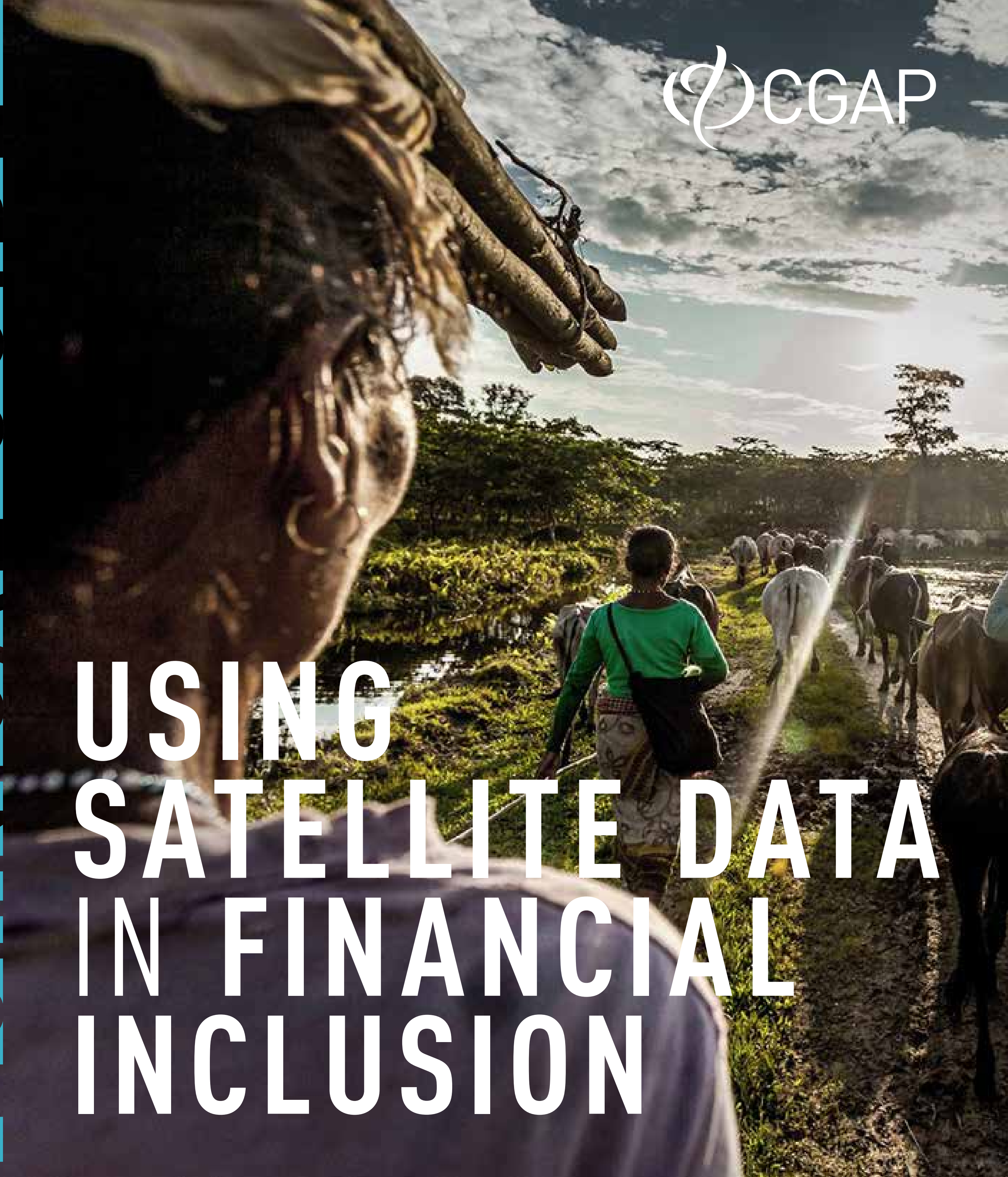
TECHNICAL GUIDE

CGAP



USING SATELLITE DATA IN FINANCIAL INCLUSION

How financial services providers can use satellite data and advanced analytics techniques to reach remote customers

July 2019

Maria Fernandez Vidal and Peter Bull

Consultative Group to Assist the Poor
1818 H Street NW, MSN F3K-306

Washington DC 20433

Internet: www.cgap.org

Email: cgap@worldbank.org

Telephone: +1 202 473 9594

Cover photo by Sujan Sarkar, India.

CONTENTS

EXECUTIVE SUMMARY

Financial services providers that see an opportunity to reach financially excluded people in rural areas can use new technology to remotely gather and analyze data on potential customers.

High-quality satellite data are becoming increasingly available. By leveraging advances in machine learning—the ability of computers to analyze data quickly and at scale—providers can gain valuable insights into customers' economic, environmental, and demographic characteristics. This guide explains foundational concepts of machine learning and how financial services providers can apply those methods to leverage information contained in satellite images for the purpose of credit scoring.

This guide focuses on smallholder finance, but providers may find it useful for other applications as well, such as estimating local infrastructure, housing, and income levels; assessing the effectiveness of farming practices; crop insurance and risk calculations; and forecasting yields to combat food security problems.

INTRODUCTION

THIS GUIDE FOCUSES ON THE USE OF satellite imagery for credit scoring algorithms that facilitate smallholder farmer finance. Those interested in other applications of satellite imagery will also find the guide useful. The guide covers the basics of machine learning and highlights the knowledge, skills, tools, and data sources necessary when using satellite imagery in machine learning.

The guide focuses on three types of organizations:

- **Financial services providers (FSPs) that serve smallholder farmers.** These providers can use models based on satellite images to estimate relevant metrics, such as timing and value of yields, at a low marginal cost.
- **Non-FSP organizations that have smallholder farmers as customers.** Other organizations in agricultural value chains—input providers, exporters, and traders—can use detailed information about type and density of crops or population to help them make decisions on products, yields, quality, or source locations.
- **Development organizations and public sector.** Development organizations and public sector actors that see the potential of financial inclusion are increasingly interested in big data and analytics. Those in agriculture and agrifinance are particularly interested in using satellite imagery.

This guide strives to do the following:

- Introduce remote sensing and its potential for financial inclusion and smallholder finance.
- Explain in simple terms how these methods work and clarify both the abilities and limits of current techniques.
- Present use cases where computation techniques applied to satellite imagery can help organizations better serve smallholder farmers.
- Equip organizations interested in exploring these methods with clear and actionable roadmaps that outline data prerequisites, problem scoping guidelines, and advice on getting started with research and development efforts.

Providers can use this guide as a tool to help them apply technologies, processes, and data analytics and machine-learning methods to improve the delivery of financial services to low-income segments.

SECTION **1**

THE OPPORTUNITY IN REMOTE SENSING

1.1 What Is Remote Sensing? Why Does It Matter for Financial Inclusion?

The availability of data is increasing exponentially. Over the past 15 years, there has been a dramatic increase in the availability of high-resolution satellite imagery data from commercial providers. These images make it possible to follow events on the ground—from changing infrastructure to the availability of water resources—that can provide insight into economic, environmental, and demographic changes over time.

Better tools make it easier to apply advanced analytics techniques. Advances in machine learning paired with an increase in affordable computational and storage capabilities of machines and networks provide new ways to leverage data. Machines can analyze imagery data quickly and at scale and can learn from patterns. For example, convolutional neural networks (CNNs) power applications like the auto-tagging of friends on Facebook or the computer-aided identification of irregular tissue in medical scans. Paired with the increased availability of high-frequency satellite images, these approaches help organizations learn about what is happening in remote locations in a scalable way.

Remote sensing can enable fast, affordable, effective, and scalable decision making. Remote sensing is the acquisition of information about an object or phenomenon without making physical contact with the object. It generally refers to the use of satellite sensor technologies to detect and classify objects on Earth. Satellite imagery allows organizations to make smart decisions at scale, while the use of traditional methods can be expensive, infeasible, and complicated. There are three important benefits of remote sensing when compared with other data collection options:

- It is often cheaper
- It can often provide intelligence more quickly
- It can scale to enormous geographic areas much more effectively—thus providing large time and cost savings benefits.

Reducing the marginal costs of operation is key for financial inclusion. One of the big challenges of financial inclusion business models is the low revenue potential per customer. Therefore, low marginal costs are a key enabler of viable business models to serve the poor. The availability of digital channels has been a first step in lowering the costs of traditional operational models and expanding reach. However, once an in-person, one-on-one relationship is replaced by a digital one, organizations lose opportunities to know their customers. Remote sensing technology, although still nascent, has the potential to enable individualized decision making in an automated way by providing new customer-level data.

Remote sensing technology has significant potential for organizations that work in financial inclusion, especially financial inclusion for smallholder farmers. Collecting data about farmers who are spread across remote areas is a slow and expensive process. Most smallholder farmers do not have formal financial information that can be used to assess their income flows or ability to repay. Remote sensing can help organizations that serve rural segments (e.g., agricultural services, microfinance institutions, off-grid energy providers) to collect and analyze data and better assess opportunities and risks in a scalable way.

1.2 Areas Where These Methods Can be Applied

The following use cases illustrate the potential for satellite imagery and machine learning to impact development.

Alternative credit scoring. Low-income people in general, and rural populations in particular, often do not have traditional credit histories and financial records, which makes it difficult for FSPs to assess their creditworthiness and to lend to them. Satellite images can provide an estimate of past and future agriculture income as well as the timing and sources of this income, and thus, they can provide key inputs to credit assessments.

Estimating local infrastructure, housing, and income levels. Reliable estimates of local infrastructure, housing, and income levels can help development organizations identify gaps and needs for investment and can help private-sector players identify markets and opportunities. Satellite data can enable assessment of large areas, effective comparisons across locations, and tracking of changes and evolution in a cost-effective way.

Assessing the effectiveness of farming practices (including inputs, methods, and crops). Satellite images can enable automated assessment and guidance on farming practices at a minimal marginal cost. This may be particularly beneficial to subsistence farmers who lack access to formal training or advisers.

Calculating crop insurance and risk. Insurance can smooth income and provide resilience against shocks, which is particularly important for low-income people who have limited assets and savings. Remote sensing has the potential to automate insurance claim evaluation, which would significantly reduce marginal cost and enable business models for microinsurance.

Forecasting yields to combat food security problems. Food security is a major challenge for many developing countries that have large population segments that depend on rainfall agriculture. This is exasperated by uncertainties stemming from climate change and the lack of early detection of low yields, which limits the ability of governments and other organizations to mitigate shortages. Remote sensing can provide automated high-frequency yield estimates to maximize their ability to react and address potential food shortages.

Although many of these use cases are new or experimental, they are becoming increasingly accessible as technology becomes cheaper and more widespread.

BOX 1. **Key terms**

Data science. A broad term that encompasses collecting, storing, processing, analyzing, and communicating data. In this paper, "data science" also refers to machine learning and statistical modeling.

Deep learning (also known as deep structured learning or hierarchical learning). This is part of a broader family of machine-learning methods based on learning data representations, as opposed to task-specific algorithms. Among deep-learning methods, convolutional neural networks (CNNs) are the most common approach used for image recognition.

GIS (geographic information systems). A system of tools that are used to store, process, analyze, and map geographic data. These systems are complex, and specially trained experts are needed to run them.

Machine learning. This term refers to using data to "give computers the ability to learn without being explicitly programmed." Much of machine learning relies on statistical models or other algorithms used to generate results from information found in data rather than in predefined explicit rules.

Prediction. Making inferences about unknown information based on data inputs. When we talk about predictions, we are not necessarily talking about the future. For example, a model can try to predict crop yields in the upcoming year or it can try to predict whether a given image is of a cat or of a dog.

Satellite imagery. Data collected from satellite-mounted sensors of different light spectra as they circle the Earth. These can simply be images at visible light, but they may also include images resulting from nonvisible wavelengths (such as infrared or ultraviolet) detected by specialized sensors.

Statistical model. Refers to an algorithm that translates input data—in this paper, this is usually satellite imagery—to a predicted result of interest. The models discussed in this paper are generated using machine-learning techniques, but statistical models can be generated using a variety of econometric techniques.

SECTION **2**

DEVELOPING YOUR REMOTE SENSING CAPABILITIES

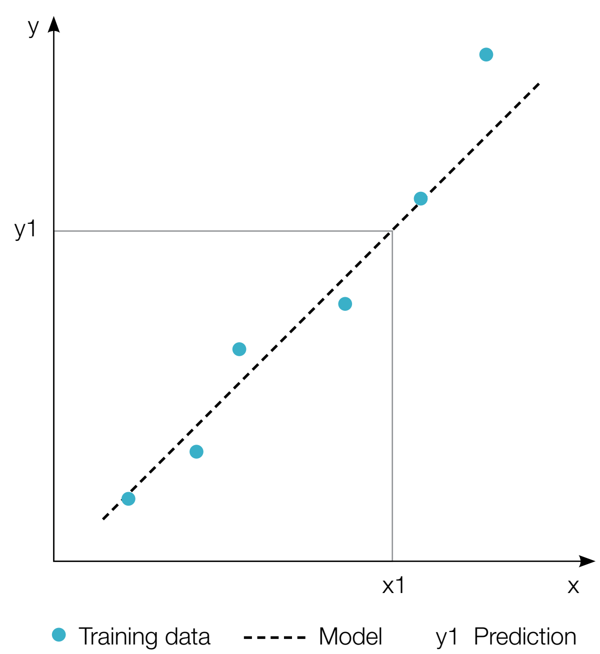
2.1 What Is Machine Learning?

Machine learning has been around since the early 20th century, and it has become increasingly relevant as advances in technology reduce the cost of the computational and storage capacity of computers. Traditional applications are set up by defining explicit rules: "if X, then Y." For example, you can tell a computer how to play tic-tac-toe (noughts and crosses) by defining rules for all possible scenarios: "If you selected top-right and player B selected top-center you should select middle-center next." For a simple game with a limited set of possible moves like tic-tac-toe, defining explicit rules is relatively easy to do. However, in some cases (e.g., a more complex game like chess), it is difficult, time consuming, and/ or expensive to write out all the rules for every situation. You may not know what the rules are ahead of time—for example, you may have data about what happened in a given situation and what the outcome was, but you may not know the steps of "if X, then Y" that produced the outcome. In this case, you want the computer to learn those rules by examining many examples or data. In the tic-tac-toe example, you would provide the computer a data set of several tic-tac-toe games that contains the sequence of selections of each player and the outcome: Player A wins, Draw, or Player B wins. The computer can use that information to learn how to play in a way that maximizes its chances of winning. This is called machine learning.

Linear regression, a standard econometrics technique used for modeling, is a simple example of machine learning (although the term is most often used to refer to more complex examples). Given two points, you can unequivocally determine the straight line that best fits them: the line that passes though both points. However, the procedure of "fitting" or "training" this line becomes increasingly complex as the number of points increases and you need to make trade-offs between being closer to one or to another.

Once you have a "fitted" or "trained" model, you can use that model to make predictions. The trained model is a formula that, for any combination of inputs, provides the most likely outcome, even if it was never trained on this exact combination of inputs. (See Figure 1.)

FIGURE 1. **An example of fitting training data**



The two basic procedures of machine learning are training a model and using that trained model to make predictions. Models can also be retrained to incorporate new data and improve predictions. This type of computation can be done in seconds by any basic computer, so the term "machine learning" tends to be used for more complex estimations involving unstructured data. For example, looking at crop yield, you can train a model that takes satellite images as inputs and generates a prediction of the crop yield of the fields in the photographs. The model can be fine-tuned with other data such as weather forecasts.

2.2 Technical Capabilities

Data science is an inherently cross-functional and interdisciplinary field, where projects ordinarily cut across many parts of the business, both technically and operationally. Successfully combining technical and operational knowledge—either by finding people with expertise in both or more commonly by achieving effective collaboration in a multidisciplinary team—is often the biggest challenge for organizations that want to use advanced analytics.

FUNCTIONAL ROLES

Functional roles define capabilities and functions rather than titles and people. A single person may have one or more roles depending on his or her expertise and the size of the operation.

- **Data scientist with deep learning experience.**[1] The core skill set for any satellite imagery machine-learning project is the ability to obtain, clean, and analyze very large heterogeneous data sets. This involves both quantitative and programming skills, including familiarity with data tools and software. At a minimum, this requires the following:
  - Strong knowledge of statistics, applied math, and computer science, usually with a masters (or higher) degree in a quantitative discipline.
  - Mastery of at least one programming language commonly used for data work. Prevailing languages used by data scientists are Python and R (R Development Core Team 2008).[2] Both of these languages are good options for modeling with satellite images.[3]
  - Experience with deep-learning methods (preferable computer vision approaches used for image analyses like CNN).
  - Experience with computer vision principles or projects,[4] where image data are analyzed using CNNs or other relevant approaches.
- **GIS practitioner.** Like data scientists, GIS practitioners specialize in cleaning, synthesizing, and analyzing varied data sets. They are experts in and focus mainly on geospatial data analysis. They generally do not need advanced mathematical backgrounds, and while many are able to script and automate certain aspects, they often use desktop environments such as ArcGIS, Mapinfo, or QGIS for their analyses rather than custom software.
- **Data engineer.** The data engineer has a traditional IT role that focuses on systems for storing and processing data. The work of the data engineer—sometimes called an infrastructure engineer—overlaps with work in development operations. The data engineer decides how to develop and manage the systems infrastructure. The data engineer needs the following:
  - An ability to strategize about "cloud" and "on premise" architectures. (See Section 4.1, IT Infrastructure for Machine Learning.)
  - Experience with storing and processing images and geospatial data, both of which have specialized database technologies.
  - Experience using application programming interfaces (APIs) to extract data and metadata from satellite providers.
  - Experience scaling systems to handle massive amounts of information (in the terabytes) of input data.
- **Business expert.** The business expert is a project manager who understands the business use case and is familiar with the "ground truth" data (see Section 3, The Data). The expert works closely with the data scientist to resolve questions about patterns in the data. Importantly, the expert

1 Because this role involves several technical skill sets that intersect several fields, it can be difficult to find qualified data scientists, and it can be expensive to bring them on as full-time employees—see "Building Capacity vs. Outsourcing," later in this section for advice on getting started without full-time data scientists.

2 Python Software Foundation, "Python Language Reference," http://www.python.org.

3 Some other data analysis languages such as SPSS, SAS, and MATLAB are not as effective and flexible when working with satellite imagery.

4 Computer vision is an interdisciplinary field that uses computers to gain high-level understanding from digital images or videos.

understands the area depicted in the satellite images and knows which patterns are expected and which are unusual. The expert understands the business case for using satellite imagery and machine learning. The business expert needs the following:

  - Domain expertise and knowledge of the business use case.
  - Understanding of the "ground truth" data that are being modeled. For example, how are the data collected? What are the anomalies? Where are the missing data? What do individual columns in the data set mean?
  - Ability to communicate across different functions and share progress, questions, and successes.

- **Senior owner or "champion."** Machine-learning projects are, at their core, research and development projects. A senior-level project owner will ensure that expectations around the work are set properly and will demonstrate that an organization understands the long-term buy-in that is necessary for success. The senior owner or champion needs to:

  - Support the project as a research and development effort that uses cutting-edge technology to solve business problems.
  - Have a firm grasp on the domain and understand what the application of the system will be—in particular, how the system will improve the business and its operations.
  - Ideally, have a technical background.

Even projects with brilliant technical staff may fail if staff cannot rely on a subject matter expert to keep the focus on issues that matter to commercial and other stakeholders and to keep assumptions realistic. Similarly, a cutting-edge research and development project that achieves impressive numerical results but cannot feasibly be run on the organization's existing IT infrastructure will not be adopted and used. Even successful and objectively high-quality technical efforts tend to flounder in larger organizations without a senior project leader acting as a champion and helping to break down barriers between disparate business units.

BUILDING CAPACITY VS. OUTSOURCING

Should you build machine-learning models with satellite imagery in-house or should you outsource this effort? Your answer depends on many aspects, including whether you have the expertise required in-house, what your project goals are, the scope of your budget, your timing requirements, and your long-term strategy.

Will machine-learning capabilities strengthen your position in the marketplace? If yes, do you have or can you find, attract, and support the right person to lead this effort? This person needs to understand the business as well as how machine learning can address business needs. The lead will need to work in a dynamic environment and invest resources in areas that have the most potential for success. However, even well-informed decisions do not guarantee success.

If you do not have the necessary conditions in-house, you may want to outsource the work. Outsourcing often means getting results faster and having lower risks because you would not need to take the time to train in-house specialists. This is especially valuable at the research and development phase, where future investments are often based on the results of initial experiments. By outsourcing this work, your organization can more effectively manage timelines, prices, and goals by negotiating these upfront with the contractor.

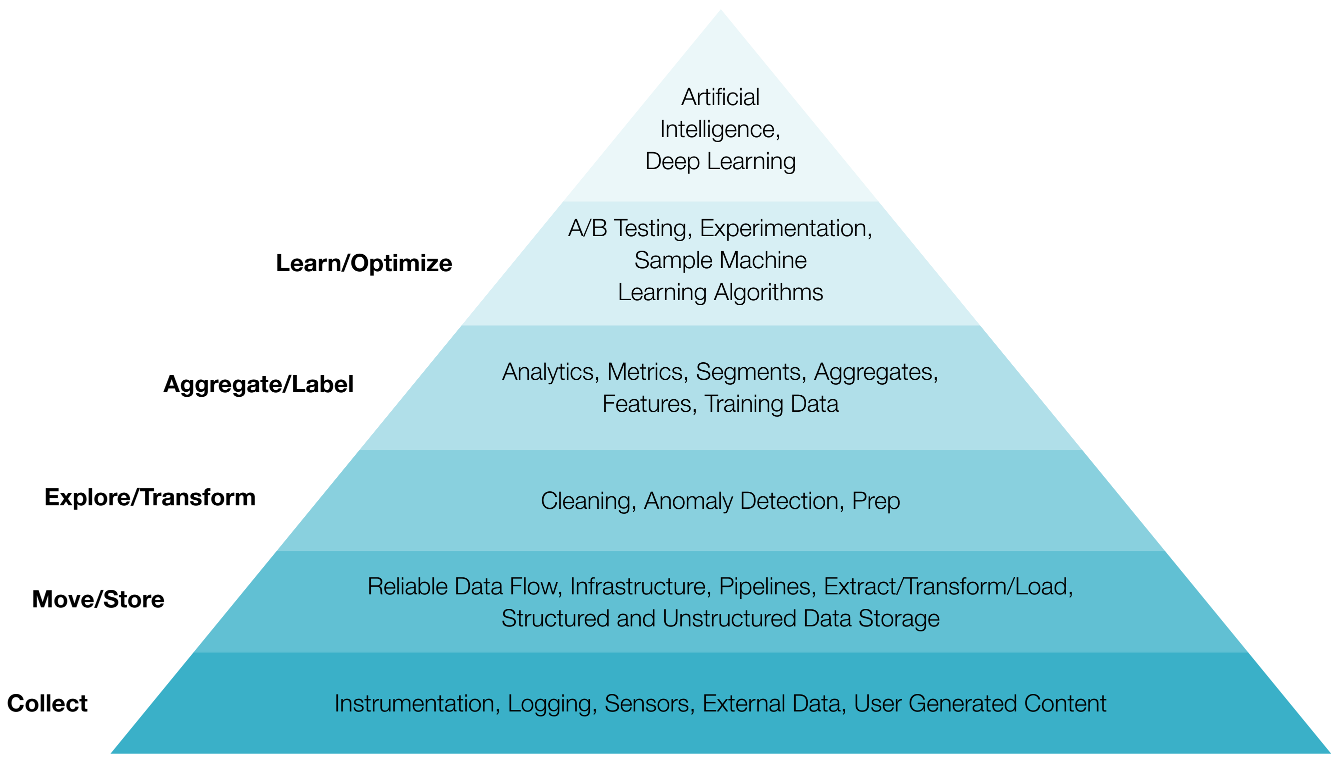
If you decide to outsource the work, you should clearly define the success (and failure) criteria of the project. Machine learning is about experimenting with data and methods, and not every experiment will work. Learning why something did not work often can be as valuable as creating a working model. This caveat should be acknowledged when assessing the value of machine learning. Creating models with extremely high accuracy involves an iterative approach and learning from both successes and failures.

Each step should be built on a solid foundation—this is referred to as the Data Science Hierarchy of Needs (after Maslow's Hierarchy of Needs, see Figure 2). When outsourcing a machine-learning project (or defining one in-house), you should review the Data Science Hierarchy of Needs with the contractor and have a frank discussion about the effort required at each level. The more information the contractor has, the more likely it is that together you can scope a successful project.

In most cases, an investment in a machine-learning model is an investment in custom software. There are currently no off-the-shelf software providers that use state-of-the-art models and near-daily satellite data. Therefore, you will need to invest in custom software, keeping in mind the need for maintenance, improvements, support, and integration of that software.

Overall, the decision to build internal capacity or to outsource comes down to your organization's unique situation. Your organization must assess strategy, current capabilities, and the expected business benefits of this technology to determine the best path forward.

FIGURE 2. **The data science hierarchy of needs**



Source: Rogati, 2017.

SECTION **3**

THE DATA

COLLECTING DATA IS AT THE BASE OF the Data Science Hierarchy of Needs. One of the mantras of the data scientist is: "Garbage in, garbage out." Without accurate, timely, and relevant data you cannot build effective models. A plan for getting the right data, in the right format, and within a reasonable timeframe is critical.

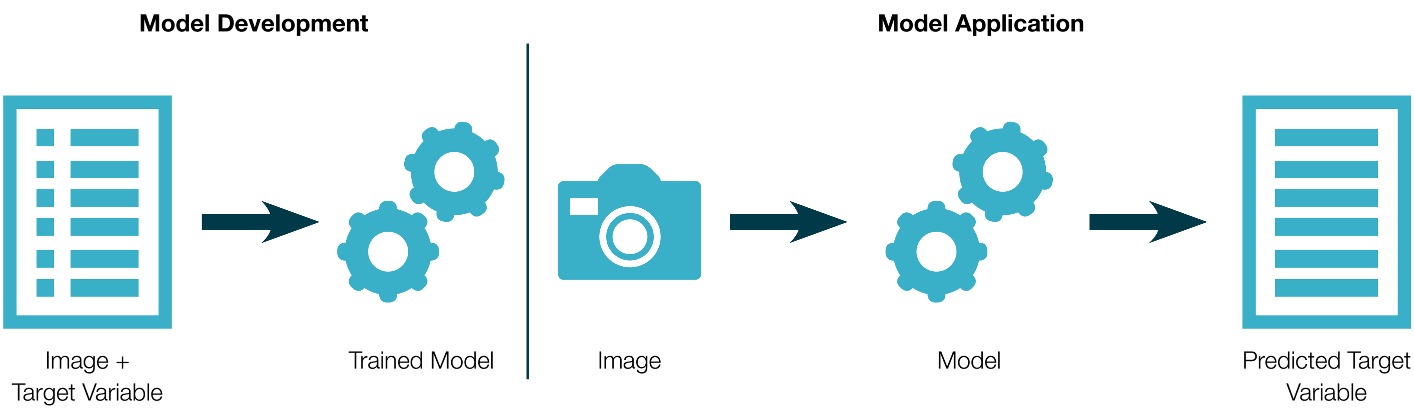
In the case of remote sensing, data need to capture three essential features: what, when, and where. While critical for working with satellite data for smallholder farmers, these features are also relevant for many other types of image and remote sensing data.

A data audit is an essential first step. In addition to ensuring that the necessary data are collected, you must make sure that data are of sufficient quality. The data scientist should provide those familiar with the data (the business expert, the data engineer, and others within the company as needed) with descriptive statistics for the different variables and identify any potential errors. Descriptive statistics may include means, medians, maximum and minimum values, distribution of data, and correlations between variables.

3.1 Selecting a Target Variable to Model (What)

The machine-learning process creates a data-driven model that takes contextual information (such as satellite images) as inputs and produces predictions about another variable.[5] For instance, a model may predict an estimate of crop yield, the rate of deforestation, the density of road traffic, or the total irrigated land. The key initial step is to produce a data set that has the variable to be predicted. This is called the "target variable." You want your model to be a function that takes images as inputs and produces estimates

FIGURE 3. **Model development and application**



5 Technically, this is a subfield of machine learning called "supervised learning." Supervised learning is the focus of this paper, but there are also methods called "unsupervised learning." These methods attempt to discover structure in data independent of a quantity of interest or variable you are looking to measure.

of this target variable. For example, to train an algorithm to predict yield based on an image, you need a data set that contains images and the yield associated to each of them. (See Figure 3.)

In the data set, this target variable will have a temporal component (when it was measured) and a spatial component (where was it measured). Knowing when and where your measurements of the target variable comes from allows you to synchronize your target data with the available satellite imagery.

CLASS OF MODEL

Classification, regression, and segmentation are the three classes of machine-learning models that are relevant when working with satellite imagery. These classes have different kinds of target variables, but all can be modeled using satellite imagery if there are enough examples of satellite imagery where the value of that target variable is known.

In a classification problem, there is a set of discrete categories to model. Figure 4 illustrates a simple example where a photograph needs to be classified as being of a cat, dog, mug, or hat. The kind of object in the picture is the target variable. The goal of classification is to identify which class (category) the image should be associated with. For satellite imagery, the algorithm might, for example, classify a satellite image as “urban” or “rural.”

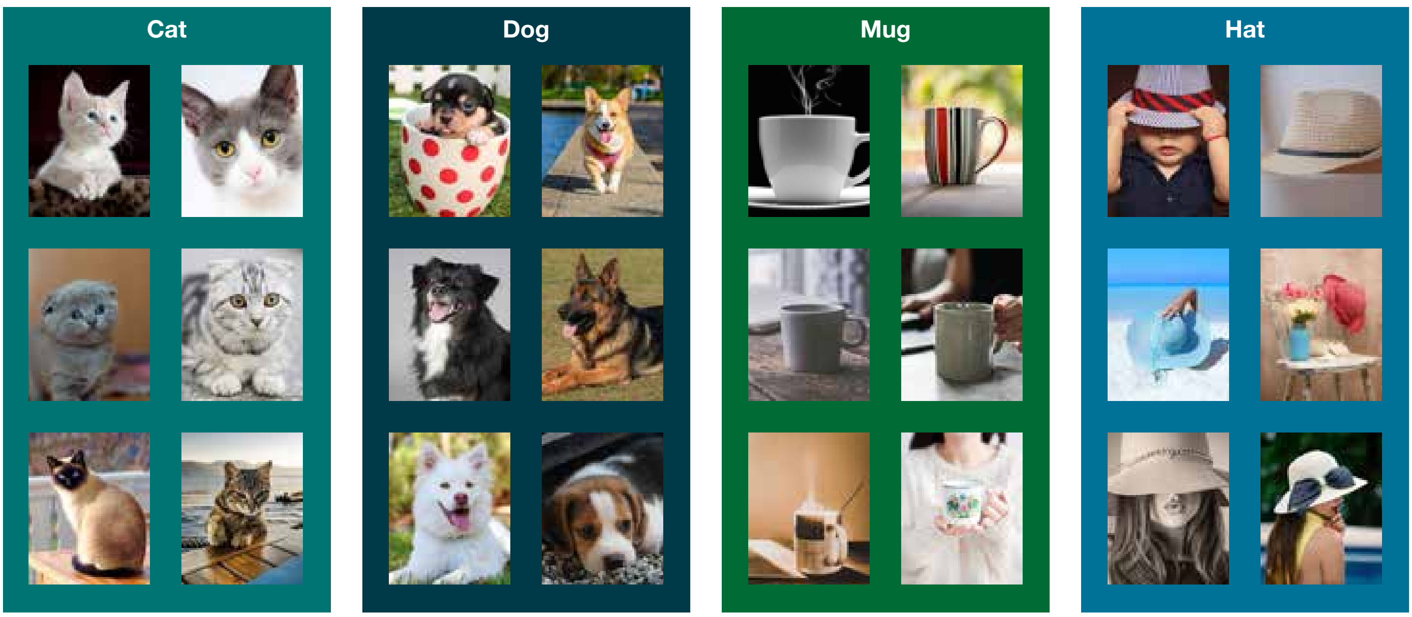
In a regression problem, a continuous variable is modeled (as opposed to a discrete category). The algorithm will predict a specific value at a given time for this continuous variable. The target variable is usually a number representing the quantity of interest. For example, a regression model might predict future house prices, where house prices can have any dollar value—not just “cheap” or “expensive.” Some examples using satellite images include changes in the number of buildings in a suburb, expected crop yield from an area of farmland, or the percentage of the image covered in water.

In an image segmentation problem, areas of interest within an image must be identified. For example, an algorithm might draw bounding boxes around cultivated land in the image or it may identify areas that are bodies of water. Image segmentation is closely related to object recognition. As Figure 5 shows, for this kind of data, example images are needed where the objects of interest have already been labeled with bounding boxes. Sometimes this can be done by manually adding these annotations to a subset of the satellite images so that an algorithm can be built to automate that process.

LEVEL OF AGGREGATION

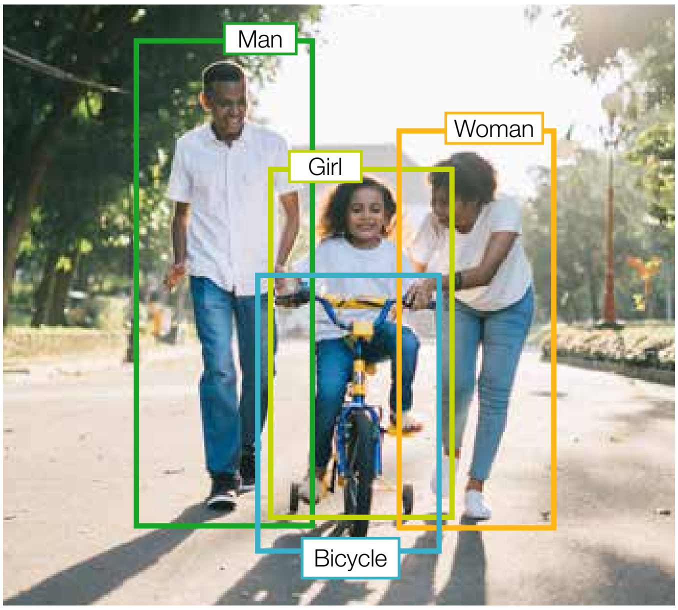
Next, the level of aggregation in the data set must be considered. The level of aggregation determines what a single row in the data set represents. When considering crop yield, there may be yield measurements at a per-farm level (e.g., Farm X produces 1,200 tons of maize). It is also common to have per-county, per-state, or per-country estimates of crop yield. The level of aggregation is considered “fine grained” if

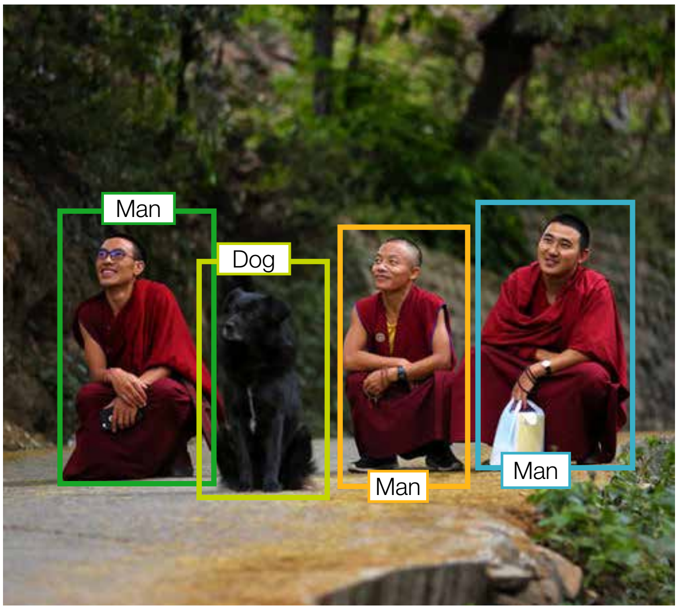
FIGURE 4. **Example of different classes for images.**



Source: Based on Karpathy, 2017.

FIGURE 5. **Example of output from an object recognition algorithm**





it is at the level of an individual and "coarse grained" if data are aggregated into larger units.

It is important to consider the level of aggregation when assessing project viability. In most cases, it is impossible to create an algorithm that makes accurate predictions at a more fine-grained level than that available to train the model. For example, if there are only crop yield measurement per state, it is not possible to build an algorithm to predict yields per county. However, the opposite, aggregating per-state predictions into per-country predictions, is possible.

Machine-learning models work best when the measurements are fine grained. There are usually more observations for fine-grained data. For example, there may be 15 states in a country, but there may be thousands of farms in those states. Having data per farm means that the algorithm can learn from many more examples. This also allows more accurate and detailed predictions, such as per specific farm. Fine-grained predictions are usually much more useful to stakeholders. It is important to note that both images and target data will need to be fine grained to fully capture their predictive power. For example, if images are at the individual farm level but estimates of yield are available only at the county level, the model's ability to predict will be weaker than if it were trained using farm level yields for each picture.

TARGET VARIABLES FOR SMALLHOLDER FARMER PREDICTIONS

Remote sensing is most needed when it is difficult or expensive to collect data in other ways, such as for smallholder farmers, many of whom live in remote areas. There are many variables of potential interest about smallholder farmers. Some of these variables are explored in the following, particularly those that may be good targets for machine-learning projects.

Economic target variables

- Farmer income
- Farmer repayment of loan

Agricultural target variables

- Farmer crops grown
- Crop mix
- Crop yield
- Soil properties
- Irrigation

The quality of these data can vary significantly; data that are self-reported (rather than measured directly) are less reliable. Some data may need to be collected, while other may be available directly from third-party sources, either commercially or publicly available. The best machine-learning insights come from data sources that are objectively measured from reliable sources.

For crop yield measurements, a crop-cutting experiment is the gold standard. In a crop-cutting experiment, a trained enumerator selects a predetermined area from a field to cut some of the crop and measure the yield. The measurements are then extrapolated for the total area of the field. For more information about the crop-cutting procedure, see "Crop Cutting Manual" from the Central Statistical Agency, Ethiopia, and the World Bank (2013).

3.2 Farm Location Data (Where)

Location data are needed to segment satellite images into areas of interest. Without specific areas of interest, working on a large geographic scale may result in very noisy models that do not deliver useful predictions. Knowing exactly where farms are located is one of the biggest hurdles in building models for smallholder farmers. These data are often expensive and time-consuming to collect, and their accuracy can be hard to verify. There are several ways to gather or approximate this data with varying levels of accuracy (see Table 1).

Different resolutions are appropriate for different modeling tasks. For example, when creating a model to determine the crop mix on a certain farm, a reliable model cannot be built unless individual farms can be identified. On the other hand, when predicting the likelihood of paying out crop insurance in a drought, understanding the impact of the drought in a larger geographical area may be sufficient.

3.3 Temporal Data (When)

The next question to ask is, "When were the key measurements taken? Depending on the questions, specific dates may even be of interest—for example, did a recent wildfire burn a certain field or not?

There are four basic categories of temporal data: event based, seasonal, trend over time, and static. Understanding

TABLE 1. **Location data by resolution**

	Approx. Spatial Resolution	Location Collection Mechanism	Comments
Most useful	< 1m	GPS measurement of a field perimeter (also called GPS polygons)	By far the most effective, but expensive, method of locating farmers. A surveyor with a GPS unit (e.g., a smartphone) walks the perimeter of each field with the GPS device. This level of detail is important for modeling variables that vary at the level of an individual field.
Very useful	10–100m	A single GPS coordinate for fields and approximate field size	This can be useful, but in practice data quality must be validated. For example, surveyors may record the point from a nearby road, but fail to note which side of the road the field is located.
Useful	100m–10km	Mobile network operator provided cellular tower pings	Towers have highly variable coverage areas based on power, location, obstructions, cell traffic, and other factors. Precision can be improved by triangulating the signal among towers of overlapping coverage. However, cellular coverage in rural areas is often limited. A major hurdle in providing these data is getting the cooperation of mobile network operators.
Somewhat useful	1–50km	Triangulation based on the time it takes to walk to 3 known locations in opposite directions	While easy to collect, e.g., with an SMS survey, it is hard to ensure the accuracy of these estimates. Furthermore, surveyed individuals cannot always provide locations that can be geocoded to a certain latitude/longitude (e.g., through a service like Google Maps). Walking times vary based on the individual.
Marginally useful	10–100km	A village or market name with exact geolocation	It is often easier to get the single closest geocodable location. However, at this resolution it becomes impossible to pick out individual farms. Corresponding statistics must be aggregated to a geographic region.
Limited usefulness	> 25km	The smallest official administrative unit for the country	This is usually the most reliable self-reported area. However, these administrative units can vary widely in size, and working with individual farms is impossible.

TABLE 2. **Types temporal data**

Temporal Category	Description	Examples	Considerations
Event-based	An event that happens at a discrete point in time.	Natural disasters, including wildfires, floods, and earthquakes; construction of individual buildings; opening of new roads and bridges.	Imagery must be matched exactly to the time window during the event to ensure reliable training.
Seasonal	A variable that changes seasonally, but year-to-year variation is relatively low.	Seasonal rains, crop harvests, tides, snow cover.	It is important to sample imagery during each point in the cycle. The year of the imagery does not need to match the year of measurement.
Trend over time	A variable that has a trend over time that is not seasonal.	Urban density, soil erosion, deforestation.	Effective data sets usually include several images over the time scale that the quantity changes. For example, for urban density, images of city edges over several years would be relevant.
Static (consistent over time)	A quantity that is relatively constant over a long period of time (long enough that only minor changes happen over the course of decades—the first images from satellites are from 1964).	Geological features, including mountains, valleys, grasslands, and deserts; paths or large bodies of water.	Matching the exact timing of the imagery for these kinds of quantities is usually not necessary.

changes in a target variable over time determines how frequently satellite coverage of an area is needed (see Table 2).

Exploring crop yield is a seasonal phenomenon based on the agricultural cycle, and yields vary year-to-year based on agricultural inputs used and weather. To build models that take these changes into account and offer robust conclusions, the data set must have enough temporal coverage to provide a representative sample of these conditions for the region of interest. When modeling farm productivity, data should cover several growing seasons, and data should be sampled at several points during the growing season.

3.4 Volume of Data (How Many)

The volume of data required to build a model depends on the task. There are two rules of thumb that can help you determine whether a project is feasible.

The first is that classification requires fewer examples than regression. It may be the case that the task can be reformulated as a classification problem. Consider crop yield, for example, and the goal of providing agricultural support for farmers with insufficient yield. If exact historical data are not available to calculate expected yield predictions, a cut-off can be created that determines whether the yield per acre is sufficient or insufficient. Because the goal has been reformulated as a classification problem, fewer examples are needed to build a reliable model.

The second is that some tasks can achieve higher accuracy through "transfer learning." Using ImageNet results to build computer vision models is one example. ImageNet is an academic data set that contains hundreds of thousands of images of everyday objects (e.g., cats, dogs, cars, balls) with the corresponding labels. Research has shown that a process of (i) training a model on ImageNet and then (ii) retraining just a small portion of the model on a specific task can produce highly accurate models. The theory is that the ImageNet training learns general features about distinguishing objects—for example, recognizing shape boundaries or different textures. By making small updates to the model, this general ability can quickly achieve good performance using a significantly smaller number of samples.

Transfer learning has been documented as effective for remote sensing, even using models that were originally trained to recognize everyday objects (Penatti, Nogueira, and dos Santos 2015). By using transfer learning, researchers often can get accurate results quickly without having to collect enormous amounts of data. As a result, transfer learning is recommended for building models for satellite imagery.

While transfer learning substantially reduces the amount of data needed, having sufficient data often means having thousands or tens of thousands of data points that represent the general population. When considering what makes a sample representative, there are some general considerations, including the following:

- **Time.** The sample should match the period when the target variable was measured. For example, satellite images from 2012, traffic congestion from 2012. It's important to have several years of data for analysis.
- **Seasonality.** Changes over the course of a year often have a significant effect on target quantities or image appearances. For instance, crops look different during different parts of the growing season, so images should be captured during different parts of the crop cycle. An exception is if predictions (e.g., crop yield) are to be made at the same time each year, in which case the satellite images should be sampled from that prediction timeframe (e.g., making predictions one month before harvest every season).
- **Geographic distribution.** The training data should match the geography where predictions will be made. Given agricultural and environmental differences, it can be difficult for models to work across geographies, even for the same crop.
- **Scale of operation.** The sample should contain examples from across the scales of operation that matter. For example, models trained on industrial scale farms will not work effectively for smallholder farmers.

The following concerns are more specific to agricultural applications:

- **Crop variety.** The kinds of crops (e.g., maize, potatoes) that the model is trained on should match the crops that the model will need to make predictions for.
- **Irregular fields.** Fields are often irregularly shaped. Using training data for large, uniform fields will not generalize well to small, irregular fields.
- **Intercropping.** Smallholder farmers often practice intercropping, where several different crops are grown in a single field. Understanding the prevailing intercropping practices and incorporating these data into the model (e.g., enabling the encoding of percentages of fields or including information about common combinations of crops) will result in models that more closely reflect realities on the ground.
- **Differing practices.** The sample should contain predominant agricultural practices in the region of interest.
- **Agricultural inputs.** There is often variance in the agricultural inputs used by individual farmers and their ability to pay for these inputs. For instance, trying to apply models trained in higher-income areas to fields in lower-income areas may fail to account for the differences in inputs, even if the land is similar.
- **Soil quality.** Soil quality can vary across areas, and this can have a profound effect on yield. Hence, soil quality should be included in the sample where possible.

Generally, a few thousand appropriate images that cover the sample of these variables are sufficient to start testing a machine-learning model.

3.5 Satellite Data

There are several important points to consider when selecting and using satellite imagery.

- **Resolution**—the number of meters an individual pixel captures. Greater resolution means fewer meters per pixel and greater detail in the images. Commercial providers generally have 5 m to 0.7 m resolution, whereas government providers tend to have resolutions of around 30 m. The resolution is often different for different spectra.
- **Frequency**—how often the area of interest is imaged. Some providers aim for more frequent, lower resolution coverage. Some have areas that are imaged daily, where others get new images of an area only a few times per year. Some providers have less frequent coverage but high-resolution images.
- **Coverage area**—the surface area of the globe that the satellite provider makes available. Not all areas are regularly imaged, so it is important to verify that the specific area of interest is covered at the desired frequency.
- **Temporal coverage**—images that cover several different seasons and samples from throughout growing seasons to build agricultural models. Some newer providers may have only limited historical imagery.

- **Spectra**—visible spectrum images include red, green, and blue spectra. In addition to the visible color spectra, some providers supply grayscale (monochromatic) images at higher resolutions. Providers also supply infrared spectra—including shortwave infrared and near infrared. Infrared can be useful for agricultural applications because it can effectively capture chlorophyll content in plants.
- **Cloud cover**—clouds included in satellite images. It is important to have strategies for when the area of interest is obscured by clouds. This is a common concern when working with satellite imagery. Many image providers offer the possibility to search by the percentage of cloud cover in an image. Having flexibility around exact dates and selecting images with low cloud cover can improve model accuracy. Additionally, some providers supply "cloud masks"—binary labels of every pixel as a cloud or not a cloud. This lets users more easily remove clouds from images or ignore these pixels in their models.
- **Paid/Free**—some government agencies—for example NASA's LANDSAT, TRMM, and MODIS missions and the European Space Agency (ESA)—provide free access to satellite data. For example, Sentinel-2, an Earth observation mission developed by ESA, provides images at 10 m, 20 m, and 60 m resolution at five-day intervals and has a free and open data policy. Companies such as DigitalGlobe and Planet provide satellite imagery commercially, often at a higher resolution or frequency than images available for free.

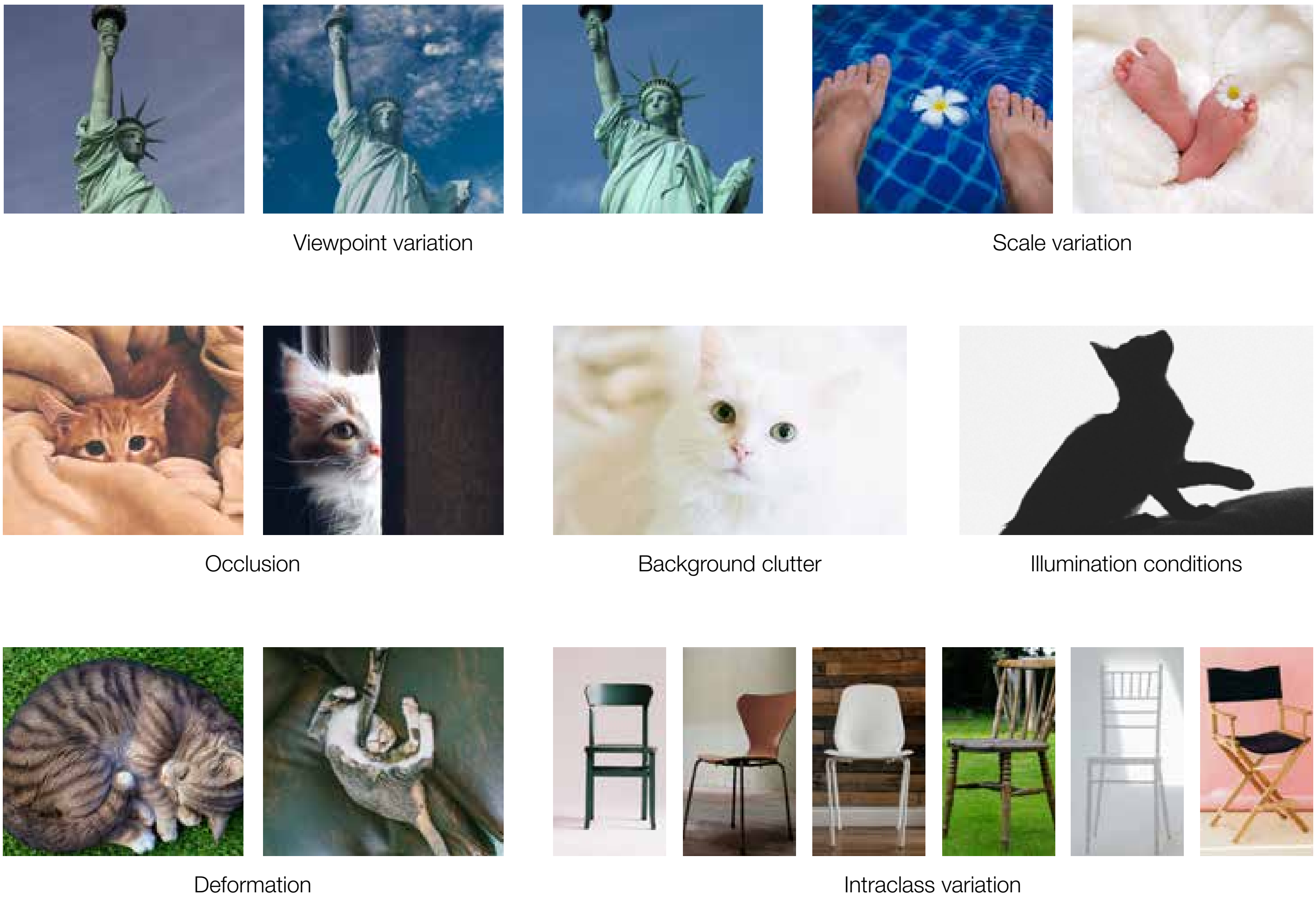
Assessing these criteria for a specific use case should help you choose the satellite imagery provider that fits the needs of your project.

SECTION **4**

THE MODELING

THIS SECTION DISCUSSES THE ALGO-rithms and approaches that are most effective for machine learning with satellite images. It will focus on a class of models referred to as “deep learning.” These models are considered the most effective machine-learning approaches for working with images. Deep learning models are based on neural networks, a computational technique inspired by the human brain. Recent work has shown that neural networks outperform other machine-learning models in many domains, including speech recognition, image classification, and natural language processing (Graves, Mohamed, and Hinton 2013; Szegedy et al. 2016; Jozefowicz et al. 2016). In particular, CNNs provide a way to incorporate both pixel-level information and spatial relationships of pixels into flexible functional forms, making them well suited for a wide array of tasks. (See Figure 6.)

FIGURE 6. **Deep learning models can deal with many types of challenges in imagery**



Source: Based on Karpathy, 2017.

Moreover, this class of model has been demonstrated to outperform traditional methods in remote sensing with satellite imagery (Nogueira, Penatti, and dos Santos 2017). CNNs are likely to be the dominant model for doing remote sensing work in the next 5–10 years, and organizations that rely on older techniques stand to fall significantly behind.

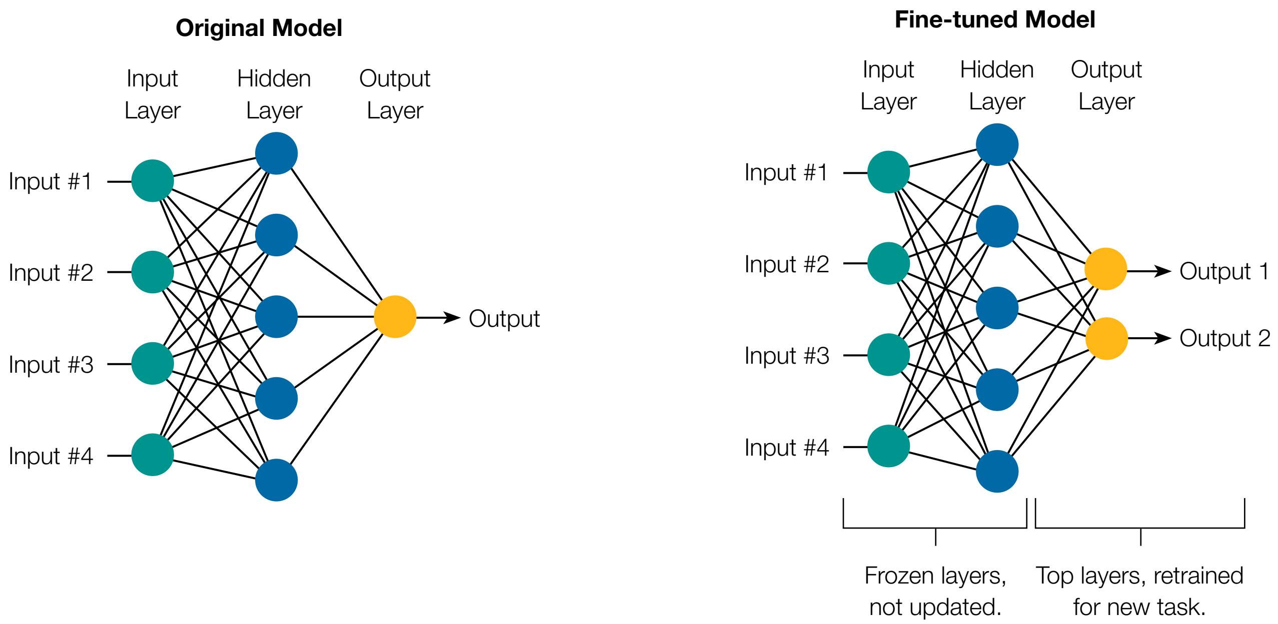
A CNN model has several interconnected layers. These layers perform different mathematical transformations on incoming data (e.g., images). Deciding which operation the layers should perform and how the layers are connected is an active area of research. These decisions are called the model architecture.

Generally, it is a best practice to use an established, published model architecture. Researchers continually share and publish codes for architectures that achieve state-of-the-art results. Using established architectures helps you get results faster, and it is unlikely that a novel architecture will produce significantly better results unless the person building it is a deep-learning researcher with a strong track record.

Most of the popular architectures used for satellite imagery are based on advances made in the field of computer vision. The current research standard for computer vision is the ImageNet competition, and architectures that have recently performed well on ImageNet are widely available.[6] Models that have performed well include GoogLeNet, VGG16, VGG19, and InceptionV3. These models are available for most libraries with pretrained weights—that is, they have been primed by being trained on the ImageNet data set.[7] The models can be fine-tuned to a specific use case. Using pretrained models has several benefits. They are much faster to train because they only need to be fine-tuned and not built from scratch. Since training models can take hours or days, quicker training time allows for much faster iteration. See Figure 7.

What does it mean to fine-tune a CNN architecture? Generally, to fine-tune a CNN, you keep the existing architecture and the existing weights, which were generated based on training from millions of images. You keep the weights that are connected with interim features that you care about, such as edges and textures, and you update the part of the model that translates these features into useful patterns for addressing the specific use case, such as crop mix or yield. To do this, you remove the topmost layer of the CNN architecture, which yields the general outputs of ImageNet (e.g., cats, dogs, hats, etc.), and replace it

FIGURE 7. **Fine-tuning models (also known as transfer learning) updates an existing model for a new task**



6 Pretrained models are available for many different libraries, for example through the Caffe "Model Zoo" (https://github.com/BVLC/caffe/wiki/Model-Zoo) or through the keras applications module (https://keras.io/applications/).

7 This is in contrast to many other machine models where the entire model is trained from scratch for a new application without any initial weights.

with one that yields as outputs the target variable you are interested in. The values, or weights, in the first part of the model are frozen, leaving some portion of the model "unfrozen" and able to update given the new data you provide. Finally, the model if fine-tuned by executing the training procedure, updating the weights in the unfrozen layers, and training them to produce the output you are interested in.

There are a few downsides to using a pretrained model. First, the architecture was not specifically designed for the use case in question. There may be instances where changing the architecture will improve the performance of the model; however, when starting from scratch, it is much more cost effective to use established architectures. Second, depending on the library and implementation, the input format for the model may be limited. Since the models are pretrained on ImageNet, they may accept only images that are 299 x 299 pixels and contain three color channels. This means that some information may have to be discarded by resizing the satellite data and removing spectra outside of the three colors. Third, choosing a model can limit the programming language and deep learning library choices to ones in which that model is available.

4.1 IT Infrastructure for Machine Learning

Machine learning is a computationally expensive process, and deep learning involves computing a huge number of parameters over a massive amount of data. Any discussion of deep learning is thus incomplete without a discussion of the hardware used to train the models.

The first order of concern for infrastructure is the scale of the data. Satellite imagery—especially from commercial providers—can quickly take up terabytes of storage. The space to store and the power to compute over this scale of data are expensive to set up and manage. As a result, organizations that do not have dedicated computational servers should use cloud resources. Providers such as Amazon Web Services, Google Cloud Engine, and Microsoft Azure provide hardware, virtual machines, and storage capable of scaling to the needs of a satellite imagery project. If you are using a commercial imagery provider, that provider may recommend a specific cloud platform—there are cost savings when a user's computation and storage is co-located with that of the imagery provider. It is cheaper and faster to transfer data from the imagery provider to your resources.

Cloud providers also use virtual machines with access to Graphics Processing Units (GPUs). GPUs are highly efficient at doing massive matrix calculations, a core computation for deep learning. Using GPU machines to train deep-learning models will significantly reduce the time required for training.

While cloud computing providers have significantly reduced the cost and expertise required to store and process large amounts of data, accessing and storing the images is still a significant technical undertaking. It can take a week or more to create, provision, and set up the resources required to host the images for the model to assess. This time and effort should be included in the requirements for resources. It would be nearly impossible for data scientists do this work on an individual laptop.

SECTION 5
CASE STUDY

THIS SECTION APPLIES THE RECOMmendations in this guide to a sample case and covers a project in which deep-learning models were trained to predict crop yield in Kenya.

5.1 Overview

FarmDrive is a microcredit provider for smallholder farmers. In collaboration with DrivenData, Impact Lab, and CGAP, FarmDrive explored the use of satellite imagery to augment its data collection and analytics for evaluating loans to smallholder farmers in Kenya. The goal of the project was to determine whether information derived from satellite data could be used as predictors of risk factors such as crop yield and income.

5.2 Geospatial Data

The first step was to evaluate data sources that could be used to predict crop yields (and thus, farmer income). Since FarmDrive did not have enough direct data from its farmers to estimate either crop yield or income, additional data sources were needed. The goal was to find a target variable data set from which to build predictive models. Crop yield data are often aggregated for geographic regions and were most readily available at the regional level for Kenya. However, because there are only 47 Kenyan regions, there were too few regions from which to build complex models of suitable geographic resolution. Smoothed models for crop yield from academic research were used instead of these government estimates.

Data provider Harvest Choice had the most relevant temporal and geographic coverage for its estimates. The model for producing these estimates is called the Spatial Production Allocation Model, and it produces estimates at a resolution of approximately 10 km x 10 km (Harvest Choice 2015). Estimates for the three most common crops from a survey of FarmDrive's farmers were used: maize, beans, and potatoes.

5.3 Satellite Modeling

The next step was to acquire the satellite imagery that corresponded with the ground truth. This was sourced from Planet, a commercial provider of satellite imagery.

IMAGE DATA

The scale of satellite image data for an area the size of Kenya made this a true "big data" project that required a cloud-based pipeline for working with the satellite data. Server instances were co-located with the satellite imagery data using Amazon Web Services, and a pipeline in Python software was created to identify the imagery "tiles" that intersected the area of interest for a given time frame and to activate and download these images to local storage.[8]

Images were downloaded based on several dimensions:

- Timeframe—against predefined seasons.
- Area of interest—all of Kenya, wards in Kenya, and defined areas.
- Cloud cover—tiles were limited to those with less than 5 percent cloud cover.
- Asset type—testing both "visual" and "analytic" assets.

The tiles were assembled into images that aligned geographically with the training data set by merging the tiles into a single image. The colors of these merged images were adjusted so that differences in color balances did not affect the machine-learning algorithms. Figure 8 illustrates this concept. As a result, these images matched the areas of interest and linked

8 A tile is a single image from the satellite provider.

FIGURE 8. **Merged images often have different color balances. Using a histogram-matching algorithm, input (left) looks more consistent (right).**

directly to estimates of crop yields from the Harvest Choice data. Overall, almost 5 terabytes of satellite data passed through the cloud-based data processing pipeline.

MODELING

Several CNN models were trained to test combinations of different modeling decisions, in particular:

- Crops: maize, beans, and potatoes
- Seasons: summer and fall (the seasons for which there was sufficient imagery coverage)
- Asset type: both visual and analytic assets
- Model architectures: two state-of-the-art computer vision architectures with different levels of model complexity: InceptionV3 and VGG 16 (Szegedy et al. 2015; Simonyan and Zisserman 2015)

These models were trained on GPUs that are optimized for tasks such as image processing.

To evaluate the results, the “error rate” was studied as the models were trained. The training error measures how well the model performs on examples it has been trained on (in-sample) while the validation error measures how well the model performs on examples it has not been trained on (out of sample). In Figure 9, the model improves if the green line (validation error) continuously decreases with the blue line (training error). Since the objective is to use the model to predict new cases, for which the target value is not known, the reduction in validation error is what makes the model effective.

5.4 Results

More complex architectures (InceptionV3) and crops that have larger plants (maize) produced the most accurate models.

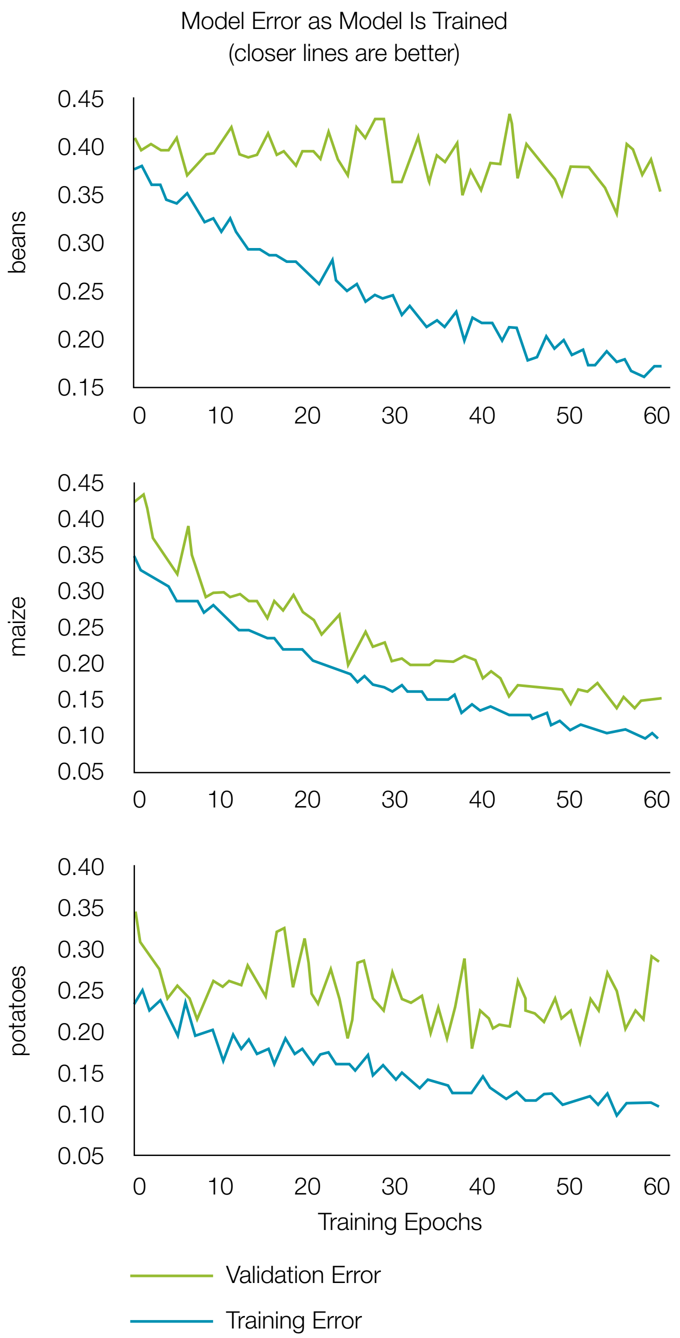
Overall, 24 different models were tested with varying degrees of success. The best performing model used the InceptionV3 architecture and red, green, and blue channels (visual assets) to predict maize yield using imagery gathered in summer 2016.

5.5 Limitations

This case study demonstrates both the power of the methods used and their limitations. The challenges that arose pointed to areas where investments are needed to make these methods viable at scale. These include the following:

- Aggregated estimates of yield at relatively large areas were used and are inappropriate for assessing the yield of individual farms that usually cover a much smaller area.
- The ground truth yield data was disjointed temporally from the satellite imagery. The crop yield estimates were over 10 years old (2005) compared with the imagery (2016).
- There were computational constraints. Even with cloud computing resources, the images were significantly scaled down from the native resolution provided by the imagery provider. This enabled a large number of models to be assessed on lower-cost machines. Using more expensive and sophisticated hardware and having a longer timeframe for the work would allow models to be trained at increased resolution to produce better predictions.

FIGURE 9. **Training passes for InceptionV3, summer, visual asset models.**



5.6 Lessons Learned

This use case demonstrated that CNN modeling could be used to quickly converge on increasingly accurate models for predicting crop yields for very different types of crops in a real-life environment. Histogram matching algorithms were successful in eliminating much of the variation in color balance from satellite imagery of the fields. Despite the limitations encountered, the satellite model shows promise and helped to indicate where FarmDrive should invest in more resources to build systems to better serve smallholder farmers. Ideally, the models would be built using objectively measured yield for target fields and a geographic shapefile that outlines the bounds of the field in question. These two pieces of data collected at sufficient scale (say 10,000 examples) would allow the models to be much more effective.

REFERENCES

Central Statistical Agency (Ethiopia) and the World Bank. 2013. "Crop Cutting Manual." Addis Ababa, Ethiopia: Central Statistical Agency and the World Bank, March. http://microdata.worldbank.org/index.php/catalog/2053/download/40405.

Graves, Alex, Abdel-rahman Mohamed, and Geoffrey Hinton. 2013. "Speech Recognition with Deep Recurrent Neural Networks." Ontario: University of Toronto. arXiv:1303.5778; https://arxiv.org/abs/1303.5778

HarvestChoice. 2015. "Maize Production (mt, 2005)." Washington, D.C.: International Food Policy Research Institute and University of Minnesota. http://harvestchoice.org/data/maiz_p.

Jozefowicz, Rafal, Oriol Vinyals, Mike Schuster, Noam Shazeer, and Yonghui Wu. 2016. "Exploring the Limits of Language Modeling." Version 2, 11 February. arXiv:1602.02410; https://arxiv.org/abs/1602.02410

Karpathy, Andrek. 2017. "CS231n Convolutional Neural Networks for Visual Recognition." Accessed 1 September. http://cs231n.github.io/classification/

Nogueira, Keiller, Otávio AB Penatti, and Jefersson A. dos Santos. 2017. "Towards Better Exploiting Convolutional Neural Networks for Remote Sensing Scene Classification." *Pattern Recognition* 61: 539–56. https://dl.acm.org/citation.cfm?id=3181486

Penatti, O. A. B., K. Nogueira, and J. A. dos Santos. 2015. "Do Deep Features Generalize from Everyday Objects to Remote Sensing and Aerial Scenes Domains?" In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Boston, Mass., 7–12 July. https://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=7293053

Python Software Foundation. "Python Language Reference." http://www.python.org.

R Development Core Team. 2008. "R: A Language and Environment for Statistical Computing." Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org.

Rogati, Monica. 2017. "The AI Hierarcy of Needs." *HackerNoon*, 1 August. https://hackernoon.com/the-ai-hierarchy-of-needs-18f111fcc007

Simonyan, Karen, and Andrew Zisserman. 2015. "Very Deep Convolutional Networks for Large-Scale Image Recognition." Version 6, 10 April. arXiv:1409.1556v6 [cs.CV]; https://arxiv.org/abs/1409.1556

Szegedy, Christian, Sergey Ioffe, Vincent Vanhoucke, and Alex Alemi. 2016. "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning." Version 2; 23 August. arXiv:1602.07261 [cs.CV]; https://arxiv.org/abs/1602.07261

Szegedy, Christian, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna. 2015. "Rethinking the Inception Architecture for Computer Vision." Version 3, 11 December. arXiv:1512.00567v3 [cs.CV]; https://arxiv.org/abs/1512.00567

APPENDIX. USING SATELLITE IMAGERY FOR FINANCIAL INCLUSION: A CHECKLIST

SATELLITE IMAGERY COMBINED WITH machine learning can be a powerful tool for improving financial and supply chain services to smallholder farmers. This guide shares a structure and the decisions that need to be made when using these tools.

As you begin your journey using satellite imagery, use this checklist to put together your plan. Not all items are required, but you should have a reason why they don't apply to your project. This checklist can help you put your organization on the path to success using state-of-the-art technology to help smallholder farmers.

Capacity

INTERNAL (OR CONTRACTED) CAPACITY

- ☐ Data scientist
- ☐ GIS specialist
- ☐ Data engineer
- ☐ Project coordinator
- ☐ Project champion

OUTSOURCED CAPACITY

- ☐ Firm with deep learning experience
- ☐ Firm with GIS experience

Shared Vision

- ☐ Agreed scope
- ☐ Clear definition of success

Data

TARGET VARIABLE

- ☐ What is your target variable?
- ☐ What is the source of these data?
- ☐ How often are data collected? When was the sample collected?
- ☐ What is the geographic range of the data?
- ☐ What is the level of aggregation of the data?
- ☐ What is the total number of examples you have?

SATELLITE IMAGERY

Choose a provider that best fits the following:

- ☐ Coverage area of the data
- ☐ Time period that matches your data
- ☐ Best resolution available
- ☐ Additional spectra, if applicable
- ☐ Suitable price for your budget

Model and Infrastructure

MODELING DECISIONS

- ☐ Select a programming language
- ☐ Select a deep learning library that has the models you will use
- ☐ Decide whether you will be using predefined architectures
- ☐ Decide whether you will be using pretrained models
- ☐ Decide how you will preprocess satellite imagery, including mosaicing, color balance, cropping, and cloud cover

INFRASTRUCTURE

- ☐ Cloud resources for storage
- ☐ Clouded resources for computing, specifically GPU machines
- ☐ Plan for managing cloud resources
- ☐ Budget for setting up and maintaining cloud resources

RESULTS EVALUATION

- ☐ Do you have enough data for a holdout validation set?
- ☐ What is your definition of success?
- ☐ Be willing to learn from what doesn't work and iterate

CGAP
cgap.org