

A PLAN for Tackling the Locust Crisis in East Africa: Harnessing Spatiotemporal Deep Models for Locust Movement Forecasting

Maryam Tabar

The Pennsylvania State University
mfg5544@psu.edu

Fei Jiang

The Pennsylvania State University
ffj5012@psu.edu

Dongwon Lee

The Pennsylvania State University
dongwon@psu.edu

Jared Gluck

The Pennsylvania State University
jpg5733@psu.edu

Derek Morr

The Pennsylvania State University
derekmorr@psu.edu

David P. Hughes

The Pennsylvania State University
dhughes@psu.edu

Anchit Goyal

The Pennsylvania State University
aqg5696@psu.edu

Annalyse Kehs

The Pennsylvania State University
annalysekeh@psu.edu

Amulya Yadav

The Pennsylvania State University
amulya@psu.edu

ABSTRACT

East Africa is experiencing the worst locust infestation in over 25 years, which has severely threatened the food security of millions of people across the region. The primary strategy adopted by human experts at the United Nations Food and Agricultural Organization (UN-FAO) to tackle locust outbreaks involves manually surveying at-risk geographical areas, followed by allocating and spraying pesticides in affected regions. In order to augment and assist human experts at the UN-FAO in this task, we utilize crowdsourced reports of locust observations collected by PlantVillage (the world's leading knowledge delivery system for East African farmers) and develop PLAN, a Machine Learning (ML) algorithm for forecasting future migration patterns of locusts at high spatial and temporal resolution across East Africa. PLAN's novel spatio-temporal deep learning architecture enables representing PlantVillage's crowdsourced locust observation data using novel image-based feature representations, and its design is informed by several unique insights about this problem domain. Experimental results show that PLAN achieves superior predictive performance against several baseline models - it achieves an AUC score of ~ 0.9 when used with a data augmentation method. PLAN represents a first step in using deep learning to assist and augment human expertise at PlantVillage (and UN-FAO) in locust prediction, and its real-world usability is currently being evaluated by domain experts (including a potential idea to use the heatmaps created by PLAN in a Kenyan TV show). The source code is available at <https://github.com/maryam-tabar/PLAN>.

CCS CONCEPTS

• Applied computing → Agriculture; • Computing methodologies → Supervised learning by classification.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

KDD '21, August 14–18, 2021, Virtual Event, Singapore.

© 2021 Association for Computing Machinery.

ACM ISBN 978-1-4503-8332-5/21/08...\$15.00

<https://doi.org/10.1145/XXXXXX.XXXXXX>

KEYWORDS

AI for Social Good, Spatio-temporal Deep Learning, Locust Crisis

ACM Reference Format:

Maryam Tabar, Jared Gluck, Anchit Goyal, Fei Jiang, Derek Morr, Annalyse Kehs, Dongwon Lee, David P. Hughes, and Amulya Yadav. 2021. A PLAN for Tackling the Locust Crisis in East Africa: Harnessing Spatiotemporal Deep Models for Locust Movement Forecasting. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '21)*, August 14–18, 2021, Virtual Event, Singapore. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/XXXXXX.XXXXXX>

1 INTRODUCTION

Since 2020, several parts of the world (especially East Africa and the Middle East) have been struggling with the worst Desert Locust (*Schistocerca gregaria*) swarm infestation in over 25 years [25]. The Desert Locust is a highly destructive pest during its swarming phase. Each 2g adult locust can move as much as 100 kilometres/day, consume its own weight in vegetation each day, and each swarm can contain billions of locusts [23, 24].

The current Desert Locust crisis has significant economic, human, and environmental impacts. For example, according to World Bank, the damages of the locust crisis could reach US\$8.5 billion in East Africa and Yemen. Apart from the economic costs, the decimated crops left by these locust swarms could jeopardize the food security of millions of people [23]. Thus, it is critically important to accurately predict the movement of these swarms, so that appropriate mitigation measures can be taken.

Currently, this is achieved by the Desert Locust Information Service (DLIS) at UN-FAO in Rome, Italy [24]. DLIS has historically relied on highly trained staff conducting field surveys in at-risk geographical areas, followed by governments allocating and spraying pesticides in affected regions. However, due to limited numbers of trained staff conducting field surveys, especially in countries (e.g., Kenya) where desert locusts are not normally present, the DLIS aims to augment its data collection and decision making through crowdsourced data. As a result, in 2020, PlantVillage¹, at the request of UN-FAO, developed eLocust3m², a smartphone application

¹<https://plantvillage.psu.edu/>

²<https://play.google.com/store/apps/details?id=plantvillage.locustsurvey>

which was designed for non-experts to use to crowdsource records of locust observations. The introduction of this eLocust3m application into a well established system of surveillance by the DLIS offers opportunities to enhance current locust mitigation operations.

Motivated by the widespread use of a crowdsourcing application for data collection on Desert Locusts (eLocust3m), we propose PLAN (**P**redictor of **L**ocust **A**ctivity and **m**oveme**N**t), a spatiotemporal deep neural network model that leverages real-world insights to accurately forecast locust presence/absence at high spatial and temporal resolutions across Kenya, Ethiopia, and Somalia (three countries in East Africa which have suffered great losses due to the Desert Locust crisis). *To the best of our knowledge, there is no existing comparable ML based tool for the prediction of locust presence/absence.* In particular, this paper makes four novel contributions: (i) through PlantVillage, a partner of the UN-FAO, we utilize data from eLocust3m, a first-of-its-kind tool which has been deployed in the field to crowdsource data about locust observations in Kenya, Ethiopia, and Somalia. (ii) Leveraging subject matter expertise and findings of prior studies in the agriculture domain, we identify ten environmental factors that contribute to locust breeding, migration, and survival, and fetch remote-sensed data for each of these ten factors. (iii) We propose a novel deep neural network architecture, called PLAN, that takes as input a single geographical location (in terms of latitude and longitude) and outputs accurate n -day forecasts of locust presence/absence at that location. PLAN explicitly models the spatio-temporal relationships inside locust movement data, and the impact of environmental factors on locust movement using a combination of Convolutional Neural Networks (CNNs) [19] and Long Short Term Memory (LSTM) models [12]. (iv) Finally, we comprehensively evaluate the effectiveness of PLAN for this problem domain. The experimental results show that PLAN outperforms several classical ML baseline models (in terms of predictive performance) on the n^{th} -step ($n \in \{1, 2, 3, 4\}$) forecasting tasks. For example, PLAN is the only model which achieves an AUC score of ~ 0.9 for next-day locust movement forecasts. More importantly, PLAN shows a significant improvement (23% higher F1 score) over the best performing baseline in a cross-region test (i.e., when we test PLAN’s performance on geographical regions which are far away from the regions where training data was collected), which illustrates PLAN’s capability of learning useful locust migration patterns.

PLAN is meant to be an assistive tool, which can augment the human expertise of highly trained staff at DLIS and PlantVillage in their locust prediction and mitigation efforts. Currently, PLAN’s real-world usability is being evaluated by domain experts, and a potential idea is to use heatmaps (of forecast predictions) generated by PLAN in a popular Kenyan TV show (which has an average viewership of five million farmers every week).

2 RELATED WORK

Historically, locust swarm migration has been studied from multiple perspectives in prior work. One line of prior research focuses on exploring the role of climatic factors in the outbreak of migratory locust swarms [6, 28]. For example, various studies have reported that different meteorological factors can have different levels of impact on locust breeding, maturation, migration, and survival;

e.g., (i) high precipitation can make a region suitable for locust breeding [28], similarly, (ii) soil moisture was also found to be a strong indicator of locust breeding areas [10, 11], (iii) wind can facilitate locust migration, (iv) green vegetation plays a key role in locust survival [28], and finally, (v) increased temperature resulting from climate change tends to exacerbate the problem of locust swarm infestation [28].

In addition, few data-driven studies at the intersection of agriculture and AI have addressed the locust crisis. Ye et al. [31] employed CNN-based models to detect locust species from imagery data. Kimathi et al. [16] used the Maximum Entropy (MaxEnt) model to identify potential locust breeding spots from several environmental factors. Moreover, in January 2021, the Selina Wamucii company³ has announced the ongoing development of an AI tool (called *Kuzi*) for predicting locust occurrence and breeding, however, the details of their underlying model and its predictive performance have not been released yet [22]. Therefore, to the best of our knowledge, there is no prior publicly available research on forecasting locust movement patterns at high spatial and temporal resolutions. To address this research gap, we propose PLAN, a ML algorithm that leverages recent advances in the field of spatio-temporal forecasting [29, 30], as well as findings of prior studies (in the agriculture domain) on locust outbreaks to generate accurate predictions of future movement patterns of locust swarms.

3 THE PROPOSED FRAMEWORK: PLAN

We now describe the data that is used to develop PLAN, followed by a discussion on PLAN’s architecture.

3.1 Dataset

Raw Data Sources. PLAN utilizes two sources of raw data: (i) crowdsourced locust survey data; and (ii) remote-sensed environmental data. Our crowdsourced locust survey data is collected through the “eLocust3m” (or eL3m) Android application, which has been developed by PlantVillage for the UN-FAO. This smartphone application enables users to record observations of locust presence/absence at a particular geographical location (given by latitude and longitude) on a given date (see Figure 1). Since 2020, eL3m has been deployed in many countries around the world, and the various groups (PlantVillage, county governments, charities) have employed local community members to scout the areas and provide geocoded observations of locust presence/absence via eL3m. In this paper, we only focus on eL3m locust presence/absence data collected from Kenya, Ethiopia, and Somalia (three countries which have been badly hit by the locust crisis) between March 1, 2020 to September 30, 2020. In total, during this time period, $\sim 21,000$ locust presence/absence reports were recorded in these three countries via eL3m. We use all these reports as our first raw data source. Note that data collection through eL3m has not stopped since September 30, 2020, and in future work, we will use these larger and improved datasets to design operational systems.

The second raw data source used by PLAN is related to the environmental factors that can affect locust breeding, migration, and survival. For this purpose, we take advantage of subject matter

³<https://www.selinawamucii.com>

expertise and prior work in the agriculture discipline [10, 28], and fetch raw remote-sensed data (between the dates of March 1, 2020 to September 30, 2020) for the following ten environmental factors from publicly available data sources cited below: (1) soil moisture [2, 17, 20], (2) sand content of soil [13], (3) precipitation [9], (4) land elevation [27], (5) wind speed at 10 meters [21], (6) wind speed at 50 meters [21], (7) U wind speed at 10 meters [15], (8) V wind speed at 10 meters [15], (9) total biomass productivity in 2019 (TBP_19) [26], and (10) actual evapotranspiration (AET) [26].

Rationale for the Choice of Environmental Factors. Each environmental factor chosen by us has been reported in prior work as potentially having an impact on locust breeding, migration, or survival. For example, high sand content in soil, and soil moisture is conducive for locust egg-laying [28]; as a result, precipitation, soil moisture, sand content of soil, and AET (which measures the amount of water evaporated from the ground and transpired by crops) can serve as strong indicators of potential locust breeding spots, which can help in forecasting their movement patterns. Further, wind is regarded as the main means of locust migration [28]. The wind heights most important to locust movement are 1,000 and 1,500m above sea level. Here, we use wind speed at 10/50 meters and directions (i.e., U/V wind) as they were readily available. In future work, we will integrate winds at higher elevations. Finally, certain land characteristics are conducive to locust presence; e.g., high locust activity is seen at lower elevations [18], and green vegetation is needed for locust survival [28]. As a result, land elevation, and TBP_19 can play important roles in forecasting locust presence/absence in different regions.

Data Characteristics. Our eL3m data has certain characteristics, which mainly stem from the nature of crowdsourced data collection. First, as locust presence/absence is voluntarily reported by human eL3m users, the total number of reports received in each day varies across time. Users often submit multiple records in close succession resulting in temporal and spatial aggregation. Figure 2 represents the total number of locust presence/absence reports received across Kenya, Ethiopia, and Somalia over time. As illustrated in this figure, a large number of locust presence/absence reports were received each day from the beginning of June until mid July. In particular, most of the reports received in June are locust presence (or, positive) reports, whereas majority of the reports received in July are locust absence (or, negative) reports. Second, the spatial distribution of the data is not uniform over time; e.g., on several

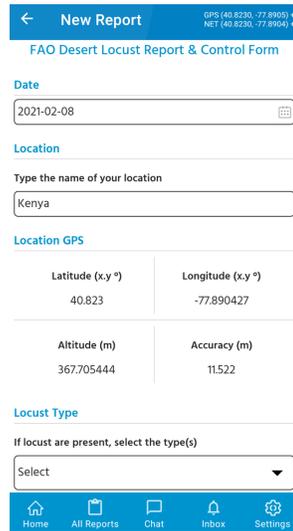


Figure 1: The eL3m user interface for crowdsourced data collection.

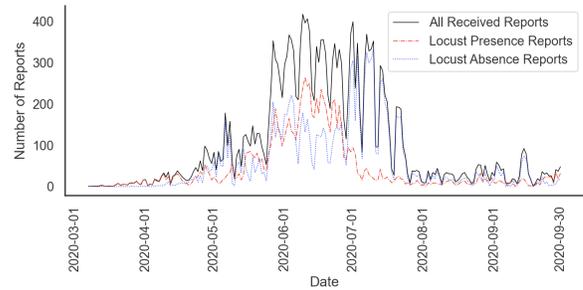


Figure 2: Distribution of eL3m locust presence/absence reports received from Ethiopia, Kenya, and Somalia over time.

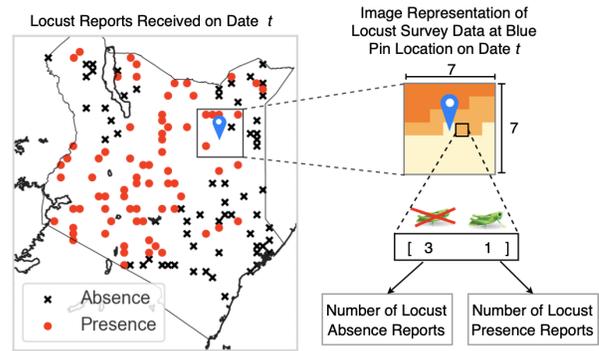


Figure 3: Schema for image representation of a single locust presence/absence report received on date t .

days in June, there are many regions in Kenya from which no locust reports were received. Third, we acknowledge the presence of some noise in the data, because people voluntarily report the locust presence/absence, and they might not report the ground-truth intentionally/unintentionally, e.g., there are false positive reports where users have considered it important to submit positive records even if locusts are not present (we elaborate on this in Section 6).

In addition, remote-sensed environmental factors are available at different temporal resolutions. For example, wind speed, soil moisture, and precipitation are available at a daily resolution, whereas AET is only available at a dekadal resolution⁴. On the other hand, sand content, TBP_19, and land elevation are static features which do not vary with time. Next, we describe data preparation steps.

Data Preparation. In our dataset, each data point corresponds to a single eL3m locust report with a binary (present/absent) label. Each of these data points is recorded by an eL3m volunteer at a particular geographical location (lat, long) and date/time t . For example, Figure 3 illustrates all such data points recorded in Kenya on date t (similar maps can be drawn for different dates and countries).

In order to represent each individual data point in Figure 3 (without loss of generality, we denote an arbitrary point by the blue GPS pin), we create an image based feature representation which captures the spatio-temporal movement of locusts in nearby regions

⁴In a dekadal dataset, three data points is available for each month. Depending on the length of the month, the length of each dekad ranges between 8 to 11 days.

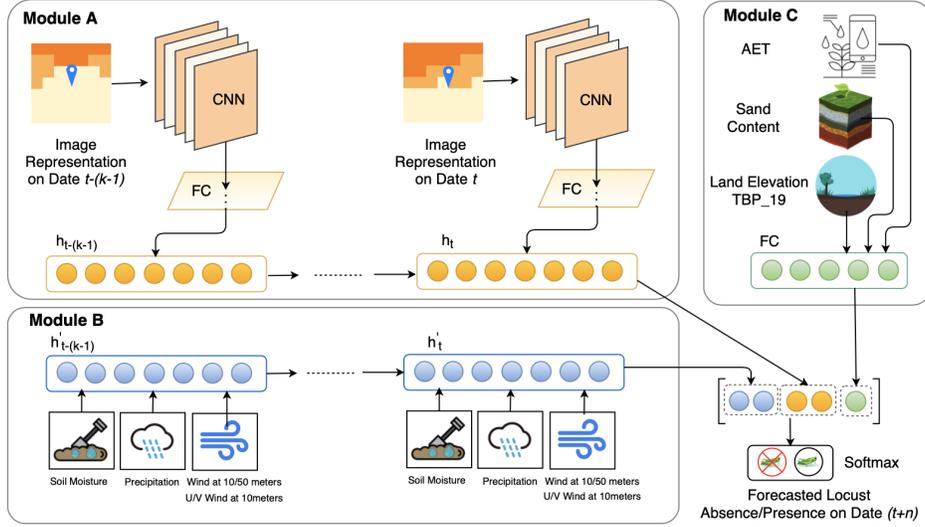


Figure 4: The architecture of PLAN.

(surrounding the blue pin location) over the previous k days. In particular, for each of the previous k days, we create a separate $7 \times 7 \times 2$ image representation which summarizes all eL3m locust reports (both presence and absence) received from surrounding areas which lie in a 7×7 grid centered on the blue pin location.

More formally, the feature representations for a data point corresponding to location $l = (\text{lat}, \text{long})$ and date $(t + n)$ are created by generating k images of size $7 \times 7 \times 2$, one for each date $t' \in \{t, t - 1, t - 2, \dots, t - (k - 1)\}$. In order to build the image for date t' , we grid the geographical area surrounding location l and create a 7×7 image, in which each pixel corresponds to a square geographical area of size $d^\circ \times d^\circ$ (in spatial resolution degrees). This image is centered on location l , hence the central pixel corresponds to a region of size $d^\circ \times d^\circ$ centered on location l , and other pixels correspond to nearby $d^\circ \times d^\circ$ regions. Finally, each pixel contains two pieces of information: (1) the total number of locust presence reports from that $d^\circ \times d^\circ$ region on date t' , and (2) the total number of locust absence reports from that $d^\circ \times d^\circ$ region on date t' . Using this procedure, we create k images of size $(7 \times 7 \times 2)$.

Intuitively, this time-varying image representation of data points enables us to explicitly capture the movement of locusts across space and time which can serve as important predictors for future locust movement, e.g., locust presence in a region increases the likelihood of locust presence in nearby regions in the near future and vice versa. Thus, this set of k images (one for each of the previous k days) forms the first part of feature representation for each data point in our dataset.

The second part of feature representation for each data point comprises of time-series values for six remote-sensed environmental variables (i.e., precipitation, soil moisture, U wind at 10 meter, V wind at 10 meters, wind speed at 10 meters, and wind speed at 50 meters) over the previous k days. Finally, the third part of feature representation for each data point comprises of single values for our static environmental variables (i.e., sand content, TBP_19, land

elevation, and AET of the last dekad). We normalize each of these features independently via Min-Max normalization.

Our final dataset consists of 21,012 data points, out of which 42.35% correspond to locust presence reports (i.e., positive class). Each data point consists of the following input features: (i) k matrices of size $(7 \times 7 \times 2)$, which correspond to the image representation of locust survey data in each day of the past k days, (ii) six time-series data of length k , each of which corresponds to the historical pattern of an environmental factor, and (iii) a vector of four elements which corresponds to values of our four static variables.

3.2 PLAN's Network Architecture

We now describe PLAN, a deep learning framework for generating accurate forecasts of locust presence/absence at different geographical locations in Kenya, Ethiopia and Somalia (our three countries of interest). PLAN takes the (latitude, longitude) of the target location and the current date t as input, and generates as output a binary forecast about whether locusts will be present (or not) at that (latitude, longitude) n days into the future (i.e., on day $t + n$).

Figure 4 illustrates the architecture of PLAN. At a high level, PLAN consists of three components: (i) a CNN+LSTM network for capturing spatio-temporal relationships from our image-based feature representations; (ii) an LSTM network for capturing temporal relationships in time-series environmental variables; and (iii) a Feed-Forward neural network (FNN) for extracting relevant features from the static environmental factors. In the following paragraphs, each component is explained in detail.

Module A: CNN + LSTM Model. We model spatio-temporal relationships in eL3m locust reports as follows. (i) For each data point, we build k image representations of locust report data (as described in Section 3.1) to summarize the locust reports received from surrounding regions over the last k days. (ii) Each image is passed through a separate CNN network followed by a fully-connected (FC) layer, which outputs dense latent representations of the spatial

relationships that exist in that image. (iii) The output from each FC layer is then fed as input to the hidden state of an LSTM network which captures locust migration patterns over space and time. Each of our CNN networks (in Figure 4) consists of three convolutional layers with 16 filters of size 3×3 and the FC layer has 64 neurons. Similarly, the hidden state size of the LSTM network is 256.

Module B: LSTM Model. We model the impact of environmental factors on locust movement as follows. (i) For each data point, we concatenate the time-series values of six environmental factors (i.e., soil moisture, precipitation, wind speed at 10/50 meters, and U/V wind at 10 meters) at that data point’s geographical location over the previous k days. (ii) This $6 \times k$ time-series data is passed through an LSTM network with h hidden states (with hidden state size = 64), which enables capturing time-varying patterns of environmental factors at a specific geographical location.

Module C: FNN Model. Prior studies in the agriculture domain show that locust presence could be associated with land characteristics. For example, sandy soil is favourable for locust breeding, and high locust activity is seen at lower elevations, etc. [18, 28]. As a result, PLAN takes four of such factors (i.e., land elevation, sand content of soil, TBP_19, and AET in the last dekad) as input and uses a FNN consisting of a FC layer to extract relevant features of the target region from these factors.

Finally, the output representations discovered by the last hidden layers of LSTMs in Modules A and B as well as the output of Module C are fed into a softmax layer to generate a predicted forecast of locust presence/absence n days into the future.

4 EXPERIMENTS

First, we discuss our baselines, evaluation approach, and experimental settings. Then, we evaluate PLAN’s performance as follows: (1) We compare PLAN with several baseline models to show its superior predictive performance on various forecasting tasks. (2) We conduct ablation analysis to show the impact of different parts of PLAN’s architecture on its predictive performance. (3) We conduct cross-region tests to evaluate PLAN’s performance when being tested on the data of a distant geographical region which is far away from the training region. (4) To tackle data sparsity, which stems from the unavailability of reports from many geographic regions, we propose a model-agnostic data augmentation algorithm, and then, assess its effectiveness in this problem domain.

Baselines. To evaluate the effectiveness of PLAN, we compare its performance with the following baselines: (i) Logistic Regression (Logit), (ii) Support-Vector Machine (SVM) with RBF kernel [4], (iii) AdaBoost [8], and (iv) Extreme Gradient Boosting (XGBoost) [3]. Building these baseline models requires one further pre-processing step as they cannot handle imagery data; i.e., we flatten the image representations of the eL3m locust report data, and concatenate them with all environmental factors to build the input feature representations for these baseline models.

Note that we choose these classical ML models as baselines, as there is no comparable prior work on sophisticated deep learning models to predict locust movement. Thus, any deep learning model that we compare PLAN against would have to be developed from scratch. Further, we note that during ablation analysis, we compared PLAN

against several variant neural network architectures to evaluate the contribution of different modules in PLAN’s superior performance.

Evaluation Approach. To evaluate the performance of various models, we take advantage of the walk-forward testing approach [1, 5, 14] and adapt it to our problem domain. At a high level, walk-forward testing extends the idea behind K-fold cross-validation to sequential time-series data. This method enables a more robust and trustworthy assessment of the performance of various forecasting models, because each model is evaluated under a series of time-varying conditions. In our domain, given the heterogenous distribution of eL3m locust reports over time (see Figure 2), walk-forward testing enables us to evaluate our models performance across a number of sequenced and time-shifted train/test splits.

However, in walk-forward testing, the overall predictive performance of a forecasting model (in terms of F1) is computed by averaging the F1 score achieved by the model across different time-shifted test sets (similar to macro-averaging in multi-class classification). Unfortunately, in our problem domain, the total number of eL3m locust reports per day changes over time (see Figure 2). Consequently, the different time-shifted test sets created during walk-forward testing have different number of data points. Therefore, it is not fair to report the average F1 score (or other evaluation metrics) computed on each time-shifted test set as the overall predictive performance of the model. To address this challenge, we combine all time-shifted test sets (and the predictions on those test sets) into a single larger test set. We compute all evaluation metrics on this single test set, and use these metrics to evaluate and compare different forecasting models. Prior literature has shown that this approach to computing the overall performance produces unbiased estimates of predictive performance in several situations, e.g., this approach is commonly used with k-fold cross validation, etc. [7].

Experimental Setup. All experiments are run on a machine with one NVIDIA Tesla T4 GPU, 4 vCPUs, and 15 GB RAM. Except for Table 1, the window length w of walk-forward testing is set to 7 days in all experiments. Finally, all experiments are run five times, and the average performance over all runs is reported.

The hyper-parameters are set as follows. All fully connected layers in PLAN’s architecture use Sigmoid activation functions. The batch size is set to 64, and Adam optimizer with a learning rate of 0.001, β_1 of 0.9, and β_2 of 0.999 is used. For all our experiments, we set the value of $k = 7$, i.e., both LSTM networks in Modules A and B of PLAN’s architecture (Figure 4) take the data of the past 7 days as input. Further, we set the value of $d = 0.2^\circ$, i.e., the spatial resolution of each pixel in our image representations (Figure 3) is set to 0.2° , which makes each pixel correspond to a $22.2 \text{ km} \times 22.2 \text{ km}$ geographical region. The value of d was set via hyperparameter tuning, as shown in the appendix.

Predictive Performance. Table 1 compares the predictive performance of PLAN against baseline models on 1^{st} -step prediction tasks (i.e., next day forecasts) with different choices of window lengths $w \in \{7, 14, 21\}$ (for walk-forward testing). The best model’s performance is shown in bold, whereas the second best model’s performance is shown with an asterisk. According to the results, PLAN consistently outperforms all baseline models; in particular, PLAN achieves an F1 score of ~ 0.792 with a window length $w = 7$,

Table 1: The predictive performance of different ML models on the 1st-step prediction task with various window lengths (w).

Model	$w = 7$ days			$w = 14$ days			$w = 21$ days		
	Accuracy	F1	AUC	Accuracy	F1	AUC	Accuracy	F1	AUC
Logit	0.7417	0.7026	0.7810	0.7366	0.6958	0.7652	0.7233	0.6823	0.7398
SVM	0.7772	0.7303	0.8433	0.7678	0.7104	0.8076	0.6853	0.5545	0.7870
AdaBoost	0.7585	0.7317	0.8282	0.7492	0.7247	0.8137	0.7408	0.7049	0.8002
XGBoost	0.7848*	0.7612*	0.8650*	0.7730*	0.7436*	0.8493*	0.7516*	0.7206*	0.8338*
PLAN	0.8174	0.7918	0.8904	0.8060	0.7750	0.8781	0.8052	0.7814	0.8798
Improv.	+4.15%	+4.01%	+2.93%	+4.26%	+4.22%	+5.21%	+7.13%	+8.43%	+5.51%

which improves upon XGBoost’s (the best performing baseline) performance by $\sim 4\%$. *This is a significant finding, as the high-stakes nature of decision-making in this domain means that any increases in predictive accuracy over baseline models could potentially lead to widespread impact (in terms of increased food security, better management of the locust crisis, etc.) at the scale of nations.*

In addition, PLAN tends to be more robust to increasing window lengths w as compared to baseline models. In particular, the percentage improvement achieved by PLAN over XGBoost (in terms of F1) significantly increases with increasing window length sizes. For example, PLAN improves upon XGBoost’s F1 score by 4.01%, 4.22%, and 8.43% with window length sizes of $w = 7, 14$ and 21 days, respectively. This finding illustrates that with increases in the window length size w , the distribution of training and test sets are more likely to differ from each other; as a result, the performance of all models is likely to degrade. However, Table 1 shows that PLAN is significantly less sensitive to potential covariate shift problems as compared to baseline models, e.g., when the window length is increased from $w = 7$ to $w = 21$, PLAN’s F1 score minimally degrades by $\sim 1\%$, whereas XGBoost’s F1 score degrades by $\sim 4\%$.

Next, we evaluate the predictive performance of different models on the n^{th} -step prediction task, as forecasting farther ahead into the future tends to be a more difficult task. Table 2 compares the predictive performance of different models on 2nd-step, 3rd-step, and 4th-step forecasting tasks. As expected, the performance of all ML models degrade with increasing forecast horizons. However, PLAN consistently outperforms baseline models at all forecast horizons. In particular, PLAN achieves an average AUC of 0.84 (across all horizon values) which shows its high capability of distinguishing positive/negative samples even when forecasting farther ahead into the future. *In summary, Table 1 establishes PLAN’s superior performance against strong classical ML baseline models on a real-world task for which no comparable prior deep learning models exist.*

Ablation Study. Having established PLAN’s superior performance, we now conduct two sets of ablation studies to investigate the impact of different parts of PLAN’s architecture on its overall performance. Our first ablation study evaluates the impact of different input features on PLAN’s performance. We build the following variants of PLAN: (i) PLAN\Env: All ten input environmental variables (both time-series and static ones) along with Modules B and C are removed from PLAN’s architecture. (ii) PLAN\eL3m: All eL3m locust report data along with Module A is removed from PLAN’s architecture. (iii) PLAN\Labs: Instead of using dual-channel image representations of eL3m locust reports (where we store both the

numbers of locust presence and absences reported at each pixel), we experiment with single-channel image representations by only storing locust presence numbers at each pixel in the image; as a result, our input images become $7 \times 7 \times 1$ sized.

Our second ablation study investigates the impact of different components of PLAN’s architecture on its predictive performance (i.e., all input features are used for the prediction task, but the architecture is changed). We build the following variants of PLAN: (i) PLAN\CNN: CNNs are removed from the architecture of PLAN; instead, all image data is flattened and is passed through the FC layers in Module A, the output of these FC layers is passed into the LSTM network in Module A. Modules B and C are unchanged in PLAN\CNN. (ii) PLAN\LSTM: Both LSTMs are removed from the architecture of PLAN; instead, the outputs of FC layers in Module A, the inputs of Module B, and the output of Module C are concatenated and fed into the output layer. (iii) PLAN\CNLS: All LSTMs and CNNs are removed from the architecture, and instead, a FC layer with the same number of neurons as the size of the LSTM hidden state is used to replace those networks. Thus, PLAN\CNLS is similar to a Multi-Layer Perceptron model.

Table 3 compares the predictive performance of our different ablations on the 1st-step forecasting task. The results show that PLAN\eL3m (which ignores eL3m data along with Module A) leads to the greatest decrease in PLAN’s predictive performance by reducing F1 scores by $\sim 17\%$. This illustrates the importance of the crowdsourced eL3m data in the superior predictive performance of PLAN. Further, PLAN\Labs, which removes locust absence information from the input images (of Figure 3) results in a 5.52% decrease in F1 score, which shows that locust presence reports (by themselves) are not enough to generate accurate forecasts, and incorporating locust absence reports in image based feature representations has a significant impact on the performance of PLAN. Additionally, PLAN\Env, which removes environmental factors, results in a 1.17% decrease in F1 score, which is consistent with domain insights on the role of environmental factors in locust activity and movement. Results from our second ablation study show that removing CNNs (i.e., PLAN\CNN) or LSTMs (i.e., PLAN\LSTM) from the architecture leads to $\sim 1\%$ reduction in F1 score. Additionally, removing both CNNs and LSTMs results in further decrease ($\sim 2.1\%$), in F1 score. This shows that different components of PLAN play roles of differing importance in its overall predictive performance.

Cross-Region Test. Until now, we trained PLAN on a portion of the data collected from Kenya, Ethiopia, and Somalia, and tested

Table 2: The predictive performance of different ML models for 2nd-step, 3rd-step, and 4th-step prediction tasks.

Model	2 nd -step prediction			3 rd -step prediction			4 th -step prediction		
	Accuracy	F1	AUC	Accuracy	F1	AUC	Accuracy	F1	AUC
Logit	0.7297	0.6933	0.7691	0.7210	0.6907	0.7692	0.7166	0.6855	0.7634
SVM	0.7569	0.7050	0.8244	0.7424	0.6887	0.8114	0.7407	0.6956	0.8017
AdaBoost	0.7438	0.7158	0.8028	0.7303	0.7014	0.7834	0.7302	0.7049	0.7992
XGBoost	0.7717	0.7522	0.8507	0.7578	0.7374	0.8346	0.7507	0.7329	0.8380
PLAN	0.7908	0.7588	0.8637	0.7726	0.7429	0.8497	0.7692	0.7340	0.8427

Table 3: The results of ablation study.

Model	Accuracy	F1	AUC
PLAN\Env	0.8057	0.7801	0.8822
PLAN\eL3m	0.6381	0.6162	0.7039
PLAN\LAbs	0.7398	0.7366	0.8512
PLAN\CNN	0.8050	0.7822	0.8816
PLAN\LSTM	0.8109	0.7820	0.8835
PLAN\CNLS	0.7960	0.7708	0.8686
PLAN	0.8174	0.7918	0.8904

it on another portion of the same data. Now, we evaluate the performance of PLAN when trained and tested on datasets from two geographically distant regions. We hypothesize that in this cross-region test, Module A, (i.e., the component designed for capturing spatio-temporal patterns of locust movement) should still be able to learn useful location-agnostic patterns of locust migration.

For this purpose, in addition to the data from Kenya, Ethiopia, and Somalia, we use eL3m data collected from Iran during the same time-period (i.e., March 1, 2020 to September 30, 2020) which consists of 5,117 locust reports. To check the aforementioned hypothesis, in each iteration of walk-forward validation, we use the same training portion of the data from Kenya, Ethiopia, and Somalia to train the PLAN model. Then, we replace the test data with the locust reports received from Iran in that particular test period, and evaluate the performance of the trained model on this new test set.

Table 4 shows the predictive performance achieved by PLAN and XGBoost in our cross-region test on the 1st-step prediction task. As expected, the predictive performance of both ML models degrades in this cross-region test. However, PLAN consistently outperforms XGBoost on each evaluation metric, e.g., PLAN achieves ~23% higher F1 score than XGBoost in this cross-region test. More importantly, comparing PLAN with PLAN\Env shows that removing environmental factors from PLAN results in a significant improvement in its predictive performance when being tested on the data of Iran. This improvement (that results from removing environmental factors) makes sense because the climatic conditions in Iran differ completely from conditions in Kenya, Ethiopia, and Somalia. Consequently, training our models on environmental variables from East Africa could add noise to the model’s forecasts when tested on Iran. Additionally, PLAN\Env achieves an AUC of ~0.8, which indicates its capability in learning useful locust movement patterns that can help it generate relatively accurate forecasts about locust presence in regions located far away from the training region.

Table 4: The results of cross-region test (i.e., the models are trained on the data of three East African countries and tested on the data of Iran).

Model	Accuracy	F1	AUC
XGBoost	0.5276	0.3322	0.6464
PLAN	0.6576	0.4115	0.7480
PLAN\Env	0.7363	0.4819	0.8062

Model-Agnostic Data Augmentation. As locust observations were voluntarily reported by human eL3m users, locust reports are not available for many geographical regions on any given day. Thus, there are many 0’s in the image representations of eL3m locust report data, as each image summarizes the total number of locust (presence/absence) reports received from a specific region on a particular day. To account for this data sparsity, we implement a model-agnostic linear interpolation approach for data augmentation, and evaluate its impact on model predictive performance.

Our linear interpolation based data augmentation approach relies on the following intuition about locust movement: if locusts are reported to be present (absent) in location l on two separate days (t_1 and t_2) that are close in time, it is highly likely that locusts are present (absent) at location l on all the days between t_1 and t_2 .

More formally, in our data augmentation procedure, to forecast locust presence/absence in location l at date t , we take the following steps after creating image representations of eL3m reports received by date $(t - 1)$: (1) if no locust report (neither locust presence nor locust absence) is available for a specific region, we set the value of the corresponding pixel to NULL in both image channels (locust presence and absence channels of the image). (2) For each region (i.e., pixel), we impute the time-series data of locust presence (absence) at each pixel separately using linear interpolation. (3) If no reports are available from specific regions, all elements of the time-series data could be NULL. Therefore, the remaining NULL values are replaced with 0 again. This procedure enables us to impute values for pixels which contain $[0, 0]$ (i.e., pixels which have no locust presence and absence reports at all). Furthermore, to forecast locust presence/absence in location l at date t , we do not rely on the reports received after date $(t - 1)$, and therefore, this data augmentation approach is consistent with the time-series nature of the problem.

Table 5 shows the impact of data augmentation on the predictive performance of ML models on the 1st-step forecasting task. Each evaluation metric’s value before/after data augmentation is

Table 5: Impact of data augmentation on the predictive performance of different ML models.

Model	Accuracy			F1			AUC		
	Before	After	Gain (%)	Before	After	Gain (%)	Before	After	Gain (%)
Logit	0.7417	0.7709	+3.93%	0.7026	0.7303	+3.94%	0.7810	0.8130	+4.09%
SVM	0.7772	0.7859	+1.11%	0.7303	0.7259	-0.60%	0.8433	0.8544	+1.31%
AdaBoost	0.7585	0.7978	+5.18%	0.7317	0.7608	+3.97%	0.8282	0.8651	+4.45%
XGBoost	0.7848	0.8099	+3.19%	0.7612	0.7749	+1.79%	0.8650	0.8819	+1.95%
PLAN	0.8174	0.8306	+1.61%	0.7918	0.8036	+1.49%	0.8904	0.9021	+1.31%
Avg			+3.00%			+2.11%			+2.62%

reported in the Before/After columns, respectively. The percentage of improvement achieved by applying data augmentation is reported in the Gain column. Table 5 shows that incorporating data augmentation results in a significant improvement in the predictive performance of all ML models; in particular, it improves the accuracy and F1 score by about 3.0% and 2.1%, respectively (on average), which shows this data augmentation technique’s effectiveness in this domain. Importantly, PLAN achieves an AUC of ~ 0.9 with this data augmentation technique. Thus, we propose to use PLAN with this data augmentation technique in future operational systems.

5 REAL-WORLD USAGE OF PLAN

One possible way in which PLAN can be used to assist farmers, policymakers, and human experts at UN-FAO is through the generation of high-resolution heatmaps (containing accurate forecasts of locust swarm movement). These heatmaps can give all three stakeholders an improved understanding of the future susceptibility of locust swarm infestation for different geographical regions, which in turn, can hopefully help them make a more well-informed locust mitigation plan. For example, these heatmaps can help decision makers in strategically allocating scarce resources (e.g., helicopters, pesticides, etc.) among high-risk geographical areas in order to ensure efficient resource usage and a corresponding reduction in locust swarm populations. In particular, our collaboration with PlantVillage enables us to potentially broadcast PLAN’s heatmaps daily on Shamba Shape Up⁵, a popular farming based television show which reaches five million farmers in Kenya every week.

Figure 5 illustrates a heatmap of 1st step forecasts (for June 10th, 2020) generated by PLAN. This heatmap is generated by running PLAN’s prediction model for each geographical location in Kenya on June 10th, 2020. This heatmap shows North West Kenya and East Kenya as two potential hotspots of locust presence on June 10th (characterized by high predicted likelihood of locust presence), whereas it shows Central Kenya as a potential source of locust absence reports (characterized by low predicted likelihood of locust presence). In this figure, the light blue circles and light green crosses shows the eL3m locust presence and absence reports (respectively) received on June 10th, 2020 across Kenya. These circles align well with our forecasted hotspot in North West Kenya, whereas the crosses align well with Central Kenya. Thus, this indicates that PLAN’s predictions have high recall in this example.

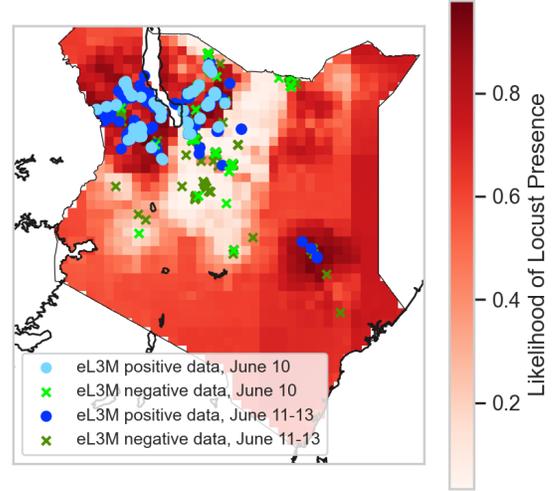


Figure 5: Plan’s forecasts about the likelihood of locust presence across Kenya in June 10, 2020 along with the ground truth reports received from Kenya at this particular date.

In order to understand why PLAN forecasted East Kenya as another hotspot, we plot dark blue circles and dark green crosses to represent eL3m locust presence and absence reports (respectively) received from June 11th to 13th, 2020 across Kenya. Interestingly, the dark blue circles align extremely well with the forecasted hotspot in East Kenya, whereas most of the dark green crosses align well with Central Kenya. We hypothesize that this is due to delays in data reporting by human volunteers, i.e., locusts arrived in East Kenya on June 10th, but they were reported by eL3m users on June 11th to 13th. Since we don’t have ground truth information, it is impossible to completely validate this hypothesis. However, we argue that the forecasted hotspot in East Kenya should not be viewed as false positives output by PLAN, as eL3m locust presence reports are recorded from the East Kenya hotspot within a period of 24 hours of our day of forecast. This illustrates that PLAN’s predictions also possibly have high precision.

6 IMPLEMENTATION CHALLENGES

The ubiquity of smartphones offers the possibility that well designed mobile apps (such as eL3m) can enable the collection of large amounts of data in a short period of time. For humanitarian

⁵<https://shambashapeup.com/>

challenges like locusts (but also including floods, droughts and other pests that damage crops), the potential benefits of such an application are very high. However, the major trade-off is data quality. Here we sought to use the data received from the crowd as-is in order to develop a machine learning model that could effectively use noisy data. We found that PLAN has led to an increased predictive performance over baseline models. While this is recognized, we understand that a major implementation challenge is the acceptance of such approaches by local actors such as governments in charge of the control operations. It would be expensive in both resources and time to deploy control operations to areas where locusts do not generally occur, but the model predicts their presence. As such, we think a major challenge will be to familiarise the decision makers with the opportunities and pitfalls associated with ML augmented desert locust predictions. We think one important role that PLAN could play is helping human experts more readily spot false records submitted by the crowd. This would reduce time spent in cleaning up databases which is currently a major task for staff at DLIS and PlantVillage. Thus, we hope that the use of PLAN would lead to a greater acceptance of ML to augment the human expertise at PlantVillage, UN-FAO, and other stakeholders.

7 CONCLUSION

This paper proposes PLAN, a ML algorithm for forecasting locust movement at high spatial and temporal resolutions. PLAN utilizes a unique crowdsourced dataset along with remote-sensed environmental data, and relies on a modular neural network architecture to provide accurate predictions of locust movement. Experimental results show that PLAN achieves superior predictive performance against several strong classical ML baseline models on a wide variety of locust movement forecasting tasks. PLAN represents a first step in using deep learning to assist and augment human expertise at PlantVillage (and UN-FAO) in locust prediction, and its real-world usability is currently being evaluated by domain experts.

ACKNOWLEDGMENTS

This work was in part supported by NSF award #1909702 and the Bill and Melinda Gates Foundation Grant #141840. We are also grateful for support from the UN-FAO.

REFERENCES

- [1] L. J. Cao and F. E.H. Tay. 2003. Support Vector Machine with Adaptive Parameters in Financial Time Series Forecasting. *Trans. Neur. Netw.* 14, 6 (Nov. 2003), 1506–1518.
- [2] J. Case, J. Mungai, Vincent Sakwa, B. Zavodsky, J. Srikishen, A. Limaye, and C. Blankenship. 2016. Transitioning Enhanced Land Surface Initialization and Model Verification Capabilities to the Kenya Meteorological Department (KMD). *American Meteorological Society Fall Meeting* (2016).
- [3] Tianqi Chen and Carlos Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.
- [4] Corinna Cortes and Vladimir Vapnik. 1995. Support-Vector Networks. *Machine Learning* 20, 3 (1995), 273–297.
- [5] D. Falessi, Jacky Huang, L. Narayana, J. Thai, and Burak Turhan. 2020. On the need of preserving order of data when validating within-project defect classifiers. *Empir. Softw. Eng.* 25 (2020), 4805–4830.
- [6] Food, Agriculture Organization of the United Nations, and World Meteorological Organization. 1965. *Meteorology and the Desert Locust: Proceedings of the WMO/FAO Seminar on Meteorology and the Desert Locust, Tehran, 25 Nov. - 11 Dec., 1963*. Secretariat of WMO.
- [7] George Forman and Martin Scholz. 2010. Apples-to-apples in cross-validation studies: pitfalls in classifier performance measurement. *Acm Sigkdd Explorations Newsletter* 12, 1 (2010), 49–57.
- [8] Yoav Freund and Robert E Schapire. 1997. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *J. Comput. System Sci.* 55, 1 (1997), 119–139.
- [9] Chris Funk, Pete Peterson, Martin Landsfeld, Diego Pedreros, James Verdin, J. Rowland, Bo Romero, Gregory Husak, Joel Michaelsen, and Andrew Verdin. 2014. A Quasi-Global Precipitation Time Series for Drought Monitoring Data Series 832. *USGS professional paper Data Series* (01 2014).
- [10] D. Gómez, Pablo Salvador, Julia Sanz, C. Casanova, D. Taratiel, and J. Casanova. 2018. Machine learning approach to locate desert locust breeding areas based on ESA CCI soil moisture. *Journal of Applied Remote Sensing* 12 (2018).
- [11] Diego Gómez, Pablo Salvador, Julia Sanz, and José Luis Casanova. 2020. Modelling desert locust presences using 32-year soil moisture data on a large-scale. *Ecological Indicators* 117 (2020), 106655.
- [12] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.
- [13] ISRIC - World Soil Information. 2020. SoilGrids250m 2.0 - Sand content. <https://data.isric.org/geonetwork/srv/eng/catalog.search#/metadata/713396fa-1687-11ea-a7c0-a0481ca9e724>.
- [14] Ieabeling Kaastra and Milton Boyd. 1996. Designing a neural network for forecasting financial and economic time series. *Neurocomputing* 10, 3 (1996), 215–236.
- [15] E. Kalnay, M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, A. Leetmaa, B. Reynolds, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. C. Mo, C. Ropelewski, J. Wang, Roy Jenne, and Dennis Joseph. 1996. The NCEP/NCAR 40-Year Reanalysis Project. *Bulletin of the American Meteorological Society* 77, 3 (March 1996), 437–472.
- [16] Emily Kimathi, Henri Tonnang, Sevgan Subramanian, Keith Cressman, Elfatih Abdel-Rahman, Mehari Tesfayohannes, Saliou Niassy, Baldwin Torto, Thomas Dubois, Chrysantus Tanga, Menale Berresaw, Sunday Ekesi, David Mwangi, and Segenet Kelemu. 2020. Prediction of breeding regions for the desert locust *Schistocerca gregaria* in East Africa. *Scientific Reports* 10 (07 2020).
- [17] S.V. Kumar, C.D. Peters-Lidard, Y. Tian, P.R. Houser, J. Geiger, S. Olden, L. Lighty, J.L. Eastman, B. Doty, P. Dirmeyer, J. Adams, K. Mitchell, E.F. Wood, and J. Sheffield. 2006. Land information system: An interoperable framework for high resolution land surface modeling. *Environmental Modelling & Software* 21, 10 (2006), 1402–1415.
- [18] Marion Le Gall, Rick Overson, and Arianne Cease. 2019. A Global Review on Locusts (Orthoptera: Acrididae) and Their Interactions With Livestock Grazing Practices. *Frontiers in Ecology and Evolution* 7 (2019), 263.
- [19] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. 1989. Backpropagation Applied to Handwritten Zip Code Recognition. 1, 4 (Dec. 1989), 541–551.
- [20] A. McNally, K. Arsenault, S. Kumar, S. Shukla, P. Peterson, Shugong Wang, Chris Funk, C. Peters-Lidard, and J. Verdin. 2017. A land data assimilation system for sub-Saharan Africa food and water security applications. *Scientific Data* 4 (2017).
- [21] NASA. 2021. POWER Data Access Viewer. <https://power.larc.nasa.gov/data-access-viewer/>.
- [22] Space in Africa. 2021. AI Tool Will Help African Farmers Fight Locusts Using A Warning And Prediction System. <https://africanews.space/ai-tool-will-help-african-farmers-fight-locusts-using-a-warning-and-prediction-system>.
- [23] The World Bank. 2020. THE WORLD BANK GROUP AND THE LOCUST CRISIS. <https://www.worldbank.org/en/topic/the-world-bank-group-and-the-desert-locust-outbreak>.
- [24] UN-FAO. 2015. FAO Desert Locust Information Service (DLIS). <http://www.fao.org/3/a-i4353e.pdf>.
- [25] UN-FAO. 2020. Massive, border-spanning campaign needed to combat locust upsurge in East Africa. <http://www.fao.org/news/story/en/item/1257973/icode/>.
- [26] UN-FAO. 2021. WaPOR: The FAO portal to monitor WAter Productivity through Open access of Remotely sensed derived data. <https://wapor.apps.fao.org/catalog/1>.
- [27] U.S. Geological Survey. 2013. GMTED2010 Viewer. https://topotools.cr.usgs.gov/gmted_viewer/viewer.htm.
- [28] WMO and UN-FAO. 2016. *Weather and Desert Locusts*. Technical Report WMO-No. 1175. World Meteorological Organization (WMO).
- [29] Huaxiu Yao, Xianfeng Tang, Hua Wei, Guanjie Zheng, and Zhenhui Li. 2019. Revisiting Spatial-Temporal Similarity: A Deep Learning Framework for Traffic Prediction. *Proceedings of the AAAI Conference on Artificial Intelligence* 33, 01 (Jul. 2019), 5668–5675.
- [30] Huaxiu Yao, Fei Wu, Jintao Ke, Xianfeng Tang, Yitian Jia, Siyu Lu, Pinghua Gong, Jieping Ye, and Zhenhui Li. 2018. Deep multi-view spatial-temporal network for taxi demand prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32.
- [31] Sijing Ye, Shuhan Lu, Xuesong Bai, and Jinfeng Gu. 2020. ResNet-Locust-BN Network-Based Automatic Identification of East Asian Migratory Locust Species and Instars from RGB Images. *Insects* 11, 8 (2020).

A APPENDIX ON REPRODUCIBILITY

In this section, we provide more details on our implementation and hyperparameter tuning. All codes are implemented in Python. We use keras (v. 2.0.8) with the tensorflow (v. 1.3.0) backend to implement neural network models. We also use scikit-learn (v. 0.23.2), numpy (v. 1.18.5), and pandas (v. 1.1.4) libraries. Table 6 summarizes the value of all hyperparameters.

We conduct hyperparameter tuning to set the value of d (the pixel width). To this end, we consider the data of the last w ($w = 7$) days of each training set as the validation set (similar to the nested cross-validation) and evaluate the performance of PLAN on the validation set. Figure 6 represents the accuracy of PLAN on the validation set with different pixel width ($d \in \{0.01, 0.02, \dots, 0.09, 0.1, 0.2, \dots, 1\}$). According to the results, $d = 0.2^\circ$ leads to the highest accuracy on the validation set, and thus, we set the pixel width to 0.2 in our experiments.

Table 6: The details of hyperparameters.

Hyperparameter	Value
Batch size	64
β_1 (Adam)	0.9
β_2 (Adam)	0.999
Learning rate	0.001
d (pixel width)	0.2
kernel size (CNN)	3×3
Number of filters (CNN)	16
Size of FC (Module _A)	64
k (length of time-series data)	7
size of hidden state (LSTM _A)	256
size of hidden state (LSTM _B)	64

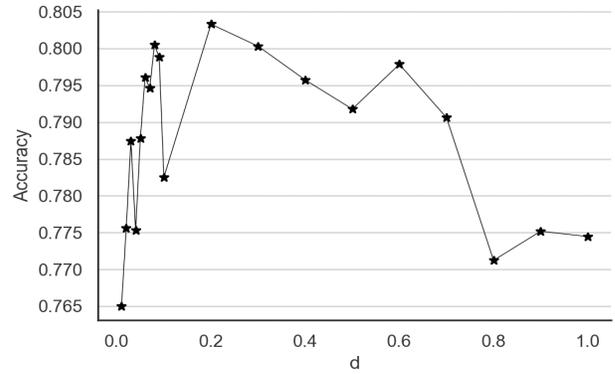


Figure 6: Performance of PLAN on the validation set with various pixel width ($d \in \{0.01, 0.02, \dots, 0.09, 0.1, 0.2, \dots, 1\}$).