INTEGER

Endnotes

1 De Witte, Melissa. 2019. Stanford scholars show how machine learning can help environmental monitoring and enforcement. April 8. Accessed November 2021. https://news.stanford.edu/press-releases/2019/04/08/how-machine-learning-can-help-the-environment/.

2 Handan-Nader, Cassandra and Ho, Daniel E. 2019. "Deep learning to map concentrated animal feeding operations." *Nature Sustainability* 2:298-306 https://doi.org/10.1038/s41893-019-0246-x

3 MacDonald, James, and William McBride. 2008. *The Transformation of U.S. Livestock Agriculture: Scale, Efficiency, and Risks.* Economic Information Bulletin Number 43, Washington D.C.: United States Department of Agriculture.

4 Government Accountability Office. 2008. *Concentrated Animal Feeding Operations: EPA Needs More Information and a Clearly Defined Strategy to Protect Air and Water Quality from Pollutants of Concern.* Report to Congressional Requesters, Washington D.C.: Government Accountability Office.

5 Food and Water Watch. 2020. Factory Farm Nation: 2020 Edition. Issue Brief, Food and Water Watch.

6 Government Accountability Office. 2008. *Concentrated Animal Feeding Operations*.

7 Miller, D. Lee, and Gregory Muren. 2019. *CAFOs: What We Don't Know is Hurting Us*. R: 19-06-A, Washington D.C.: Natural Resources Defense Council.

8 Government Accountability Office. 2008. *CAFOs*.

9 Hribar, Carrie. 2010. *Understanding Concentrated Animal Feeding Operations and Their Impact on Communities*. Bowling Green: National Association of Local Boards of Health.

10 Miller and Muren. 2019. *CAFOs: What We Don't Know*.

11 MacDonald and McBride. 2008. *U.S. Livestock Transformation*.

12 Hribar. 2010. *Understanding CAFOs Impacts.*

13 United States Geologic Survey. 2021. "Agricultural Contaminants." Accessed October 11, 2021. https://www.usgs.gov/mission-areas/water-resources/science/agricultural-contaminants?qt-science_center_objects=0#qt-science_center_objects.

14 Hribar. 2010. *Understanding CAFOs Impacts*.

15 Environmental Protection Agency. 2013. *Literature Review of Contaminants in Livestock and Poultry Manure and Implications for Water Quality*. EPA 820-R-13-002, Environmental Protection Agency.

16 Hribar. 2010. *Understanding CAFOs Impacts*.

17 Miller and Muren. 2019. *CAFOs: What We Don't Know*.

18 Hribar. 2010. *Understanding CAFOs Impacts*.

19 Preece, Sharon L. M., N. Andy Cole, Richard W. Todd, and Brent W. Auvermann. 2012. Ammonia Emissions from Cattle Feeding Operations. E-632, College Station: Texas A&M AgriLife Extension.

20 Miller and Muren. 2019. *CAFOs: What We Don't Know*.

21 Ibid.

22 Burnett, John. 2009. "New Mexico Dairy Pollution Sparks 'Manure War'." National Public Radio, December 9.

23 Eastman, Chelsea, Marc B. Schenker, Diane C. Mitchell, Daniel J. Tancredi, Deborah H. Bennett, and Frank Mitloehner. 2013. "Acute pulmonary function change associated with work on large dairies in California." *Journal of Occupational and Environmental Medicine* 74-79.

24 Lusk, Jayson, and Ranveer Chandra. 2021. "Farmer and farm worker illnesses and deaths from COVID-19 and impacts on agricultural output." *PLoS One*.

25 Douphrate, David, David Gimeno, Matthew Nonnenmann, Robert Hagevoort, Cecilia Rosas-Goulart, and John Rosecrance. 2014. "Prevalence of work-related musculoskeletal symptoms among US large-herd dairy parlor workers." *Am J Ind Med* 370-379.

26 Ibid.

27 Adcock, Flynn, David Anderson, and Parr Rosson. 2015. *The Economic Impacts of Immigrant Labor on U.S. Dairy Farms.* College Station: Center for North American Studies, Texas A&M AgriLife Research.

28 Leach, Ashley, and Mark Flaherty. 2014. New Mexico Agriculture. Regional Review, New Mexico Department of Workforce Solutions, Business Services Division.

29 Miller and Muren. 2019. *CAFOs: What We Don't Know*.

30 New Mexico Environment Department. 2021. Dairy Discharge Permitting. Accessed November 2021. https://www.env.nm.gov/gwqb/dairy/.

31 Government Accountability Office. 2008. *CAFOs*.

32 Miller and Muren. 2019. *CAFOs: What We Don't Know*.

33 Ibid.

34 Hagevoort, Robert. 2021. "New Mexico Dairy Industry Key Indicators (January 2021)." New Mexico State University Dairy Extension.

35 Ibid.

36 Ibid.

37 Burnett, John. 2009. "New Mexico Dairy Pollution."

38 Food and Water Watch. 2020. The Urgent Case for a Moratorium on Mega-Dairies in New Mexico. Factsheet, Food and Water Watch.

39 New Mexico Environment Department. 2021. Water Resources & Management. Accessed November 2021. https://www.env.nm.gov/water/.

40 Miller and Muren. 2019. *CAFOs: What We Don't Know*.

41 Nicole, Wendee. 2013. "CAFOs and Environmental Justice: The Case of North Carolina." *Environmental Health Perspectives* A82-A189.

42 Food and Water Watch. 2020. Moratorium on Mega-Dairies in New Mexico.

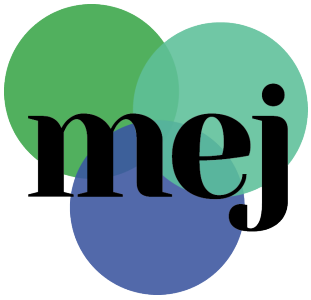
43 Iowa Department of Natural Resources. 2017. 2017 Annual Report for Work Plan Agreement between Iowa Department of Natural Resources and the Environmental Protection Agency Region 7.

44 Handan-Nader, C. and Ho, D. E. 2019. "Deep learning to map concentrated animal feeding operations."

45 Amazon Web Services. 2021. Registry of Open Data on AWS. https://registry.opendata.aws/.

46 Descartes Labs. 2021. Descartes Labs. https://descarteslabs.com/.

47 Counterglow. 2021. Counterglow. https://www.counterglow.org/.


mej