Cattlelogue: Modeling Climate-Driven Livestock Migrations with Machine Learning

Jieruei Chang
Massachusetts Institute of Technology
jieruei@mit.edu

Julia Kurnik
World Wildlife Fund, Markets Institute
julia.kurnik@wwfus.org

***Abstract*—Changing global climates are likely to result in shifts of livestock production, potentially resulting in significant ecological impacts due to land conversion. Modeling of this shift is complicated by natural, economic, and social factors; prior efforts use complex feedback loops or account for only one climate variable. In this paper, we leverage machine learning models and large-scale climate projections to determine shifts in the distribution of cattle, validating accuracy on historical data. We show that this data-driven framework is a viable alternative to handcrafted expert systems in solving similar climate problems.**

***Index Terms*—livestock modeling, machine learning**

I. Introduction

As climate change accelerates, livestock around the world will be affected. Cattle cannot survive beyond a certain temperature regime. The crops that feed them may experience reduced yields. The grasslands in which they graze may disappear. It is known that climatic parameters such as surface temperature, relative humidity, solar radiation and rainfall patterns either directly or indirectly influence livestock survival, disease, parasitism, and feed resource availability [1], [2]; increased periods of high heat are already causing or are poised to cause declines in livestock yields in sub-Saharan Africa [3]. Climate change will therefore drive shifts in the distribution of cattle around the world. Determination of such shifts is relevant to securing the world's food supply as well as the livelihood of farmers. It is relevant to national and political actors who wish to determine how to best deploy economic resources such as infrastructure subsidies and microloans to incentivize ranchers. It is relevant to environmental groups, as ecosystems are cleared to make way for feedlots and pasture; conversion of natural landscapes into grazing ground is one of the leading contributors of food production to GHG emissions. Over half of GHG emissions from grassland cattle feed comes from pasture expansion in Latin America [4].

It is therefore imperative to estimate where cattle production is likely to relocate in response to changing climate conditions, so that appropriate actions may be taken. However, such systems are difficult to model due to the complex interplay of ecological, climatic, and socioeconomic variables, many of which exhibit nonlinear dependencies and spatial heterogeneity. The reaction of livestock distribution in the face of climate change is therefore poorly understood, and modeling it is considered by some to be beyond current capabilities [5]. In this work, we leverage machine learning models due to their ability to capture complex relations mapping climate and human variables to livestock suitability. Preliminary results show improved productivity in northern latitudes, migrations and ecosystem risks to the Pantanal and the Congo Basin, and losses in India and other equatorial regions.

Therefore, our contributions are as follows:

- We apply machine learning techniques to project climate-driven shifts in the spatial distribution of cattle. To the best of our knowledge, this is the first time any machine learning research has been done on this task.
- We provide initial regions of interest where livestock shifts may occur and particularly where such shifts may have ecological impacts, in order to facilitate focused deeper research, anticipatory land-use planning, or policy intervention.
- We provide a viable machine learning framework for future use in solving similar climate problems. While our results are promising, they should not be used to inform policy decisions or taken as absolute truth; rather, we hope they will be used to motivate further research into machine learning for modeling responses to climate change.

II. Related Work

A. Crop Yield Projection

A variety of statistical, biophysical, and machine learning approaches have been applied to determine impacts of climate change on agricultural outputs. In particular, the Agricultural Model Intercomparison and Improvement Project (AgMIP) has run a machine learning competition to project crop yields in the face of a high-emissions climate change scenario, with reasonably impressive results [6]. AgMIP itself focuses on process-based models that simulate yields by encoding

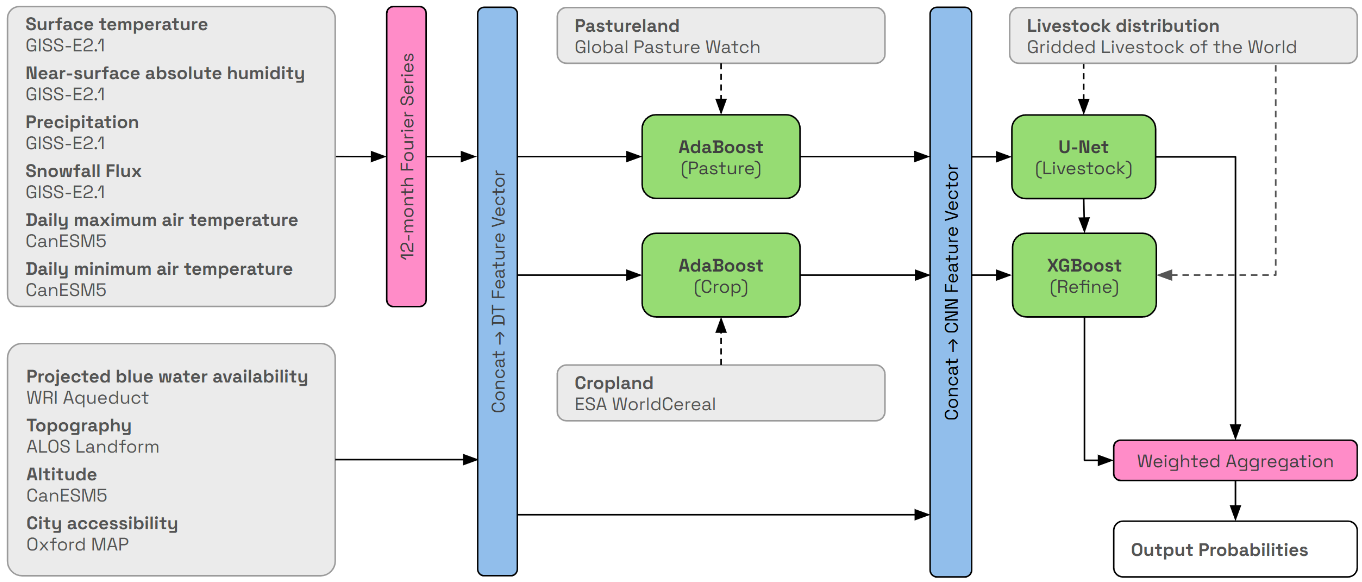

Fig. 1. Overall Model Architecture. Solid arrows represent forward propagation. Dashed arrows represent ground truth references used during training.

crop physiology, soil dynamics, and management practices. These kinds of biophysical models offer detailed insight but require extensive expert calibration. Additionally, a recent reduced-form empirical model uses longitudinal farm-level data to estimate the net effects of weather on crop yields, taking socioeconomic factors and climate adaptation into account [7]. This kind of model bypasses the need to simulate biophysical processes explicitly, instead capturing both observed climate impacts and the aggregate effects of real-world producer adaptations. One global-scale application estimates that adaptation offsets up to one-third of climate-related crop losses by 2100, but substantial residual declines persist. While these approaches are powerful for crop systems with extensive data, they have yet to be widely applied to livestock, where outputs are shaped by more complex interactions of climate, land, and management. This motivates our use of scalable machine learning models to identify emerging patterns in cattle distribution under climate change, without requiring detailed assumptions about underlying biophysical mechanisms.

B. *Livestock Yield Projection*

There exists prior work that tries to determine the effect of climate change on livestock production. Rahimi *et al.* [3] estimate decreases in cattle survival due to increasing heat stress in East Africa. To develop estimates of crop yields in Sub-Saharan Africa, Srivatastava *et al.* [8] simulate crop-vegetation-livestock interactions with detailed feedback loops involving variables such as the availability of feed and impacts on soil health.

However, the objective of our model is not to estimate changes in yields or optimal locations in which to raise livestock; instead, we attempt to model how livestock distribution may realistically shift due to climate change. This subtle difference in the problem statement means we account for both environmental suitability and human-driven constraints, such as historical land use patterns and existing agricultural infrastructure, and allows us to quantify the potential ecological impact of these shifts by identifying areas at risk of new pasture or feedlot expansion, particularly in biodiverse or previously undisturbed ecosystems, under future climate scenarios.

III. Methodology

We develop a multi-stage machine learning framework to project future shifts in cattle distribution under climate change. In the first stage, we use AdaBoost regressors to estimate future cropland and pasture suitability based on climate and infrastructure data. In the second stage, these projections, along with the original input data, are fed into a U-Net convolutional neural network that classifies livestock presence at high spatial resolution. A final third stage refines the U-Net's projections to create a higher-resolution map. The overall model architecture is shown in Fig. 1.

A. *Datasets and Inputs*

We use Gridded Livestock of the World (GLW4) data as livestock ground truth. ESA WorldCereal and Global Pasture Watch data are used as cropland and pasture ground truth respectively. Projections of surface temperature, precipitation, snowfall, and relative humidity are taken from GISS-E2.1 [9], which is part of the Coupled Model Intercomparison Project (CMIP6) ensemble. We supplement climatic variables with city distance data from Oxford MAP (as a measure of human infrastructure) and surface elevation from CanESM5.

B. *Data Preprocessing*

In order to better handle time-series variables such as surface temperature or precipitation over time, we apply a 12-month Fourier series decomposition to each time-series variable in order to better encode periodic patterns and smooth temporal variation.

To prepare data for U-Net inference, we tile the global input maps into overlapping square windows of size 16×16, corresponding to a visual field of approximately 300 miles per side at the equator. Overlapping tiles (50% overlap) are used to reduce edge artifacts during convolutional inference and to ensure smoother predictions across tile boundaries. Each tile includes the full set of processed input channels (both static and time-transformed dynamic variables) stacked into a multi-channel input tensor.

During training, tiles are randomly shuffled to improve generalization. Data augmentations (rotations and flips) are applied to encourage spatial/rotational invariance and mitigate overfitting. Labels for each tile are generated from the processed GLW4 livestock distribution maps, thresholded to binary presence/absence labels for classification.

C. *Model Architecture*

Since pasture and cropland are strong predictors of livestock presence, we aim to incorporate such data into our modeling pipeline. However, direct future projections for cropland and pastureland are either limited or unavailable altogether. We therefore introduce a multi-stage approach in order to first estimate future land-use suitability based on climate and infrastructure-based predictors, and then use these estimates as inputs to a downstream livestock classification model. We model both pastureland and cropland using decision trees, since their structure is well-suited for capturing rules that often underlie human land-use decisions. Because decision trees are prone to overfitting and high variance [10], we employ ensemble methods to improve robustness. Specifically, we use AdaBoost to combine multiple weak learners into an aggregate model that generalizes better to unseen data [10]. The cropland model is a regressor with 20 decision trees, each with a maximum depth of 4 nodes; the pasture model has 100 decision trees with a maximum depth of 4 nodes.

For the livestock projection stage, we choose a U-Net for its ability to leverage spatially distant data while still retaining high resolution [11]; the intuition is that the model needs to be able to access information about crops that are not in the immediate vicinity. The U-Net consists of four convolutional blocks with SiLU activation [12] and max pooling, followed by deconvolutions with skip connections. A detailed description of the U-Net submodel is shown in Fig. 2. Though the U-Net is designed to maintain spatial detail, it still tends to be unable to create finer details. We therefore follow it with a final XGBoost classifier [13] to create refined projections, and take a weighted sum of the XGBoost and U-Net results as the overall model output.

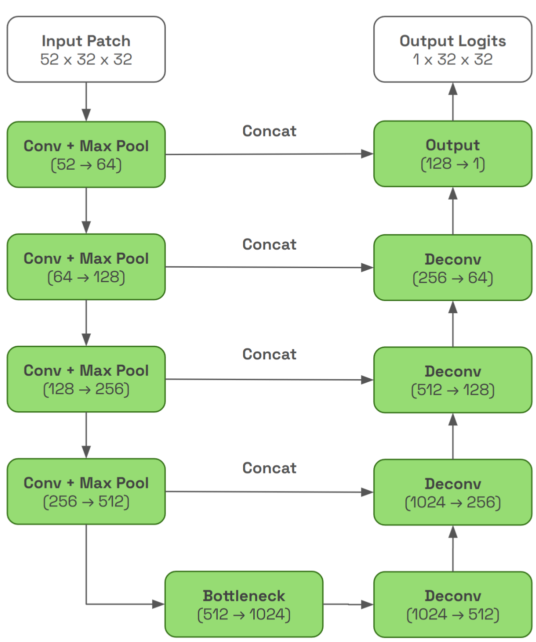

Fig. 2. Detailed Architecture of U-Net Model.

IV. Results

A. *Implementation*

AdaBoost models are implemented with scikit-learn and U-Net models are implemented with PyTorch. For the U-Net, we use AdamW as our optimizer at an initial learning rate of 10^{-4} and a StepLR scheduler with a step size of 10 epochs and $\gamma = 0.1$. To help mitigate overfitting, we apply dropout with $p = 0.3$ after each convolutional and deconvolutional block. We train for 50 epochs on a batch size of 256. All models were run using an nVIDIA GeForce RTX 4060 GPU.

B. *Model Validation*

Since samples overlap and there is high likelihood of cross-contamination, we run evaluation on historical data from the Annual Global Gridded Livestock of the World (AGLW) study [14]. Note that the AGLW data is generated from a machine learning model extrapolating

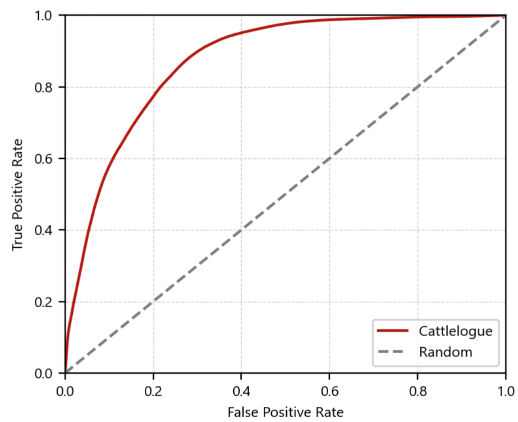

Fig. 3. ROC curve for livestock model.

from annual national averages and may not represent absolute ground truth. However, it has the earliest start date of any high-resolution global historical livestock dataset available; this allows us to reduce cross-contamination. We use the Area Under Receiver Operating Characteristic curve (AUC-ROC) metric to evaluate the livestock model against thresholded AGLW data from 1961. The ROC curve is shown in Fig. 3. The combined model achieves an AUC-ROC of 90.26%.

C. Ablation Study

In order to show the contribution of each component of our machine learning architecture to the overall model performance, we evaluate a series of other architectures: (1) a decision tree-based regressor that attempts to directly produce livestock projections, (2) a U-Net model that attempts to directly produce livestock projections, and (3) a two-stage model that uses a decision tree regressor in place of the U-Net. We evaluate these models against (4) our proposed architecture in Table I.

D. Analysis of Feature Contributions

We try to show the relative importance of each input variable in Fig. 4. Since the U-Net is essentially a black box model, it is difficult to analyze how the individual components are considered. Instead, Fig. 4 is generated by training a decision tree-based classifier instead of the U-Net, corresponding to the architecture in row (2) in the ablation study. Decision

TABLE I
Evaluation and Ablation Study (higher numbers are better)

Climatic/Infra	U-Net	Crop/Livestock	AUC-ROC (%)
✓			73.61
✓		✓	79.72
✓	✓		86.24
✓	✓	✓	90.26

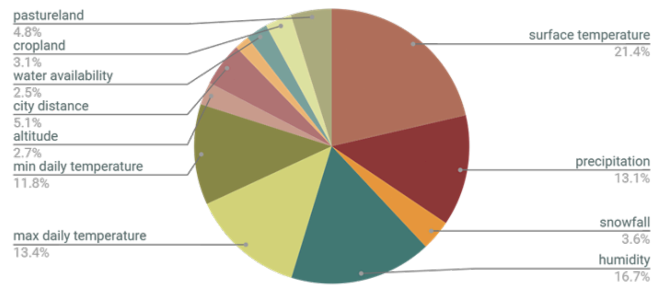

Fig. 4. Breakdown of feature importance.

trees offer interpretability by explicitly partitioning the input space according to feature values, allowing us to compute feature importances based on how often and how effectively each variable is used to split the data. While the ablation study shows that the surrogate model is not as accurate as the full U-Net, it allows us to approximate which variables the model relies on most heavily. The resulting chart should be interpreted as a proxy for feature importance in the U-Net. We see that it is dominated by climatic parameters as expected, but that the outputs from the intermediate crop and pasture projections still contribute to the overall model performance. In fact, it likely underestimates their importance, since the decision tree model lacks the spatial context and nonlinear compositional capabilities of the U-Net, which may be better suited to integrating these intermediate land-use variables in combination with climate signals. Thus, the contribution of crop and pasture projections may be more substantial within the full model than is captured here.

E. Future Projections

We generate future livestock distribution projections by applying our trained multi-stage model to climate inputs from the GISS-E2.1 model under the CMIP6 SSP4-6.0 scenario from the years 2015 to 2100. WRI Aqueduct data is used under the business-as-usual scenario, with linear interpolation used to create yearly data that covers the requisite timespan. All other inputs are held constant. Projections are made at the same spatial resolution as the historical data GLW4 data. We then calculate delta maps by subtracting averaged 2015-2024 model outputs from 2091-2100 projections. The map of livestock distribution shift is shown in Fig. 5, with intermediate drivers (cropland and pasture suitability) in Fig. 6 and Fig. 7.

V. Discussion

A. Analysis of Trends

Our projections reveal several notable patterns of cattle redistribution under climate change scenarios.

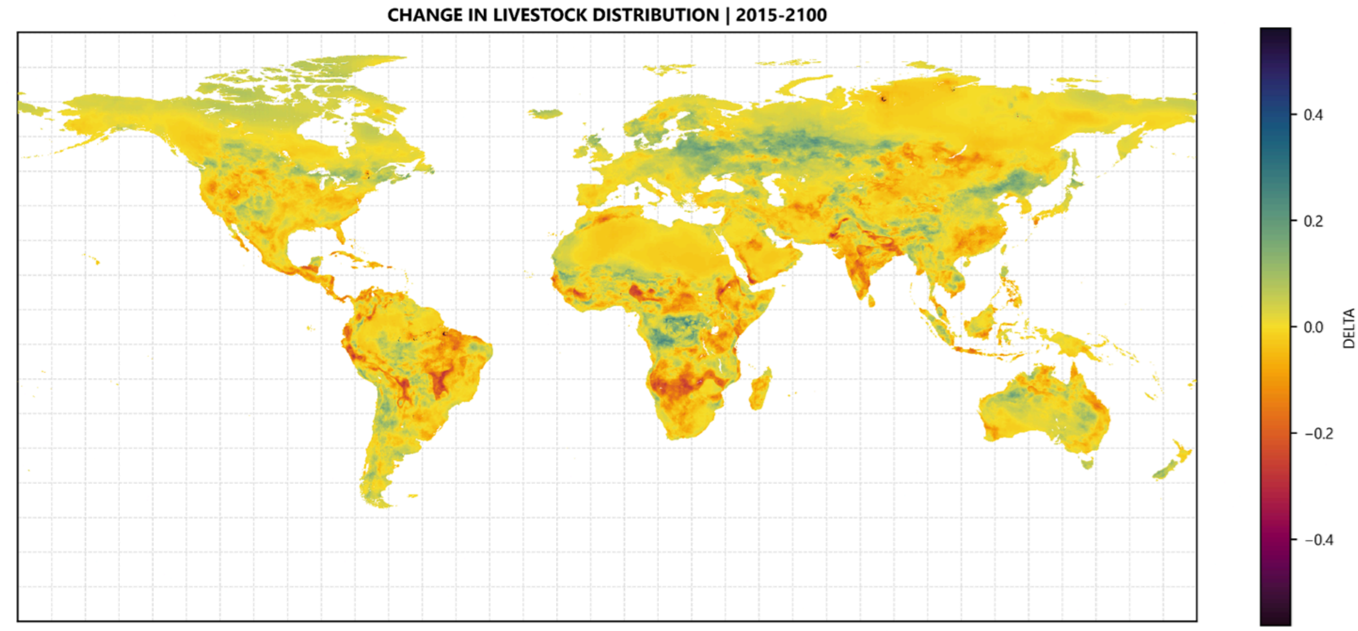

Fig. 5. Projected change in cattle distribution from 2015 to 2100. Deltas are changes in aggregated model confidence score.

The Pantanal wetlands in Brazil show a marked increase in projected cattle suitability, correlated with expanding pasture areas. This can be attributed to a drying trend in the region, reducing seasonal flooding and making land more amenable to grazing. Such transformations raise concern due to the Pantanal's exceptional biodiversity and ecological fragility. High-latitude regions, namely Canada, Northern Europe, and Northern Asia (particularly Siberia) are projected to experience increased cattle suitability. Rising temperatures are reducing the severity of winter conditions and thawing permafrost in these areas, thus increasing vegetative growth periods and potentially opening up new areas for pasture. Livestock migrations towards the Congo Basin area are evident in the model. This is a trend that is already being observed; Mbororo herders are moving to settle in the northern provinces of the DRC, causing regional conflicts and potentially leading to damage in the region's natural savannahs and forests [15]. India and other equatorial regions show significant declines. These are already hot regions expected to become more prone to heat stress, drought, and vegetation loss; increased climate variability and potential for disease outbreaks (e.g. ticks and parasites) may further reduce viability. In the United States, livestock distributions shift towards the Midwest and away from both the American Southeast and the West Coast. This trend coincides with changes in crop distribution observed in the intermediate stage, and aligns with projections from recent crop shift models [7].

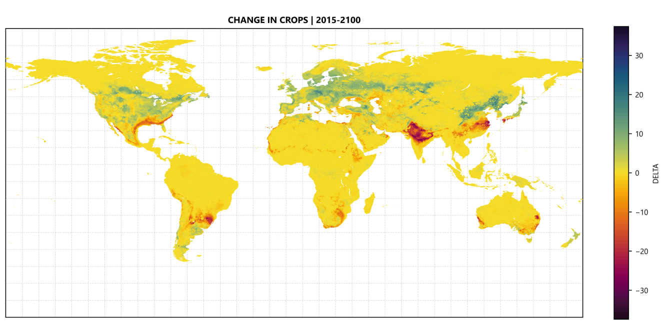

Fig. 6. Projected change in cereal crop distribution from 2015 to 2100. Deltas are percentage point changes in fraction of land projected to be covered by crops.

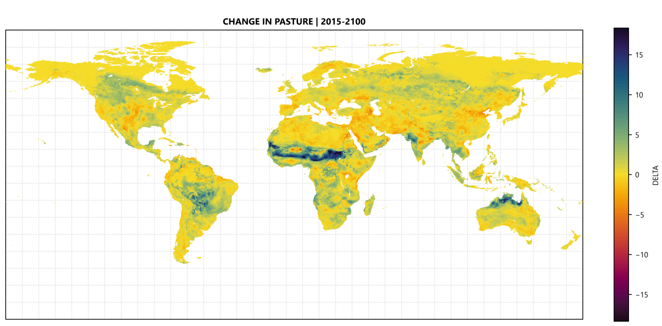

Fig. 7. Projected change in cultivated pastureland distribution from 2015 to 2100. Deltas are percentage point changes in fraction of land projected to be covered by pasture.

B. *Limitations*

Our model assumes that non-climatic inputs such as infrastructure and land topography remain static; in practice, these may also change under future socioeconomic scenarios. Second, while Fourier decomposition provides a compact representation of seasonality, it may miss abrupt or nonlinear temporal shifts, especially under extreme events or regime changes. Additionally, although we validate using historical datasets such as AGLW, these datasets themselves are not ground truth and include modeled estimates; the absence of high-resolution, ground-truthed temporal data limits the precision of our model's evaluation. We treat livestock relocation as a binary classification problem, which may oversimplify the nuances of livestock density, herd composition, and mixed-farming systems. Lastly, livestock distribution is often not fully determined by regions of best suitability. In the United States, the location of cattle is also dictated by market forces and government subsidies; although the current model includes an input that proxies as a metric for existing infrastructure, a more sophisticated socioeconomic model may be necessary to better capture these different factors.

VI. Conclusion

Our results illustrate that machine learning can be used to project livestock redistribution under future climate scenarios. The preliminary model projects increases in cattle suitability in high-latitude regions such as Canada, Northern Europe, and Siberia, as well as parts of southern Brazil and South Africa, while indicating declines across much of equatorial Africa, India, and the U.S. Southeast. Particularly concerning are projected expansions into ecologically sensitive areas like the Pantanal and the Congo Basin. By flagging regions at risk of new pasture and feedlot expansion, our framework can help guide anticipatory policy efforts and conservation planning.

A. Further Research

While this work focuses on cattle distribution due to their disproportionately large climate footprint [4], we believe that similar modeling approaches could be extended to other types of livestock, such as sheep, goats, pigs, or poultry. These animals have distinct physiological thresholds, feeding systems, and economic roles, and their redistribution patterns under climate change could differ substantially from cattle. However, we are confident that the data-driven nature of our approach makes it readily adaptable for solving these problems.

VII. Data and Code Availability

The software used in this study is available at https://github.com/knosmos/cattlelogue. Any additional data or code is available upon reasonable request.

Acknowledgements

We thank David Thau and Jordi Laguarta Soler for machine learning ideas, and Sam Wildman for insights into livestock production. We further acknowledge Shaun Martin, Taryn Skinner, Emily Moberg, Jessica Beck, and other colleagues at the World Wildlife Fund for support.

References

[1] V. Sejian, R. Bhatta, J. Gaughan, F. Dunshea, and N. Lacetera, "Review: Adaptation of animals to heat stress," *Animal*, vol. 12, pp. s431–s444, 2018, doi: https://doi.org/10.1017/S1751731118001945.

[2] E. Kimaro and O. Chibinga, "Potential impact of climate change on livestock production and health in East Africa: A review," *Livestock Research for Rural Development*, vol. 25, no. 7, p. 2013, 2013.

[3] J. Rahimi, J. Y. M. Mutua, A. M. O. Notenbaert, K. Marshall, and K. Butterbach-Bahl, "Heat stress will detrimentally impact future livestock production in East Africa," *Nature Food*, vol. 2, 2021, doi: https://doi.org/10.1038/s43016-021-00226-8.

[4] D. Kim, M. Burns, E. Moberg, and G. Thoma, "Measuring and Mitigating GHGs: Beef," *World Wildlife Fund*.

[5] C. Godde, D. Mason-D'Croz, D. Mayberry, P. Thornton, and M. Herrero, "Impacts of climate change on the livestock food supply chain; a review of the evidence," *Global Food Security*, vol. 28, p. 100488, 2021, doi: https://doi.org/10.1016/j.gfs.2020.100488.

[6] B. Groenke *et al.*, "The FutureCrop Challenge." 2024.

[7] A. Hultgren *et al.*, "Impacts of climate change on global agriculture accounting for adaptation.," *Nature*, vol. 642, pp. 644–652, 2025, doi: https://doi.org/10.1038/s41586-025-09085-w.

[8] A. K. Srivastava *et al.*, "Modelling mixed crop-livestock systems and climate impact assessment in sub-Saharan Africa," *Scientific Reports*, vol. 15, 2025, doi: https://doi.org/10.1038/s41598-024-81986-8.

[9] M. Kelley *et al.*, "GISS-E2.1: Configurations and Climatology," *Journal of Advances in Modeling Earth Systems*, vol. 12, no. 8, p. e2019MS002025, 2020, doi: https://doi.org/10.1029/2019MS002025.

[10] Y. Freund and R. E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, 1997, doi: https://doi.org/10.1006/jcss.1997.1504.

[11] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds., Cham: Springer International Publishing, 2015, pp. 234–241.

[12] S. Elfwing, E. Uchibe, and K. Doya, "Sigmoid-weighted linear units for neural network function approximation in reinforcement learning," *Neural Networks*, vol. 107, pp. 3–11, 2018, doi: https://doi.org/10.1016/j.neunet.2017.12.012.

[13] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, in KDD '16. San Francisco, California, USA: Association for Computing Machinery, 2016, pp. 785–794. doi: 10.1145/2939672.2939785.

[14] Z. Du *et al.*, "Annual global gridded livestock mapping from 1961 to 2021," *Earth System Science Data Discussions*, vol. 2025, pp. 1–20, 2025, doi: 10.5194/essd-2025-175.

[15] C. Assogba, "Herders caught between conflict and climate change in the northern Congo Basin." [Online]. Available: https://news.mongabay.com/2024/08/herders-caught-between-conflict-and-climate-change-in-the-northern-congo-basin/