A Deep Learning Approach For Deforestation Prediction

Varun Goyal* — *Moja Global* — gvarun@iitk.ac.in
Akshat Sharma* — *Moja Global* — akshatsh@iitk.ac.in
Priyanshu Yadav* — *Moja Global* — pryadav@iitk.ac.in
Ayush Gupta* — *Moja Global* — ayushgup@iitk.ac.in
Naishadh Parmar * — *Moja Global* — naishadh@iitk.ac.in
Tushar Goswamy* — *Moja Global* — tgoswamy@iitk.ac.in
Vatsalya Tandon* — *Moja Global* — vatsalt@iitk.ac.in
Mithil Mandape* — *Moja Global* — mithil@iitk.ac.in
Deval Pandya — *Moja Global* — deval.pandya@vectorinstitute.ai

***Abstract*—The concern for climate change is rising but do we have the right tools for predicting the future changes and taking the best preventive measures? Deforestation is one of the primary sources of greenhouse gas emissions from the land sector, supporting climate change. Efforts to reduce deforestation are currently hampered because of a lack of accurate deforestation predictions. This project develops an algorithm based on machine learning (ML) to predict the likelihood of a particular area in the landscape being deforested over the projection period. The predicted deforestation for each area can be aggregated to obtain expected deforestation for a project area, administrative region, or a whole country, which will be more informative for the policy-building agencies.**

I. Introduction

Agriculture, forestry, and other types of land use account for 23% [1] of global greenhouse gas emissions. Deforestation is one of the main driving factors of climate change in the land sector. Large amounts of carbon are stored in trees, absorbed from the atmosphere over decades, if not centuries. When the trees are cut down and used as firewood (or other short-term uses), the carbon stored in the trees is released into the atmosphere. Moreover, the dead trees stop removing CO_2 from the atmosphere leading to global warming and climate change. Investment to avoid deforestation is essential to reduce climate change and will become more important as we reduce other emissions. Such investment is currently hampered because accurate deforestation predictions are not available as the forecasts often rely on expert knowledge and statistical models. The project develops an algorithm based on machine learning (ML) to predict the likelihood of a particular pixel in the landscape being deforested over the projection period. The predicted deforestation for each pixel can be aggregated to obtain expected deforestation for a project area, administrative region, or a whole country, which will be more informative for the policy-building agencies. The problem is modeled as a time series forecasting problem, for which machine learning offers robust solutions like LSTMs to work with. The outputs of our project are compatible with emissions Monitoring, Reporting, and Validation (MRV) systems that conform with the United Nations Framework Convention on Climate Change (UNFCCC) and Intergovernmental Panel on Climate Change (IPCC) guidelines.

Uganda, an ideal country to sharpen the algorithm, was selected based on the need, data suitability, and local forestry support. According to the Worldbank Report [2], the rate of forest cover loss in Uganda stands at 2.6% annually, one of the highest in the world. As per the Uganda National Household Survey[3], more than 80% of Uganda's rural households use firewood for cooking. The high demand for wood fuel, increase in urbanization, and limited access to energy-saving alternatives means that forest cover is at risk of continued degradation. Previously Uganda did not had a unified approach for tracking deforestation/forest cover gain. Therefore, the ML models developed in this project will be a valuable tool for pertinent agencies such as the National Forestry Association (NFA) and National Environment Management Authority (NEMA) in carrying out their policy, tracking, and enforcement activities towards afforestation.

II. Past Work

Current techniques blend historical trends at the aggregated level with drivers of deforestation [4]. Some researchers have employed modeling techniques such as local forecasts that do not require past information at an aggregation level [5]. All of these measures are crude and do not establish a level playing field for investment in order to prevent deforestation. The methodology for deforestation projections varies from country to country (and often from one project to the next), adding to the uncertainty. Compared to these methods, AI can help improve deforestation prediction accuracy[6] while also reducing the time required to make projections by automating the entire process.

III. Progress and Approach

A. Data Collection And Preprocessing

We procured Landsat 7 images from Google Earth Engine in 30-meter resolution as input data from 2000 to 2020 for training the model. The data collected was pre-processed by filling in existing nan values. We tried various methods like

Project funded by Google AI for Social Good

* Equal contribution

forward fill, average fill, and flood fill for filling out nan values. Since the data was highly biased, we shortlisted data patches where deforestation has happened in considerable amounts during the last two decades for training purposes. We shard the data into 15 Hierarchical Data Format (HDF5)[7] files for faster data I/O. HDF5 data format stores and organizes large amounts of data in a hierarchical, filesystem-like data format. The Google Earth Engine data is composed of 11 spectral bands. While experimenting with dimensionality reduction techniques such as Principal component analysis (PCA), we found that 93.5% of the variance in the dataset can be explained using only three principal components. Hence we reduced the effective dimensionality of the data for improving training time. The target data was procured from the Mat Hansen dataset, also present on Google Earth Engine.

B. Approach

We applied AI to solve the core problem by implementing baseline models (RandomForest Regressor, Logistic Regressor, SVM Classifier, Bayesian Framework) and Deep Learning models (CNN-LSTM) to datasets for Uganda and the Philippines and have already made headway in using AI to solve the main problem. We have achieved a maximum AUC of **0.83** using CNN-LSTM models and have chosen it to be our final model. We proved that deep learning could help predict deforestation and thus help in preventive policy changes. We have applied deep learning to regions in Uganda and the Philippines and have achieved promising results for both regions.

C. Model architecture

The model is a standard CNN-LSTM architecture[8] (see figure 1). Each CNN-LSTM layer contains LSTM cells where the input, output, and forget gates are composed of convolutional feature extractors.

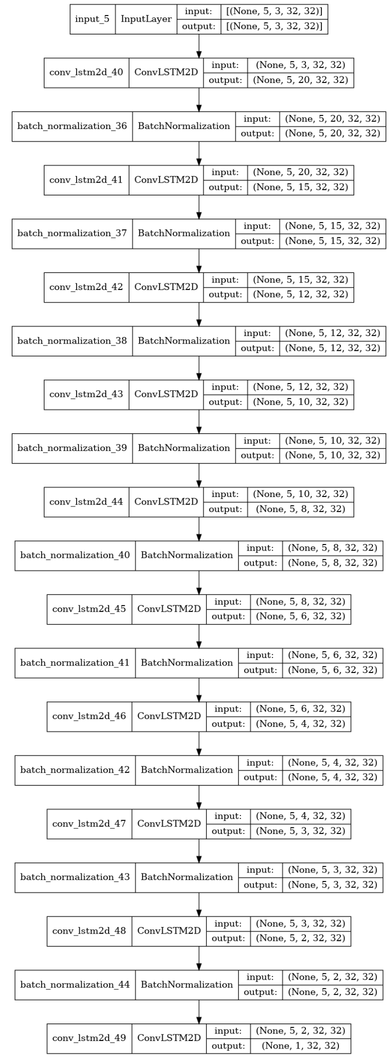

Fig. 1. Model architecture

D. Results

The following table I summarises the results obtained for various models we experimented with

TABLE I
RESULTS

Model	AUC Score
Baseline Ml techniques	0.68
ConvLSTM on binary targets	0.78
ConvLSTM on regression targets	0.83

IV. IMPACT

1) Accurate prediction of deforestation is necessary for any organization/nation that is working towards reducing the impacts of climate change by preserving the forests. The output of this project will not only inform decision making, but also monitoring, reporting and validation of climate mitigation activities at nation scales. The algorithm would help officials and NGOs alike to make more informed choices while allocating resources to maximise their efficiency. We have been focused on making our model scalable and reproducible such that it would help in spearheading deforestation prevention initiatives around the globe.
2) We have the required tools with user-friendly interfaces to make the results usable by people with a negligible technological background. Our partners are in contact with some NGOs and various national forest preservation authorities to whom we wish to explain our results. Having an insight into how their forests might change in the upcoming five years will surely help the policymakers take the best preventive measures and lead them closer to their goals for tackling climate change.
3) All the data acquired from various sources and preprocessed for this project would be made available to the scientific community. This will further help fast-track many projects in the climate change domain that might not have access to authentic and tested time series data.

V. Future Outlook

There are two ways in which we can scale our project. Firstly, we plan to extend the model and make the pipeline reusable for different regions. We plan to experiment with training data by incorporating data like proximity from roadways, waterways, and other spatial features. We hypothesize that this would improve the model performance as these factors lead to higher human intrusion into the forests, which leads to deforestation. We further plan to test the viability of transfer learning by implementing trained models in a new geographic region. Secondly, our approach for ground report collection leverages commodity mobile devices distributed to local communities. Hence it can also be applied in new regions at low cost and with additional social benefits such as improved weather, crop, and market information. The first point ensures that the project can be applied to regions where ground reports are not abundant. After implementing and testing our predictive deforestation models on Uganda, we will be in a position to understand whether general global models can be applied more broadly or whether the model would require calibration to local conditions using high quality and coverage ground reports.

Another important direction for future efforts would be creating models that empower policymakers to use AI to make informed decisions. Hence our model predictions would have to be explainable, reliable, and understandable to relatively less-trained individuals. We plan to use techniques such as computing saliency maps via guided backpropagation[9] to explain model outputs.

References

[1] *Land Use, Land-Use Change and Forestry (LULUCF)*. URL: https://unfccc.int/topics/land-use/workstreams/land-use--land-use-change-and-forestry-lulucf.

[2] *Ugandan Government Steps Up Efforts to Mitigate and Adapt to Climate Change*. URL: https://www.worldbank.org/en/news/feature/2019/05/31/ugandan-government-steps-up-efforts-to-mitigate-and-adapt-to-climate-change.

[3] *The Uganda National Household Survey 2016/17*. URL: https://sun-connect-ea.org/wp-content/uploads/2018/12/2017_UNHS_26092017-Final_Presentation.pdf.

[4] *Reference Levels: Concepts, Functions, and Application in REDD+ and Forest Carbon Standards*. URL: https://climatefocus.com/sites/default/files/reference_levels_concepts_functions_and_application_in_redd_and_forest_carbon_standards.pdf.

[5] Guifang Liu et al. *Costs and Carbon Sequestration Assessment for REDD+in Indonesia*. 2020. URL: https://www.mdpi.com/1999-4907/11/7/770/pdf.

[6] M. X. Ortega et al. "EVALUATION OF DEEP LEARNING TECHNIQUES FOR DEFORESTATION DETECTION IN THE AMAZON FOREST". In: *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* IV-2/W7.121-128 (2019). DOI: https://doi.org/10.5194/isprs-annals-iv-2-w7-121-2019.

[7] *Hierarchical Data Format*. URL: https://en.wikipedia.org/wiki/Hierarchical_Data_Format.

[8] Xingjian Shi et al. "Convolutional LSTM network: A machine learning approach for precipitation nowcasting". In: *Advances in neural information processing systems* (2015).

[9] Jost Tobias Springenberg et al. "Striving for simplicity: The all convolutional net". In: *arXiv preprint arXiv:1412.6806* (2014).