Rana, P., H. W. Fischer, E. A. Coleman, and F. Fleischman. 2024. Using machine learning to uncover synergies between forest restoration and livelihood support in the Himalayas. Ecology and Society 29(1):32. https://doi.org/10.5751/ES-14696-290132

Research

Using machine learning to uncover synergies between forest restoration and livelihood support in the Himalayas

Pushpendra Rana [1], *Harry W. Fischer* [2], *Eric A. Coleman* [3] and *Forrest Fleischman* [4]

ABSTRACT. In recent years, governments and international organizations have initiated numerous large-scale tree planting projects with the dual goals of restoring landscapes and supporting rural livelihoods. However, there remains a need for greater knowledge of drivers and conditions that enable positive social and environmental outcomes over the long term. In this study, we used interpretable machine learning (IML) to explore win–win and win–lose outcomes between livelihood benefits and forest cover using four decades of tree plantation data from northern India. Our results indicated that, in areas with a larger population of socioeconomically marginalized groups, moderate levels of education, and existing histories of community collective action, there is a higher probability of achieving joint positive outcomes. We also found that joint positive outcomes are more common within a consolidated local institutional space, suggesting that decentralized governance structures with cross-sectoral duties and functions may be better equipped to mediate conflicts between intersecting forest and land use challenges. Finally, our findings showed that non-forestry and anti-poverty interventions such as universal labor generation programs and universal education are associated with improved forest cover alongside livelihood benefits from plantations. Whereas contemporary policy discussions have given substantial attention to tree plantation schemes, our work suggests that effective restoration requires much more than planting alone. A broad mixture of socioeconomic, institutional, and policy interventions is needed to create favorable conditions for long-term success. In particular, anti-poverty programs may serve as important indirect policy pathways for ensuring restoration gains.

Key Words: *Himalayas; interpretable machine learning; lose–lose outcomes; social-ecological interactions; tree planting; win–win outcomes*

INTRODUCTION

Threats from climate change have galvanized many national governments and international organizations to invest in forest and landscape restoration to protect, enhance, and maintain forest cover. These investments are often promoted both to mitigate carbon emissions and support local livelihoods (Griscom et al. 2017, Bastin et al. 2019, Busch et al. 2019, Strassburg et al. 2020, Shyamsundar et al. 2022). Global land restoration efforts, such as the Bonn Challenge, Aichi Targets of the Convention on Biological Diversity, New York Declaration on Forests, and Intended Nationally Determined Contributions to the Paris Climate Agreement, rely heavily on forest restoration activities, including tree planting (Griscom et al. 2017, Bastin et al. 2019, Brancalion et al. 2019, Shyamsundar et al. 2022). To assist government restoration planning, many scholars have produced national or global-scale studies that estimate the potential carbon storage from natural climate solutions, including tree-based restoration (Fargione et al. 2018, Bastin et al. 2019, Brancalion et al. 2019, Roe et al. 2021).

Recent studies have attempted to identify areas where tree planting is more likely to produce livelihood benefits alongside environmental objectives (Brancalion et al. 2019, Brancalion and Holl 2020, Di Sacco et al. 2021, Rana and Varshney 2020). This work recognizes that extending forest cover without addressing local needs risks negative economic consequences for millions of forest-dependent people and compromises restoration efficacy (Erbaugh et al. 2020, Scheidel and Gingrich 2020, Pichler et al. 2021, Fleischman et al. 2022, Löfqvist et al. 2023). Many scholars have pointed out that these restoration assessments fail to adequately incorporate local governance, socioeconomic, and environmental conditions (Seddon et al. 2020, Pritchard 2021, Coleman et al. 2021a, Schultz et al. 2022). To date, there remains relatively limited empirical guidance on what variables affect win–win or win–lose outcomes, where to plant trees to maximize chances of win–win outcomes, and how to manage trade-offs between multiple resource management objectives. One promising way forward is to study past tree planting programs to identify the social-ecological factors and interactions associated with joint improvements in forest cover and sustainable livelihoods.

Past forest policy research has often relied on the qualitative identification of critical enabling conditions for sustainable resource management (Ostrom 1990, Agrawal 2001). This research often considers relatively few variables that shape win–win and win–lose relationships across multiple outcomes in forests (Chhatre and Agrawal 2009, Persha et al. 2011, Newton et al. 2016) or reports results on the basis of individual case studies rather than comparing across many cases (Agrawal 2001, Howe et al. 2014, Malkamäki et al. 2018, Miller et al. 2021). Recent literature has urged scholars to move beyond identifying individually influential variables and, instead, better understand how different suites of variables work together to influence outcomes on multiple social-ecological dimensions (Agrawal 2001, Agrawal and Chhatre 2011, Rana and Miller 2021). However, very little research has examined social-ecological processes or quantified such outcomes in forest restoration programs like tree planting (Adams et al. 2016, Malkamäki et al. 2018).

The latest machine learning research offers a methodological framework to examine such issues. Machine learning algorithms, especially predictive algorithms, rely on data-driven approaches to build models and then select the most appropriate model to predict outcomes based on cross-validation. There are three major

[1]Indian Forest Service, [2]Swedish University of Agricultural Sciences, Uppsala, [3]Department of Political Science, Florida State University, [4]University of Minnesota

strands that define how machine learning is advancing the frontiers of policy evaluation research. First, scholars have used a range of algorithms that estimate the causal impacts of natural resource and other policies and programs. These include double machine learning (Chernozhukov et al. 2018) and causal machine learning frameworks. Second, scholars have used supervised machine learning to estimate the heterogeneity in treatment effects of policies across subpopulations of studied units or regions (Athey and Imbens 2016, Rana and Miller 2019a). Finally, scholars have developed new tools, approaches, and frameworks to better explain and interpret black-box predictive algorithms to assist in decision-making and policy evaluation (Rana and Varshney 2023). This paper expands the last theme by empirically demonstrating the use of interpretable machine learning (IML), also known as explainable AI/ML, to understand the suite of variables that best explain tree restoration policy outcomes in northern India.

By extracting relevant knowledge from machine-learning models, IML techniques uncover relationships either hidden in the data or learned by the model (Murdoch et al. 2019). IML analysis produces predictive insights about domain relationships contained in the data, referred to as interpretations (Greenwell et al. 2024, Murdoch et al. 2019, Molnar 2022). IML techniques allow for identifying thresholds and variable effect reversals in a way that regression-based models cannot uncover with standard (linear or nonlinear) monotonicity assumptions (Elith et al. 2008). Specifically, IML methods help to identify key predictor variables and interactions, effect sizes, directionality of effects, zones of maximum influence, and critical thresholds associated with multiple outcomes including win–win, win–lose, or lose–lose outcomes (Molnar 2022). IML models can bring new insights hidden in the data (and not obvious to other traditional methods) in the field of social sciences (Epstein et al. 2021), ecology (Lucas et al. 2020), health (Wood et al. 2019), and human behavior (Du et al. 2020).

In this study, our objective was to understand the variables that drive joint social-ecological outcomes in tree restoration programs in northern India. We employed two methods to answer our question. First, we used boosted regression trees to identify a set of key social-ecological variables and interactions out of possible combinations of variables. These were used to predict improvement in forest cover and livelihood benefits in tree planting programs in northern India. Second, we demonstrated how IML can be a powerful toolkit for uncovering the relative importance of variables, developing hypotheses, and revealing the nature of the relationships between variables and the probability of achieving desired tree planting outcomes.

MATERIALS AND METHODS

Data were collected during 2018–2019 in the Kangra District in the northern Indian state of Himachal Pradesh. Forest cover has increased in the past decade in India, and Himachal Pradesh ranks in the top five states in terms of increase in forest cover from 2017–2019 (Forest Survey of India 2019). Data for the study came from 430 tree plantations established over the past four decades (the oldest dating back to 1980; Coleman et al. 2021). All plantations were ≥ 5 hectares, located on government-owned forest land, and were initiated as part of government programs. We worked with local key informants to map all plantations located in 60 randomly selected panchayats (villages with locally elected governments) and then conducted a socioeconomic survey of 40 randomly selected households in each panchayat. We removed observations with multiple plantations on the same site (n = 40) and those with missing values (n = 13), resulting in a study sample of 377 plantations.

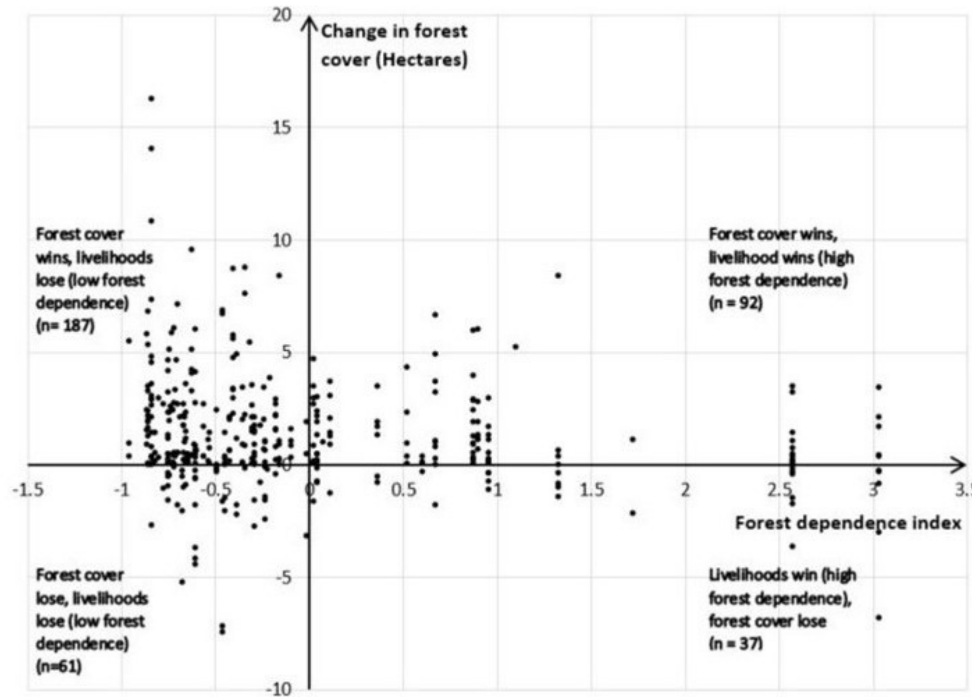
Outcome variables

For each plantation, we calculated change in tree canopy density from 2001–2017 based on national satellite data from the Forest Survey of India (2019) as an indicator of improvement in forest cover. We calculated livelihood benefits for a plantation through a forest dependence index, which was extracted through a factor analysis of the proportions of the quantity of (1) wild foods, (2) fodder, (3) grazing, and (4) timber used for domestic consumption that each plantation provided to the local forest users. We used a single-factor exploratory factor analysis to construct the forest dependence index using the minimum residual (minres) solution (see Appendix 1 for details including factor loadings, reliability, and consistency tests). We categorized plantations with positive (or zero or negative) values of forest dependence index as those with high (or low) livelihood benefits (see Appendix 1 for details). Strong positive correlation between improvement in tree cover and scores on the forest dependence index would suggest synergies whereas strong negative correlations would indicate trade-offs among the four plantation outcomes.

We used the forest dependence index as a proxy for estimating the livelihood benefits of plantations to the local communities. High values on the forest dependence index indicate high forest dependence of local communities on plantations. Higher dependence of rural people on forest plantations implies high usefulness of these plantations in terms of contributions to basic livelihood needs by providing resources that would otherwise need to be purchased on the market or would simply be unavailable to poorer households (Rana and Miller 2021). Hence, when there was a very small proportion of use (4%–14%) in our four resource use categories, the factor analysis gave a small negative value for our forest dependence index. We considered these plantations with small negative values (or zeros) as plantations that yield low subsistence use to people or where people have low forest dependence.

Although we categorized these forests as lose outcomes (or negative forest dependence index), care should be taken in interpretation because this variable does not allow us to distinguish whether the low level of forest use is caused by less useful forest resources or by alternative livelihood options in the context of the agricultural-livestock economy in our study area. Because of the lack of data, we can only say that some plantations get used more by some households. We do not know if this reflects better plantations with more livelihood benefits (i.e., with more useful products) or that these plantations are used by poorer/more forest dependent households located nearby. However, the dependence of the local communities in the study area is only on non-timber forest products such as dead and down limbs for fuelwood and standing trees and grass for fodder. People in these communities rarely harvest standing trees. Thus, local forest use by forest-dependent people is not likely to lead to a major decline in tree cover/tree density.

Fig. 1. Distribution of studied plantations (n = 377) along forest cover and livelihood outcome dimensions using a scatterplot of raw data. Change in forest cover inside plantation boundaries (in hectares) is calculated from Forest Survey of India (2019) satellite data from 2001–2017. Forest Survey of India (2019) defines forest cover as, "all lands having trees more than one hectare of area, with a tree canopy density of more than 10% irrespective of ownership, legal status of the land and species composition of trees." Livelihood contribution from each plantation is measured through a forest dependence index, calculated through factor analysis (for details, please refer to the method section and Table A1.4). Higher values of the forest dependence index indicate higher forest dependence of local communities on plantations.



We used a four-part classification to categorize the joint distribution of the outcome variables: (1) win–win outcomes, where there are high livelihood benefits and the forest cover improves; (2) livelihood win outcomes, where there are high livelihood benefits; (3) forest cover win outcomes, where there are forest cover gains or forest cover remains the same; and (4) lose–lose outcomes, where there are low livelihood benefits and the forest cover declines or remains the same. Fig. 1 shows how the joint distribution of forest cover change (y-axis) and forest dependence (x-axis) map onto these categories in the four quadrants of the graph. The final outcomes we modeled were dichotomous indicators of whether a plantation fell into each category (Tables A1.2–A1.3). For example, plantations where forest cover showed an increase and where there were high livelihood benefits were coded as win–win outcome plantations (a dichotomous indicator), and the livelihood win outcome was a dichotomous indicator of whether a plantation showed high livelihood benefits or not (Tables A1.2–A1.3).

Of the 377 forest plantations in our sample, 24.4% had win–win outcomes, 59.4% showed some combination of win–lose relationships between forest cover and livelihood outcomes, and 16.2% showed lose–lose outcomes. Within plantations with win–lose relationships, there were high livelihood benefits but decline (n = 33) or no change (n = 4) in forest cover in 37 plantations and forest cover improvement but low livelihood benefits in 187 plantations. Finally, plantations with lose–lose outcomes had a decline (n = 49) or no change (n = 12) in forest cover and low livelihood benefits. Looking at individual outcomes, we found an increase in the forest cover for 279 plantations (forest cover wins) and improvement of livelihood outcomes for 129 plantations (livelihood wins).

We note here that the forest dependence index is a cross-sectional snapshot of livelihoods (2018–2019), whereas our ecological outcome is calculated based on measuring change in tree cover from 2001–2017. Therefore, care should be taken when interpreting our results because we expect people to have variable levels of forest dependence on planted enclosures over time. In addition, there was a considerable time lag to obtain other livelihood benefits, such as fodder, wild fruits, or timber, and this time lag varied with the type of forests and species planted or naturally regenerated as well as with levels of plantation monitoring. Initially, plantations may yield higher levels of grass or even grazing for households because of protection of planted enclosures through fencing. Finally, in each planted enclosure there is pre-existing or post-plantation growth of naturally regenerated seedlings or already existing dense tree growth, which may also determine the flow of livelihood benefits to local communities depending upon the location, time, and levels of local forest dependence (Coleman et al. 2021, Rana and Miller 2021). Hence, a lack of data on livelihood changes over time is a limitation of our analysis.

Despite this, the forest dependence index is suitable for this analysis. Most of the plantations (83 of 129) that had high livelihood benefits were in areas where the dominant forest type is broadleaf or mixed. The plantations grown in these broadleaf and mixed forests with native species are likely to be associated with high livelihood outcomes because of their high utility to local populations for fodder (Coleman et al. 2021). Forest-dependent communities, especially Scheduled Caste and Scheduled Tribe groups, may be more likely to collectively act to ensure the success of these plantations because of their own high stakes in their success for fodder. As a result, we expected these plantations to show high forest dependence index values, which reasonably matched our forest dependence snapshot outcomes.

Predictor variables

Based on past research in our study region, we identified 36 variables with the potential to affect the outcome trajectories of tree planting programs (Table A1.1–A1.2). We chose our variables based on critical enabling conditions for sustainability on the forest commons (Agrawal 2001, Ostrom 2009, Miller and Hajjar 2020, Epstein et al. 2021), causal influences shaping forest conditions in the Indian Himalaya (Agrawal and Chhatre 2006), and conditions associated with long-term vegetation growth trajectories in the Kangra district of Himachal Pradesh (Rana and Miller 2019b). The variables represent a variety of socioeconomic, institutional, and environmental conditions that may predict win–win and win–lose relationships for livelihood benefits and forest cover outcomes. Despite our large set of variables, please note that this may not fully exhaust the full suite of theoretically relevant social-ecological variables (such as leadership or social capital) associated with these outcomes (Ostrom 2009).

We grouped the 36 variables into three sets: (1) socioeconomic and demographic characteristics of local communities, (2) institutional dynamics of forest governance and plantation activity, and (3) biophysical characteristics of plantations (Tables A1.1–A1.3). Socio-economic and demographic variables included level of education, poverty status, total number of households, total number of Scheduled Caste and Scheduled Tribe households, labor under the Mahatma Gandhi National Rural Employment Guarantee Act (MGNREGA), remoteness of the community, and other resource endowment-related variables. Institutional variables included indicators of community collective action, civic participation, community- and state-led plantation areas and species selection and supervision, plantation-making, land tenure, and plantation access and rule enforcement in panchayats. Biophysical characteristics included subsistence value of the plantations, plantation age, and plantation size.

We measured community collective action as the total number of days that people spent on activities involving mutual exchange of labor for forestry, agriculture, construction, and cultural activities as part of a traditional customary practice known as Juari (Vasan 2002). We measured civic participation of households across several civic groups active in each panchayat. Scholars have highlighted the importance of formal and informal civic groups (such as women and youth cultural groups, forest management committees, and forest cooperatives) in achieving favorable forest conservation outcomes (Andersson 2004, Gibson et al. 2005, Baynes et al. 2015, Chazdon et al. 2020). Scheduled Castes and Scheduled Tribe populations are recognized as more marginal socioeconomic communities in India and often are more forest dependent than other communities (Gundimeda and Shyamsundar 2012). In addition, literature on common property has highlighted how the nature of forest tenure rights (community, state, open-access), monitoring, and enforcement shape long-term social-ecological outcomes including forest restoration (Ostrom and Nagendra 2006, Coleman 2011, Coleman and Liebertz 2014). For details about the rest of the variables, please refer to Tables A1.1–A1.3.

Boosted regression trees and interpretable machine learning

We used a gradient boosting model to identify key variables that had a higher relative influence in terms of a substantial contribution in predicting joint tree cover and livelihood benefits (Elith et al. 2008, Ridgeway 2024, Greenwell et al. 2024; Fig. 2). We then used IML to estimate the magnitude and direction of variable effects and interactions between the variables. We used partial dependence plots (Greenwell 2017) and Friedman's H Statistic (Molnar et al. 2018) to explore these variable effects and interactions, respectively. We used standard machine learning evaluation methods to control for bias and overfitting in these models. These included tuning to balance fit with stratified tenfold cross-validated receiver operating characteristic curve (ROC) and predictive accuracy to avoid overfitting and to determine the optimal set of parameters (number of trees, shrinkage, interaction depth, and minimum number of observations in a terminal node) in each of our four plantation outcome models (Epstein et al. 2021).

Fig. 2. Methodological steps involved while examining the win-win, livelihood win, forest-cover win, and lose-lose plantation outcomes.

Step 1: Collection and cleaning of data for analysis; selection of appropriate theory-based key variable influencing the plantation outcome trajectories; constructing outcome variables.

Step 2: Tuning the model parameters for boosted regression tree models using ROC to obtain the optimal model.

Step 3: Using the optimized model parameters to estimate the relative influence of the variables on the probability of multiple plantation outcomes.

Step 4: Exploring the magnitude, direction and critical range of indicator values and thresholds where the effect of a particular variable is maximum on the probability of obtaining a particular plantation outcome using partial dependence plots.

Step 5: Estimating the interaction effects between two variables using Friedman's H statistic and, also estimates the zones of maximum interaction effects using bivariate interaction plots.

Step 6: Interpretating the results using local contextual knowledge to provide useful insights for policy and for developing hypotheses for further testing for causal interpretations.

We tuned the model parameters used in boosted regression trees using the caret package (Kuhn 2008) because such models are prone to overfitting (Epstein et al. 2021). The initial set of parameters to tune each model (four models—one for each binary

Table 1. Variables with high relative influence associated with multiple outcomes, their maximum predictive effect, magnitude of the predictive effect, and the variable's sample mean. Column 2 shows the relative influence of the variable in terms of its importance in changing the probability of multiple plantation outcomes. Column 3 shows the range of variable values wherein the variable's predicted effect on multiple plantation outcomes is highest. Column 4 shows the range of predicted effects corresponding to the values of a variable when its effect on outcomes is highest. Column 5 shows the mean value of the variable in the study sample.

Variables	Relative influence	Variable values with maximum predictive effect on the outcome[†]	Magnitude of predictive effect[†‡]	Mean value of the variable in the study sample
Win–win outcomes (livelihood wins, forest cover wins)[§]				
Scheduled Castes and Scheduled Tribes[|] households (number)	22.7	> 275	0.17–0.53	33
Level of education (%)	18.2	80–81	0.19–0.25	88
MGNREGA[¶] (employment scheme) labor days	12.7	> 1210	0.42–0.52	690
Community collective action (days)	8.9	358–470	0.08–0.09	418
Total households (number)	5.3	282–567	0.11	509
Land under cultivation	4.2	6.5–166.2	0.04–0.06	135
Number of civic groups	3.7	4–5	0.09	15
Livelihood wins				
Level of education (%)	29.8	80	0.81	88
Scheduled Castes and Scheduled Tribes[‡] households (number)	22.9	306–390	0.81	33
Total households (number)	14.8	242–567	0.17	509
Community collective action (days)	7.4	> 423	0.15–0.18	418
MGNREGA[¶] (employment scheme) labor days	6.7	> 1210	0.46–0.84	690
Acreage under cultivation (kanals)	6.5	6.5–36.4	0.40	135
Forest cover wins				
Plantation age (years)	10.9	13–38	0.79–0.83	19.5
Plantation size (ha)	8.8	5.6–8.7	0.81	8.9
Community collective action (days)	6.8	275–386	0.81–0.83	418
Scheduled Castes and Scheduled Tribes households (number)	6.1	3–107	0.81–0.82	33
Decrease in livestock (number)	6.0	> 20	0.78–0.79	34
Plantation equitable benefits (number)	5.2	0–10	0.75–0.79	2.8
Lose–lose outcomes (livelihood loses, forest cover loses)				
Plantation age (years)	8.9	1–21	0.12–0.14	19.5
Community collective action (days)	7.5	< 265	0.16–0.17	418
Increase in culturable waste	7.2	2–7	0.16	15.8
Community LPG[#] use	6.4	> 6.7	0.10–0.14	7.1
Total households	6.3	> 445	0.10–0.15	509
Acreage under private grasslands	6.1	6–82.1	0.11	96.8

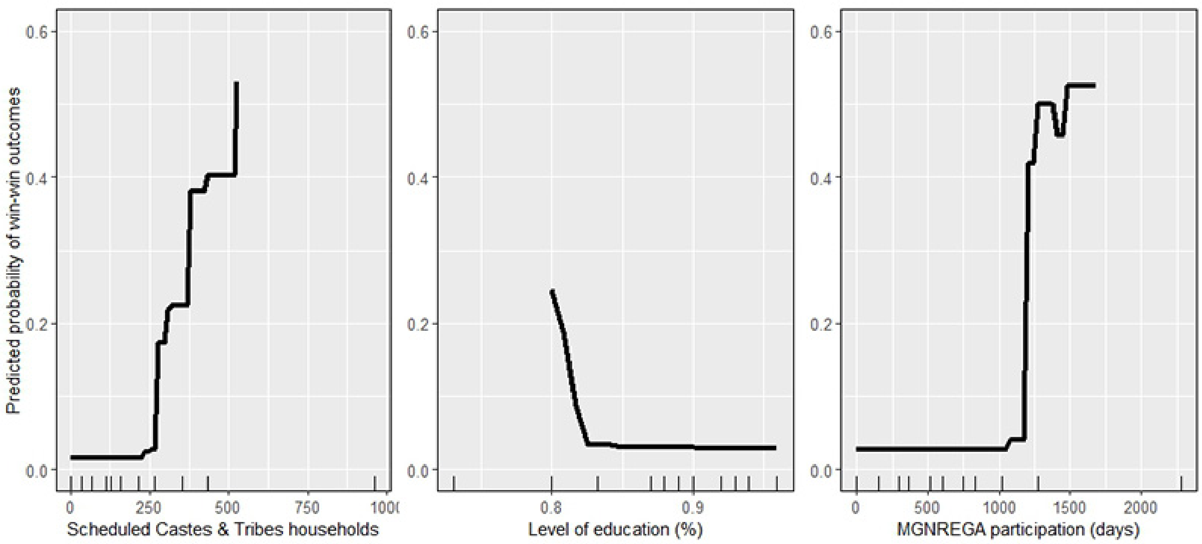
[†] These values have been determined on the basis of the visual inspection of the figures and analysis of tables.
[‡] Predictive effect possible range: 0 to 1.
[§] Livelihood wins depicts high forest dependence on plantations whereas livelihood loses indicates low forest dependence.
[|] Scheduled Castes and Scheduled Tribe households are recognized as socioeconomically disadvantaged communities in India. These communities have been provided reservations in employment and electoral seats and are given special concessions under national social welfare policies and programs.
[¶] MGNREGA refers to the Mahatma Gandhi National Rural Employment Guarantee Act.
[#] LPG refers to liquified petroleum gas.

outcome indicator) included the number of trees (100–2000 in increments of 100), shrinkage (0.001, 0.005, 0.01, or 0.05), the minimum number of observations in a terminal node (three, five, or ten), and interaction depth (one–five). We used a fixed bag fraction of 0.7, which means we selected 70% of the training set observations randomly to propose the next tree in the model expansion considering our comparatively small sample size. We used receiver operating characteristics for model selection, which adjusts for model sensitivity to imbalanced classes (Branco et al. 2016). Tuning each of our four models using similar initial sets of parameters resulted in a separate optimized set of parameters for each model. The details regarding model tuning and optimized parameters selection for all four outcomes (livelihood wins-forest cover wins, livelihood win outcomes, forest cover win outcomes, and livelihood loses-forest cover loses) are described in Appendix 1.

We used the optimized model parameters to calculate the relative influence of the variables on the probability of win–win and win–lose outcomes between tree cover and livelihood benefits outcomes using the GBM package (Greenwell et al. 2024; Table 1). The squared relative importance of the variable is the sum of squared improvements over all the internal nodes for which that variable was chosen as the splitting variable (Hastie et al. 2009). Higher values of relative influence indicate greater effects on the probability of improving plantation outcomes.

We showed the magnitude and direction of each variable on the probability of win–win or win–lose outcomes while averaging the effect of other variables using partial dependence plots (PDPs) using the pdp package in R (Greenwell 2017). These plots help to visualize the magnitude, direction, and critical range of indicator values as well as the thresholds where the effect of a given variable has a maximum impact or changes its direction of impact on the win–win, win–lose

Fig. 3. Partial dependence plots: win-win outcomes (livelihood wins, forest cover wins). The plots shows the effect of a given variable on the probability of win-win outcomes while averaging the effect of other variables. The dark lines are the predicted probabilities of win-win outcomes (probability values range from 0 to 1) as estimated by running generalized boosted regression models (GBM) iteratively. The predictive probabilities range from 0 to 1, with higher values for a particular variable suggesting its greater importance in predicting the win-win outcomes.



or lose–lose outcomes. The PDP shows how the average prediction effect in the dataset changes with the change in a j^{th} variable (Molnar 2022). The range of prediction effect lies between zero and one, and higher values for a particular variable suggest a greater importance of that variable in predicting the outcome. These plots also provide a critical range of variable values (zone of influence) where the effect of a variable is high and thresholds where a variable changes its direction of effect on the probabilities of multiple plantation outcomes (Figs. 3–6; Figs. A1.5–A1.8). To save space, we only show these plots for the three variables with highest predictive importance in each of four plantation outcomes.

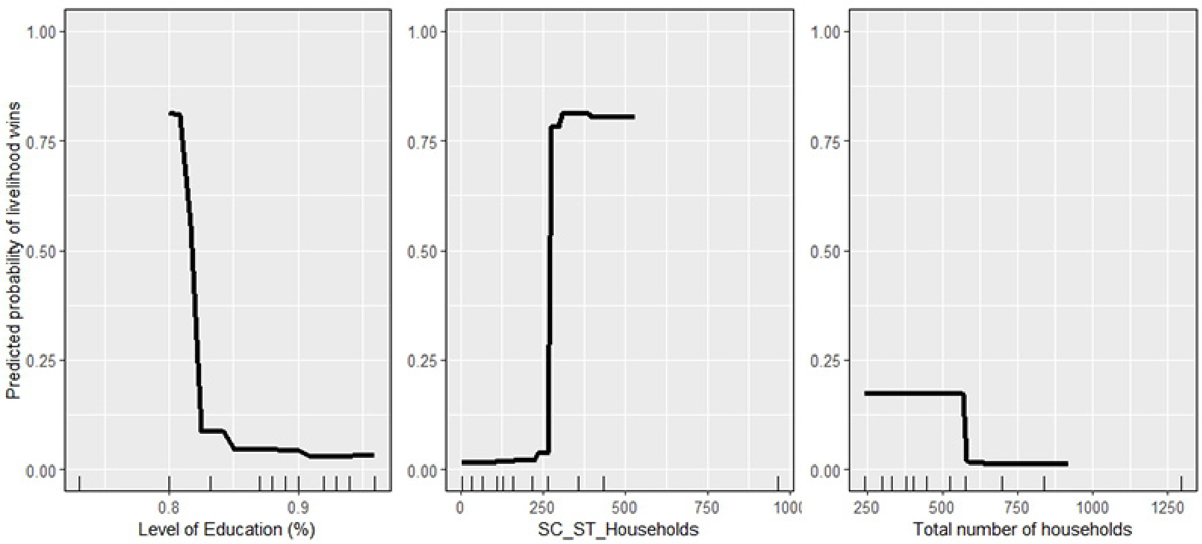
Finally, we estimated the interaction effects between variables using Friedman's H Statistic as well as the zones of maximum interaction effects through bivariate interaction plots in the iml package in R (Molnar et al. 2018; Table 2; Figs. A1.9–A1.12). A Friedman's H statistic of one indicates that the partial dependence between two variables of interest is constant, and the variables only influence the predictions of synergistic or trade-off outcomes through their interaction. On the other hand, a value of zero means there is no interaction between studied variables. We also estimated a range of indicator values for each variable where their interaction strength was highest using bivariate plots (Table 2; Figs. A1.9–A1.12). We present each of the theoretically relevant and hypothesized variables and their interactions in our results to avoid pure data mining.

Estimating constituent and interactive effects of 36 variables as well as differences in functional form in a sample of 377 can strain the data because the tests involve thousands of different combinations of variables. To account for small sample size, we used simple trees and slow learning rate and allowed at least 2000 trees. In addition, we used dummy variables (increase = 1, low forest dependence = 0) as outcome variables as per the requirements of generalized boosted regression modeling (GBM) and to facilitate this analysis given our small dataset. We used one-against-all binary classifications for win–win and lose–lose outcomes and single class models for individual livelihood and forest cover outcomes to improve estimation time and because we have a small dataset, similar set of variables, and fewer classes to train. (Murphy 2012).

GBMs are more robust to smaller datasets and less susceptible to non-normalized data (Friedman 2001, Zou et al. 2022). We addressed overfitting through tenfold stratified cross-validation as part of the tuning process by randomly dividing the set of observations into ten folds (or groups) of approximately equal sizes and then using nine folds for training, reserving one fold for testing. The procedure is repeated five times, each time reserving a different tenth fold/group for testing (Kuhn 2008). In gradient boosting, even if the model fails to accurately predict the outcome class for the first time, it gives more weighting to misclassified observations in the next iterations, thereby increasing its ability to predict the class with low cases. The gradient boosting relies on minimizing loss function of the model by adding weak learners using gradient descent. These weak learners are iteratively added in areas where strong/existing learners perform poorly, and the contribution of each of these weak learners to the final prediction is decided based on a gradient optimization process. This leads to improved ROC value and accuracy of the model through minimization of the overall error of the strong learner (Greenwell et al. 2024, Epstein et al. 2021).

Some of the limitations of the methods include the possibility of bias from parameter tuning because each of the models is individually tuned, and the differences among the models could be influenced to some extent because of parameter tuning

Fig. 4. Partial dependence plots: livelihood win outcomes. The plots shows the effect of a given variable on the probability of livelihood win outcomes while averaging the effect of other variables. The dark lines are the predicted probabilities of livelihood win outcomes (probability value ranges from 0 to 1) as estimated by running generalized boosted regression models (GBM) iteratively. The predictive probabilities range from 0 to 1, with higher values for a particular variable suggesting its greater importance in predicting the livelihood win outcomes.



(Jouffray et al. 2015). We expected this bias to be lower because we followed a systematic model tuning procedure involving 100–1000 combinations of different hyperparameters and the same initial set of parameter values in a grid, and we selected a set of hyperparameters that had the highest tenfold stratified cross-validated ROC while tuning our models. Some of the possible bias is specific to tree-based methods where continuous variables are preferred over ordinal or categorical variables because of the presence of more split points in continuous variables (Strobl et al. 2009), which affects the measures of relative influence.

RESULTS

Relative influence, zones of maximum predicted effect, critical thresholds, and directionality of effect of variables on multiple plantation outcomes

Win–win outcomes

Table 1 shows the range of predictor variables with the highest effects on the outcome variables and describes their associated predicted effects. For win–win outcomes, the presence of Scheduled Castes and Scheduled Tribes households had a relative influence of 22.7%, with its maximum predicted effect between 275 and 527 households. The probability of a win–win outcome from tree planting programs increased with the increasing number of Scheduled Castes and Scheduled Tribes households (Fig. 3). An increase of 100 Scheduled Castes and Scheduled Tribes households in a panchayat increased the average predictive probability of achieving win–win outcomes by 9.0%, and the effect was almost linear.

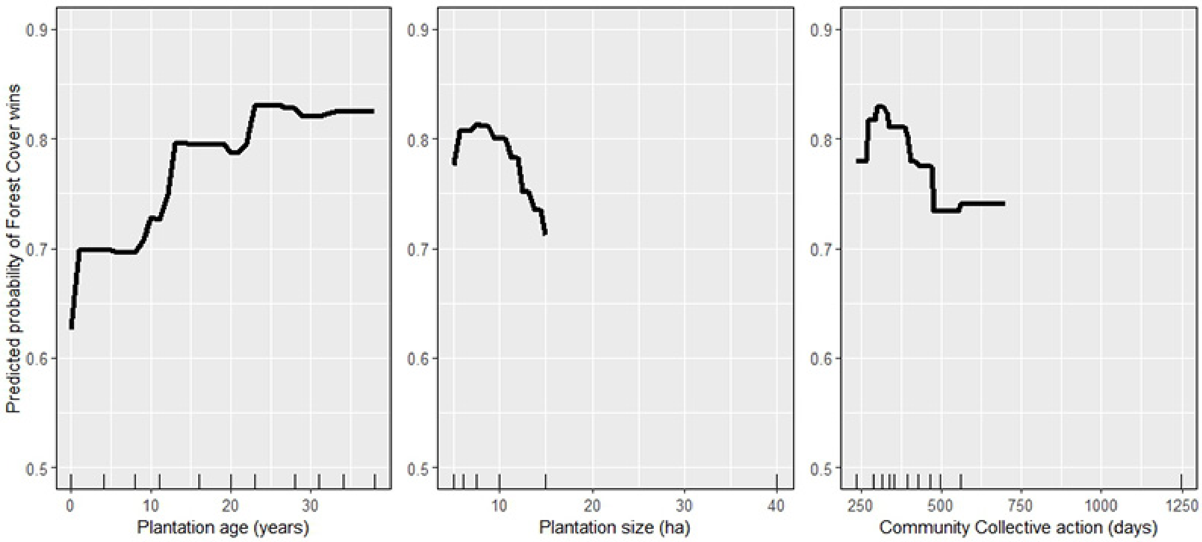
Education had the second highest predictive importance (18.2%), with its maximum predicted effect between 80% and 82%. The probability of a win–win outcome from tree planting programs declined with increasing levels of education (Fig. 3). An increase of 1% in education reduced the average predictive probability of achieving win–win outcomes by 1.4%. The effect of MGNREGA was non-linear (12.7% predictive importance), with insignificant change in the average predictive effect per 100 days until 1176 days. We found a threshold effect: as MGNREGA labor in a panchayat extended beyond 1210 days, the predicted probability of win–win outcomes drastically increased to 2.2% per 100 days, a difference of about twenty-fold (Fig. A1.5).

The other key variables most predictive of win–win outcomes included community collective action days (predictive importance: 8.9%), total households (5.3%), land under cultivation (4.2%), and number of civic groups (3.7%; Table 1). Community collective action days had a non-linear, inverted U-shaped relationship with the predicted probability of win–win outcomes. For each 100 days of increase in collective action days in a panchayat until 423 days, we found a 4% increase in the average predictive probability of achieving win–win outcomes. After 423 days, there was a decrease in the average predictive probability of obtaining win–win outcomes by 1% per 100 days. The average predictive effect for total number of households and land under cultivation was at its maximum between 282–567 households and 6.5–166.2 kanals (1 Kanal = 0.0505 hectares), respectively. Finally, the effect of number of civic groups on win–win outcomes was much higher when there were 4–5 groups. The partial dependence plot shows a negative relationship between the number of civic groups and the probability of win–win outcomes.

Livelihood win outcomes

For livelihood outcomes, we found education had a higher relative influence (29.8%) in changing the probability of win–lose outcomes (Table 1). With each 1% increase in level of education, there was a decline of 0.4% in the average predictive probability of achieving livelihood benefits (Fig. 4).

Fig. 5. Partial dependence plots: forest cover win outcomes. The plots shows the effect of a given variable on the probability of forest cover win outcomes while averaging the effect of other variables. The dark lines are the predicted probabilities of forest cover win outcomes (probability value ranges from 0 to 1) as estimated by running generalized boosted regression models (GBM) iteratively. The predictive probabilities range from 0 to 1, with higher values for a particular variable suggesting its greater importance in predicting the forest cover win outcomes.



The presence of Scheduled Caste and Scheduled Tribe households had a relative influence of 22.9%, with a threshold effect. The predicted probability of livelihood win outcomes drastically increased from 0.007% to 32% per 100 days as the Scheduled Castes and Scheduled Tribes households increased from 265 to 527 days in a panchayat (Fig. 4). The average predicted probability of achieving livelihood win outcomes declined with the increase in the total number of households (predictive importance: 14.8%) in a panchayat.

Community collective action had the next highest predictive importance (7.4%) and showed a threshold effect. The effect of community collective action on the average probability of livelihood win outcomes was 3% until 340 days and then increased to 15% between 349 and 702 days (Fig. A1.6). Similarly, MGNREGA labor days (6.7%) show a threshold effect on livelihood win outcomes, with a 6% increase in average predictive probability of achieving livelihood wins until 1176 days. It then showed a sharp increase to 69% between 1210 and 1681 days. Finally, acreage under cultivation in the panchayats showed a non-linear relationship with the average predictive effect, declining from 39% (≤ 36.4 kanals) to 6% (> 36.4 kanals; Fig. A1.6).

Forest cover win outcomes

Plantation age (10.9%) followed by plantation size (8.8%) had the highest importance in influencing the average predictive probability of forest cover wins. The effect of plantation age had a positive relationship with forest cover wins; for every one-year increase, the average predictive probability of forest cover wins increased by 0.05% (Fig. 5). Plantation size had a non-linear effect on the probability of forest cover gain. As the acreage under plantation increased by 1 ha, there was a decline of 0.05% in the average predictive probability of forest cover gains.

Community collective action days, with next highest importance (6.8%), was a critical variable in changing the probability of forest cover win outcomes. The relationship was non-linear, with an inverted U-shape. The predicted probability of forest cover wins had a 1% increase until 405 days and then a decline of 2% for every 100 days increase in community action days in a panchayat (Fig. 5).

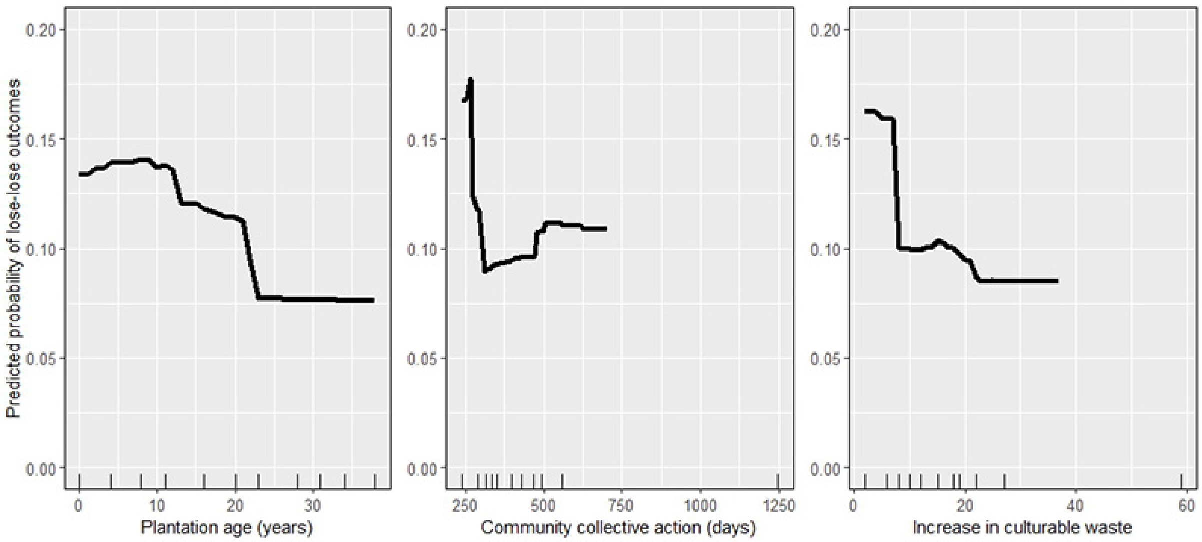
Other variables, in descending order of predictive importance, were the number of Scheduled Caste and Scheduled Tribe households, decrease in livestock, and plantation equitable benefits (Table 1). For every increase of 100 Scheduled Caste and Scheduled Tribe households in a panchayat, there was a decline of 0.08% in the average predictive probability of achieving forest cover wins. On the other hand, as the number of livestock in a panchayat declined by 10, the average predicted effect in the probability of forest gains increased by 2%. The effect of the presence of equitable plantation benefits in a panchayat on the probability of forest cover wins was largely constant.

Lose–lose outcomes

Finally, for lose–lose outcomes with low livelihood benefits and a decline or no change in forest cover, plantation age emerged as a critical variable with the highest importance (8.9%) in changing the probabilities of lose–lose outcomes (Table 1). The average predicted effect for plantation age was 14% until 12 years of age and then declined to 9% (Fig. 6).

Panchayats with a smaller number of community action days (< 265, about half of a typical panchayat) and below average acreage under private grasslands (6–82.1; sample mean: 96.8 kanals) were more likely to witness an increase in lose–lose outcomes (Fig. 6). For every 100 days of decline in a panchayat's community action, the probability of lose–lose outcomes increased by 1%.

Fig. 6. Partial dependence plot: lose-lose outcomes (livelihoods lose, forest cover lose). The plot shows the effect of a given variable on the probability of lose-lose outcomes while averaging the effect of other variables. The dark lines are the predicted probabilities of lose-lose outcomes (probability value ranges from 0 to 1) as estimated by running GBM models iteratively. The predictive probabilities range from 0 to 1 with higher values for a particular variable suggesting its greater importance in predicting the lose-lose outcomes.



In declining order of predictive importance, the other variables included increase in culturable waste, community liquified petroleum gas (LPG) use, and total households (Table 1). There was a small average predicted effect for increase in culturable waste (less increase in cultural waste was associated with a high predictive effect), community LPG use (increase), and total households (inverted U-shaped).

Key variable interactions, effect range, and zones of maximum effect of variables on multiple plantation outcomes

We identified key interactions leading to win–win or lose–lose outcomes between tree cover and livelihoods out of thousands of such possible interactions among 36 variables using Friedman's H statistic (Table 2; Figs. A1.9–A1.12). In the case of win–win outcomes, there was a positive interaction effect of acreage under cultivation and community access rights (H = 0.25), which jointly produced livelihood benefits alongside improved forest cover. There was also a positive interaction effect between acreage under cultivation and plantation species selection by co-management (jointly by local communities and forest officials; H = 0.16) and through forest department (H = 0.26; Fig. A1.9).

The results also showed a positive interaction effect between collective action days and acreage under private grasslands on win–win outcomes but only when there were at least 358 days of collective action (sample mean: 418; H = 0.15). Similarly, there was a positive interaction effect between MGNREGA days and plantation size, leading to the joint production of livelihood benefits alongside improved forest cover. However, this positive interaction effect occurred only when people employed under MGNREGA collectively got at least 1210 job days, twice the average number of MGNREGA labor days in a panchayat, and when the plantation ranged between 5–15 hectares (sample mean: 8.9 ha; H = 0.18; Table 2; Fig. A1.9).

In the case of livelihood win outcomes, plantations yielded large livelihood benefits in panchayats when people employed under MGNREGA collectively got at least 1210 job days (twice the average number of MGNREGA labor days per panchayat) and when the number of civic groups was between 3–22 (sample mean: 15; H = 0.33). Low to high numbers of civic groups (3–27; sample mean: 15) and low acreage under cultivable area (6.5–36.4 kanals; sample mean: 135 kanals) interacted to influence positive livelihood outcomes (H = 0.20; Table 2; Fig. A1.10).

There was a positive interaction effect on livelihood win outcomes when there were moderate to high collective action days in a panchayat (> 358 days; sample mean: 418) and low acreage under cultivable area (6.5–36.4 kanals; sample mean: 135 kanals; H = 0.17). In addition, plantations yielded high livelihood outcomes when the people employed under MGNREGA collectively got at least 1210 job days (twice the average number of MGNREGA labor days thin a panchayat) and when there was a higher number of Scheduled Caste and Scheduled Tribe households in the panchayat (> 275; about 8 times more than a typical panchayat; H = 0.16; Table 2; Fig. A1.10).

With forest cover wins, plantations were associated with forest cover improvement in panchayats where there were fewer households below the poverty line (< 220 households; three times the sample mean) and when there were below average equitable benefits (H = 0.28). There was also a positive interaction effect on forest cover gains when there were low to high equitable benefits from plantations to communities and a decline in the number of livestock in the panchayat by at least 20 (sample mean: 34; H = 0.27) or when the people employed under MGNREGA collectively got up to 1750 days of community collective employment under MGNREGA in the panchayat (mean MGNREGA employment in sample: 690 days; H = 0.26).

Table 2. Key two-variable interactions leading to multiple outcomes in tree plantation programs. For categorical variables, the full range of variable values is included.

Key variable interactions	H Statistic	Value of first variable where interaction effect is maximum	Value of second variable where interaction effect is maximum
Win–win outcomes[†] (livelihood wins, forest cover wins[‡])			
Acreage under cultivable area x community access rights	0.25	6.5–36.4 kanals	0–30
Level of education x increase in culturable waste	0.17	80–81	2–59
Community collective action x acreage under private grasslands	0.15	> 358 days	6–320 kanals
MGNREGA[|] labor days x plantation size	0.14	> 1210 days	5.0–15.0 ha
Livelihood wins[§]			
MGNREGA[|] labor days x civic groups	0.33	>1210 days	3–22 groups
Civic groups x acreage under cultivable area	0.20	3–27 groups	6.5–36.4 kanals
Community collective action x acreage under cultivable area	0.17	> 358 days	6.5–36.4 kanals
MGNREGA[|] labor days x Scheduled Castes and Scheduled Tribes households	0.16	> 1210 days	> 275 households
Forest cover wins			
Below poverty line households x equitable benefits	0.28	< 220 households	2.5
Decrease in livestock x equitable benefits	0.27	> 20	0–10
MGNREGA[|] labor days x equitable benefits	0.26	< 1750 days	2.5–8.0
Decrease in livestock x plantation size	0.25	> 20	< 12 ha
Lose–lose outcomes			
Community LPG[¶] use x plantation age	0.28	> 6	< 22 years
Level of education x increase in culturable waste	0.17	> 0.85	< 8
Acreage under private grasslands x equitable enforcement	0.16	< 82.1	2.5–11
Acreage under private grasslands x community collective action	0.15	< 82.1	< 265 days

[†] Among the win-win outcomes, the following categorical variable interactions are not provided because of the difficulty of determining the variable range where the effect is maximum: plantation species selection (by forest department) x land under cultivation (H = 0.26) and plantation species selection (through comanagement) x land under cultivation (H = 0.16).
[‡] High livelihoods indicate high forest dependence on plantations whereas low livelihoods indicates low forest dependence.
[§] Among the livelihood wins outcomes, the interaction effect between total households and Scheduled Castes and Scheduled Tribes households is not shown as it is not relevant to explaining multiple outcomes.
[|] MGNREGA refers to the Mahatma Gandhi National Rural Employment Guarantee Act.
[¶] LPG refers to liquified petroleum gas.

Moreover, we also found a positive interaction effect on forest cover gains when there was a decline of at least 20 livestock in the panchayat (sample mean: 34) and when the plantation size was larger than 11 hectares (sample mean: 8.9 ha; Table 2; Fig. A1.11).

For lose–lose outcomes, high community LPG use and plantation age interacted positively to influence low livelihood benefits-low forest cover outcomes (H = 0.28). Also, high education and less increase in culturable waste jointly produced lose–lose outcomes (H = 0.17). Lose–lose outcomes occurred where there was low community collective action (< 265; sample mean: 418) and where there was low acreage under private grasslands in the panchayats (< 82.1 kanals; typical in the study sample: 96.78 kanals; H =15). Finally, the presence of below average private grasslands interacted with low equitable enforcement to positively influence lose–lose outcomes (H = 0.16; Table 2; Fig. A1.12).

DISCUSSION

Our study demonstrated the use of interpretable machine learning to identify key predictors and their relative influence as well as the directionality of effects, zones of influence, and critical thresholds associated with multiple plantation outcomes in northern India. We demonstrated how IML tools and approaches can be used through several sequential steps to uncover relationships among variables. The findings can provide useful insights to develop cause and effect hypotheses, the results of which could further inform tree planting policies and programs. Our results point to a range of variables and conditions that are likely to influence livelihood and forest cover outcomes.

Win–win outcomes

The number of Scheduled Caste and Scheduled Tribe households in a panchayat had the highest relative importance in explaining, with a positive relationship to, win–win outcomes. With more Scheduled Castes and Scheduled Tribes households in a panchayat, there is likely to be higher dependence on forest resources and, therefore, high livelihood benefits from plantations planted in those panchayats. A higher proportion of Scheduled Castes and Scheduled Tribes households in a panchayat may also lead to higher group homogeneity and, therefore, higher collective action outcomes on account of common needs, interests, and priorities among these households (Agrawal and Gibson 1999, Poteete and Ostrom 2004). High collective action coupled with effective involvement of these marginalized communities in designing, implementing, and monitoring of plantations may result in improved forest cover outcomes along with positive livelihood benefits (Fleischman et al. 2022, Löfqvist et al. 2023).

The level of education in a panchayat had the second highest relative importance in explaining, with a negative relationship to, win–win outcomes. With more education, there is likely to be less dependence of households on forest resources. It is unclear whether education increases forest degradation (by undermining collective action) or decreases it (by decreasing dependence on forests). Recent studies have found education to be a critical variable promoting the adoption of cleaner fuel options such as LPG, which reduces the dependence of rural communities on forest resources (DeFries et al. 2021, Khanwilkar et al. 2021). In other words, our measure of dependence means that less livelihood benefit could simply indicate that the community is not forest-dependent, not necessarily that the forest is worse for people who are forest-dependent.

In panchayats where communities have higher access to alternative sources of income through public employment programs (especially in contexts with lower educational attainment), we found a higher probability of achieving win–win outcomes. This may be due to an overall decline in plantation-based resource use because of the presence of alternative off-farm income under conditions where access to other, more skilled employment options are limited (Rana and Miller 2019b).

Our work suggests that a more consolidated institutional space along with a smaller number of civic groups is more conducive to win–win outcomes overall. Panchayats with above average collective action and few civic groups (one-third to one-half of the average civic groups in the study sample) perform better, which may be a result of better coordination for communal monitoring, enforcement of plantation enclosures, or other management activities. This finding supports earlier arguments that a proliferation of local user groups might undermine effective decentralization (Manor 2004). Our findings suggest that a more consolidated institutional space may be particularly important where local communities have a low to moderate amount of area under agricultural cultivation—where fewer productive assets are likely associated with greater reliance on plantations, rather than markets, for subsistence needs. Higher collective action can also enable communities to bargain with local forest rangers to plant locally valued species in places that do not interfere with other land uses, which may help to support greater tree survival, a higher level of local legitimacy, and greater net livelihood gains (Rana and Miller 2021).

In sum, in low education settings and low acreage under cultivation in less-populated panchayats, the results show that universal wage generation program (MGNREGA) support marginalized populations, such as Scheduled Castes and Scheduled Tribes households, to collectively act through consolidated institutional space to achieve win–win outcomes in tree plantation programs. Moreover, such win–win outcomes are more likely if these marginalized populations have secure access rights to forest and plantation resources and if the overall population is not so high as to disincentivize individuals to contribute toward collective outcomes because of a decline in the availability of per capita resource benefits (Agrawal and Gibson 1999, Poteete and Ostrom 2004).

Livelihood wins, forest cover wins and lose–lose outcomes

We found a high predictive probability of large livelihood benefits from tree planting programs in panchayats with low levels of education in less populated panchayats having high proportions of Scheduled Caste and Scheduled Tribe households. This supports the existing evidence that people who are poor, illiterate, and socially and economically marginalized are more dependent on forests for their fuelwood, fodder, and small timber needs and, therefore, are likely to get higher livelihood benefits from plantations (Rana and Miller 2019a, 2021, Löfqvist et al. 2023).

Our results also indicated improved livelihood benefits in panchayats where community collective action was very high (> 423 days), people collectively got high labor days under MGNREGA (>1210 days), and communities had low acreage under cultivation area (about one-fourth of the sample mean). High collective action may enable effective management of forest and plantation resource use for subsistence needs (Rana and Miller 2021) and may also empower local people to demand higher wage employment under MGNREGA (Carswell and De Neve 2014, Fischer and Ali 2019). In sum, there are improved livelihood benefits for marginalized populations (Scheduled Castes and Scheduled Tribes households) in areas with low cultivation acreage and where they are able to organize into fewer civic groups and collectively act to get more access to MGNREGA employment. Local communities, especially low-income groups, are likely to have higher bargaining power vis-a-vis local forest officials in the presence of a higher level of collective action in a panchayat, which enables these communities to extract high levels of livelihood benefits from planted enclosures to meet requirements for sustaining livestock-based livelihoods (Carswell and De Neve 2014).

In the case where plantations yielded high forest cover, we found a greater influence of higher age of plantations, high collective action, and small-sized plantations. The older the plantation, the higher are the chances of the survival of the planted seedlings and, therefore, the higher is the likely forest cover (Rana and Miller 2021). Collective action may enable effective management of small-sized plantations, especially for subsistence needs, despite no improvement in forest cover (Rana and Miller 2021). Communities may use their own resources for livestock production in places where they have high resource endowments, thereby reducing overall livelihood dependence on resources. Nevertheless, there was still a decline or no change in forest cover in some contexts, likely due to high community resource use, which may not be met through private resources.

We found an increase in forest cover where there were fewer resource users who were likely to have high levels of dependence, such as Scheduled Caste and Scheduled Tribe households, and where there was a decline in livestock. In cases where there is not adequate attention to providing alternative grazing or access to alternative forest areas for resource use, tree plantations may be less likely to survive (Rana and Miller 2019a, 2021). Rapid decline in the number of livestock is one of the critical reasons for reduced grazing inside forest areas and, therefore, for higher success rates of plantations and improved forest cover (Rana and Miller 2021). Forest cover improvement was more likely in places where livestock numbers were declining, plantations were grown on small-sized plantations, and where low-income and marginal populations were expected to get equitable benefits as well as higher access to MGNREGA wage employment.

We found that low levels of community collective action, limited private grassland availability, and large-sized plantations in high populated panchayats were all associated with lose–lose outcomes. Lower levels of community collective action may be associated with a lower ability of communities to collectively mobilize for tree species that are valuable for local livelihoods and to effectively protect larger-sized plantations (Rana and Miller 2021). Moreover, communities may not be able to fully divert their livestock grazing from plantation enclosures, especially where plantations are larger in size and where they have less acreage under private grasslands, even despite increasing LPG usage (Rana and Miller 2019a, 2021).

POLICY RECOMMENDATIONS AND CONCLUSION

Our analysis leads to some practical suggestions for designing and effectively implementing win–win nature climate solutions wherein tree planting programs are likely to contribute to the goals of sequestering carbon while benefiting local communities. First, scholars and practitioners should assume that there are trade-offs between forest restoration and livelihood goals unless programs are explicitly designed to improve both. Involving multiple stakeholders including local communities, governance institutions, private enterprises, and non-governmental organizations in tree planting interventions is likely to increase the probability that these trade-offs are considered and addressed in program design (Sarin et al. 2003, Rana and Miller 2019a, Ramprasad et al. 2020).

Second, governments may better promote livelihood benefits and improvement in forest cover simultaneously by supporting existing collective action practices specific to the areas under consideration. This aligns with findings from existing research, and it affirms that supporting local management is relevant not just for protecting forests but also for active restoration and plantation activities. Millions of hectares of forested land have been transferred to local communities under forest decentralization in the global south (MacDicken et al. 2016), wherein a large emphasis is given to creating or strengthening existing formal or informal institutions or traditional community practices to ensure effective management of forests (Ostrom 1990, Agrawal and Chhatre 2006, Chhatre and Agrawal 2008). Thus, aside from large-scale projections of restoration potential (Bastin et al. 2019, Busch et al. 2019), our results suggest that the quality and extent of local stakeholder engagement are likely to be among the most important variables in the success of restoration activities in many rural landscapes.

Third, our results have insight for the growing literature on restoration governance (Chazdon et al. 2020, Mansourian et al. 2020). In particular, we found that the presence of a large number of civic institutions risks fragmenting decision-making space, creating rivalries and conflicts among users due to contradictory objectives. It may also exhaust participants due to multiple and diverging institutional platforms (Sarin et al. 2003, Manor 2004, Lubell et al. 2010, Mewhirter et al. 2019). On the other hand, a more consolidated institutional space under forest decentralization reforms, national forestry programs, and other global efforts may help to foster engagement across divergent interests and promote more frequent interactions to coordinate management activities across different social groups within a community (Poteete and Ostrom 2004, Adhikari and Lovett 2006).

Fourth, our results show that existing community resource endowments, especially private grasslands, are critical in regions such as northern India where local communities still depend upon livestock. In such contexts, people need substantial grasslands for their livestock, and conversion of grasslands into woodlots can lead to lower livelihood benefits where alternative options are not available. This underscores the continuing need to balance multiple needs through different landscape types (Rana and Miller 2019a, Ramprasad et al. 2020) and to strengthen community tenure in order to ensure community access to forest resources and support greater local investment in collective management (Agrawal et al. 2008).

Finally, our results suggest that other, non-forest policy mechanisms to promote rural welfare may also help amplify existing supportive conditions for joint positive outcomes (Fischer and Ali 2019, Ferraro and Simorangkir 2020, DeFries et al. 2021). In particular, we found that a more robust social safety net (such as MGNREGA) may help to reduce dependence on forest resources for the poor and marginal, thus making it more possible to support livelihoods at a base level while also achieving forest growth. Similar results have been observed in other contexts where conditional cash transfers to reduce poverty also led to a decline in deforestation (Ferraro and Simorangkir 2020), and providing free LPG to rural communities for cooking also led to protecting forests as a side benefit (DeFries et al. 2021).

Our results show that low forest dependence (low livelihood benefits) mostly co-occurs with positive reforestation outcomes. This may mean that interventions or factors that reduce forest dependence can support forest restoration. For example, we found that tree plantations that have occurred in places where people have a high level of education may be more likely to lead to improved forest cover over the long term (Rana and Miller 2019b). Higher levels of education may also promote higher use of LPG in household cooking (Khanwilkar et al. 2021), more off-farm employment, and increased participation in labor employment programs, which may lower firewood and other resource dependence on plantation enclosures and lead to improved forest regeneration. This is indeed an interesting outcome, because it points to the potential importance of non-forestry interventions (in this case access to cooking fuel) that may help to support restoration outcomes. We argue that governments or international agencies should spend resources on improving education, helping to provide pathways toward more remunerative off-farm livelihoods, and promoting alternative clean cooking fuel options (Khanwilkar et al. 2021), which may help to increase tree cover of degraded forest landscapes.

Care should be taken to interpret our findings because our work is IML-driven prediction analysis and does not itself imply causal effects. The results here are illustrative of models that might be applied in other contexts to uncover new associations and build theory on underlying factors and interactions that shape human-environmental outcomes through plantation and restoration. Future research should focus on developing new methodologies to improve the causal interpretation of machine learning-based research (Rana and Miller 2019a, Hofman et al. 2021). By allowing us to probe different constellations of variables, assess their relative importance and direction of relationships, and uncover critical thresholds and key interactions, IML-based, data-driven methods hold great potential for moving beyond mono-causal explanations of forest and landscape change. These methods can help generate new questions and hypotheses, which can then be tested or validated using causal inference tools and approaches to accelerate social-ecological research.

Author Contributions:

P. R. conceived of the study, designed the analysis, and analyzed the data; H. F., E. A. C., and F. F. provided critical analytical insights; P. R., H. F., E. A. C., and F. F. wrote the paper.

Acknowledgments:

This work has been supported by NASA Land Cover and Land Use Change grants NNX17AK14G and 21-SARI-21-0005, FORMAS grant 2022-00651, and the Swedish Research Council (Vetenskapsraädet) grants 2018-05875 and 2022-04581.

Data Availability:

All data, code, and materials used in the analyses will be available to any researcher on request to the corresponding author on publication of the paper.

LITERATURE CITED

Adams, C., S. T. Rodrigues, M. Calmon, and C. Kumar. 2016. Impacts of large-scale forest restoration on socioeconomic status and local livelihoods: what we know and do not know. Biotropica 48(6):731-744. https://doi.org/10.1111/btp.12385

Adhikari, B., and J. C. Lovett. 2006. Institutions and collective action: does heterogeneity matter in community-based resource management? Journal of Development Studies 42(3):426-445. https://doi.org/10.1080/00220380600576201

Agrawal, A. 2001. Common property institutions and sustainable governance of resources. World Development 29(10):1649-1672. https://doi.org/10.1016/S0305-750X(01)00063-8

Agrawal, A., and A. Chhatre. 2006. Explaining success on the commons: community forest governance in the Indian Himalaya. World Development 34(1):149-166. https://doi.org/10.1016/j.worlddev.2005.07.013

Agrawal, A., and A. Chhatre. 2011. Against mono-consequentialism: multiple outcomes and their drivers in social-ecological systems. Global Environmental Change 21(1):1-3. https://doi.org/10.1016/j.gloenvcha.2010.12.007

Agrawal, A., A. Chhatre, and R. Hardin. 2008. Changing governance of the world's forests. Science 320(5882):1460-1462. https://doi.org/10.1126/science.1155369

Agrawal, A., and C. C. Gibson. 1999. Enchantment and disenchantment: the role of community in natural resource conservation. World Development 27(4):629-649. https://doi.org/10.1016/S0305-750X(98)00161-2

Andersson, K. P. 2004. Who talks with whom? The role of repeated interactions in decentralized forest governance. World Development 32(2):233-249. https://doi.org/10.1016/j.worlddev.2003.07.007

Athey, S., and G. Imbens. 2016. Recursive partitioning for heterogeneous causal effects. Proceedings of the National Academy of Sciences 113(27):7353-7360. https://doi.org/10.1073/pnas.1510489113

Bastin, J.-F., Y. Finegold, C. Garcia, D. Mollicone, M. Rezende, D. Routh, C. M. Zohner, and T. W. Crowther. 2019. The global tree restoration potential. Science 365(6448):76-79. https://doi.org/10.1126/science.aax0848

Baynes, J., J. Herbohn, C. Smith, R. Fisher, and D. Bray. 2015. Key factors which influence the success of community forestry in developing countries. Global Environmental Change 35:226-238. https://doi.org/10.1016/j.gloenvcha.2015.09.011

Brancalion, P. H., and K. D. Holl. 2020. Guidance for successful tree planting initiatives. Journal of Applied Ecology 57 (12):2349-2361. https://doi.org/10.1111/1365-2664.13725

Brancalion, P. H., A. Niamir, E. Broadbent, R. Crouzeilles, F. S. Barros, A. M. A. Zambrano, A. Baccini, J. Aronson, S. Goetz, and J. L. Reid. 2019. Global restoration opportunities in tropical rainforest landscapes. Science Advances 5(7):eaav3223. https://doi.org/10.1126/sciadv.aav3223

Branco, P., L. Torgo, and R. P. Ribeiro. 2016. A survey of predictive modeling on imbalanced domains. ACM Computing Surveys (CSUR) 49(2):1-50. https://doi.org/10.1145/2907070

Busch, J., J. Engelmann, S. C. Cook-Patton, B. W. Griscom, T. Kroeger, H. Possingham, and P. Shyamsundar. 2019. Potential for low-cost carbon dioxide removal through tropical reforestation. Nature Climate Change 9(6):463. https://doi.org/10.1038/s41558-019-0485-x

Carswell, G., and G. De Neve. 2014. MGNREGA in Tamil Nadu: a story of success and transformation? Journal of Agrarian Change 14(4):564-585. https://doi.org/10.1111/joac.12054

Chazdon, R. L., S. J. Wilson, E. Brondizio, M. R. Guariguata, and J. Herbohn. 2020. Key challenges for governing forest and landscape restoration across different contexts. Land Use Policy 104:104854. https://doi.org/10.1016/j.landusepol.2020.104854

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins. 2018. Double/debiased machine learning for treatment and structural parameters. Oxford University Press, Oxford, UK. https://doi.org/10.3386/w23564

Chhatre, A., and A. Agrawal. 2008. Forest commons and local enforcement. Proceedings of the National Academy of Sciences 105(36):13286-13291. https://doi.org/10.1073/pnas.0803399105

Chhatre, A., and A. Agrawal. 2009. Trade-offs and synergies between carbon storage and livelihood benefits from forest commons. Proceedings of the National Academy of Sciences 106 (42):17667-17670. https://doi.org/10.1073/pnas.0905308106

Coleman, E. A. 2011. Common property rights, adaptive capacity, and response to forest disturbance. Global Environmental Change 21(3):855-865. https://doi.org/10.1016/j.gloenvcha.2011.03.012

Coleman, E. A., and S. S. Liebertz. 2014. Property rights and forest commons. Journal of Policy Analysis and Management 33 (3):649-668. https://doi.org/10.1002/pam.21766

Coleman, E. A., B. Schultz, V. Ramprasad, H. Fischer, P. Rana, A. M. Filippi, B. Güneralp, A. Ma, C. Rodriguez Solorzano, V. Guleria, R. Rana, and F. Fleischman. 2021a. Limited effects of tree planting on forest canopy cover and rural livelihoods in Northern India. Nature Sustainability 4:997-1004. https://doi.org/10.1038/s41893-021-00761-z

Coleman, E., B. Schultz, V. Ramprasad, H. Fischer, P. Rana, A. Filippi, B. Güneralp, A. Ma, C. R. Solorzano, and V. Guleria. 2021b. Data for decades of tree planting in Northern India had little effect on forest density and rural livelihoods. Repository for the University of Minnesota, Minneapolis-Saint Paul, Minnesota, USA. https://doi.org/10.13020/j6sj-jw18

DeFries, R., M. Agarwala, S. Baquie, P. Choksi, S. Khanwilkar, P. Mondal, H. Nagendra, and J. Uperlainen. 2021. Improved household living standards can restore dry tropical forests. Biotropica 54:1480-1490. https://doi.org/10.1111/btp.12978

Di Sacco, A., K. A. Hardwick, D. Blakesley, P. H. Brancalion, E. Breman, L. Cecilio Rebola, S. Chomba, K. Dixon, S. Elliott, and G. Ruyonga. 2021. Ten golden rules for reforestation to optimize carbon sequestration, biodiversity recovery and livelihood benefits. Global Change Biology 27(7):1328-1348. https://doi.org/10.1111/gcb.15498

Elith, J., J. R. Leathwick, and T. Hastie. 2008. A working guide to boosted regression trees. Journal of Animal Ecology 77 (4):802-813. https://doi.org/10.1111/j.1365-2656.2008.01390.x

Epstein, G., G. Gurney, S. Chawla, J. M. Anderies, J. Baggio, H. Unnikrishnan, S. V. Tomas, and G. S. Cumming. 2021. Drivers of compliance monitoring in forest commons. Nature Sustainability 4(5):450-456. https://doi.org/10.1038/s41893-020-00673-4

Erbaugh, J. T., N. Pradhan, J. Adams, J. A. Oldekop, A. Agrawal, D. Brockington, R. Pritchard, and A. Chhatre. 2020. Global forest restoration and the importance of prioritizing local communities. Nature Ecology & Evolution 4(11):1472-1476. https://doi.org/10.1038/s41559-020-01282-2

Fargione, J. E., S. Bassett, T. Boucher, S. D. Bridgham, R. T. Conant, S. C. Cook-Patton, P. W. Ellis, A. Falcucci, J. W. Fourqurean, and T. Gopalakrishna. 2018. Natural climate solutions for the United States. Science Advances 4(11):eaat1869. https://doi.org/10.1126/sciadv.aat1869

Ferraro, P. J., and R. Simorangkir. 2020. Conditional cash transfers to alleviate poverty also reduced deforestation in Indonesia. Science Advances 6(24):eaaz1298. https://doi.org/10.1126/sciadv.aaz1298

Fischer, H. W., and S. S. Ali. 2019. Reshaping the public domain: decentralization, the Mahatma Gandhi National Rural Employment Guarantee Act (MGNREGA), and trajectories of local democracy in rural India. World Development 120:147-158. https://doi.org/10.1016/j.worlddev.2018.09.013

Fleischman, F., E. Coleman, H. Fischer, P. Kashwan, M. Pfeifer, V. Ramprasad, C. Rodriguez Solorzano, and J. W. Veldman. 2022. Restoration prioritization must be informed by marginalized people. Nature 607(7918):E5-E6. https://doi.org/10.1038/s41586-022-04733-x

Forest Survey of India. 2019. The state of forest report. Ministry of Environment and Forests and Climate Change, Dehradun, India. https://static.pib.gov.in/WriteReadData/userfiles/ISFR2019%20Vol-I.pdf and https://static.pib.gov.in/WriteReadData/userfiles/ISFR2019%20Vol-II.pdf

Friedman, J. H. 2001. Greedy function approximation: a gradient boosting machine. Annals of Statistics 29:1189-1232. https://doi.org/10.1214/aos/1013203451

Gibson, C. C., J. T. Williams, and E. Ostrom. 2005. Local enforcement and better forests. World Development 33 (2):273-284. https://doi.org/10.1016/j.worlddev.2004.07.013

Greenwell, B., B. Boehmke, J. Cunningham, and G. B. M. Developers. 2024. gbm: generalized boosted regression models. R package version 2(5). R Foundation for Statistical Computing, Vienna, Austria. https://cran.r-project.org/web/packages/gbm/gbm.pdf

Greenwell, B. M. 2017. pdp: an R package for constructing partial dependence plots. R Journal 9(1):421-436. https://journal.r-project.org/archive/2017/RJ-2017-016/RJ-2017-016.pdf https://doi.org/10.32614/RJ-2017-016

Griscom, B. W., J. Adams, P. W. Ellis, R. A. Houghton, G. Lomax, D. A. Miteva, W. H. Schlesinger, D. Shoch, J. V. Siikamäki, and P. Smith. 2017. Natural climate solutions. Proceedings of the National Academy of Sciences 114(44):11645-11650. https://doi.org/10.1073/pnas.1710465114

Gundimeda, H., and P. Shyamsundar. 2012. Forests, sustainability and poverty in India. Environment and Development Economics 17(3):373-378. https://doi.org/10.1017/S1355770X12000162

Hastie, T., R. Tibshirani, and J. Friedman. 2009. The elements of statistical learning: data mining, inference, and prediction. Springer, New York, New York, USA.

Hofman, J. M., D. J. Watts, S. Athey, F. Garip, T. L. Griffiths, J. Kleinberg, H. Margetts, S. Mullainathan, M. J. Salganik, and S. Vazire. 2021. Integrating explanation and prediction in computational social science. Nature 595:181-188. https://doi.org/10.1038/s41586-021-03659-0

Howe, C., H. Suich, B. Vira, and G. M. Mace. 2014. Creating win-wins from trade-offs? Ecosystem services for human well-being: a meta-analysis of ecosystem service trade-offs and synergies in the real world. Global Environmental Change 28:263-275. https://doi.org/10.1016/j.gloenvcha.2014.07.005

Jouffray, J.-B., M. Nyström, A. V. Norström, I. D. Williams, L. M. Wedding, J. N. Kittinger, and G. J. Williams. 2015. Identifying multiple coral reef regimes and their drivers across the Hawaiian archipelago. Philosophical Transactions of the Royal Society B: Biological Sciences 370(1659):20130268. https://doi.org/10.1098/rstb.2013.0268

Khanwilkar, S., C. F. Gould, R. DeFries, B. Habib, and J. Urpelainen. 2021. Firewood, forests, and fringe populations: exploring the inequitable socioeconomic dimensions of liquified petroleum gas (LPG) adoption in India. Energy Research & Social Science 75:102012. https://doi.org/10.1016/j.erss.2021.102012

Kuhn, M. 2008. Building predictive models in R using the caret package. Journal of Statistical Software 28(5):1-26. https://doi.org/10.18637/jss.v028.i05

Löfqvist, S., F. Kleinschroth, A. Bey, A. de Bremond, R. DeFries, J. Dong, F. Fleischman, S. Lele, D. A. Martin, and P. Messerli. 2023. How social considerations improve the equity and effectiveness of ecosystem restoration. BioScience 73(2):134-148. https://doi.org/10.1093/biosci/biac099

Lubell, M., A. D. Henry, and M. McCoy. 2010. Collaborative institutions in an ecology of games. American Journal of Political Science 54(2):287-300. https://doi.org/10.1111/j.1540-5907.2010.00431.x

MacDicken, K., Ö. Jonsson, L. Piña, S. Maulo, V. Contessa, Y. Adikari, M. Garzuglia, E. Lindquist, G. Reams, and R.

D'Annunzio. 2016. Global forest resources assessment 2015: how are the world's forests changing? Second edition. Food and Agriculture Organization of the United Nations, Rome, Italy. https://www.fao.org/3/i4793e/i4793e.pdf

Malkamäki, A., D. D'Amato, N. J. Hogarth, M. Kanninen, R. Pirard, A. Toppinen, and W. Zhou. 2018. A systematic review of the socio-economic impacts of large-scale tree plantations, worldwide. Global Environmental Change 53:90-103. https://doi.org/10.1016/j.gloenvcha.2018.09.001

Manor, J. 2004. User committees: a potentially damaging second wave of decentralisation? European Journal of Development Research 16(1):192-213. https://doi.org/10.1080/09578810410001688806

Mansourian, S., J. Parrotta, P. Balaji, I. Bellwood-Howard, S. Bhasme, R. P. Bixler, A. K. Boedhihartono, R. Carmenta, T. Jedd, and W. de Jong. 2020. Putting the pieces together: integration for forest landscape restoration implementation. Land Degradation & Development 31(4):419-429. https://doi.org/10.1002/ldr.3448

Mewhirter, J., E. A. Coleman, and R. Berardo. 2019. Participation and political influence in complex governance systems. Policy Studies Journal 47(4):1002-1025. https://doi.org/10.1111/psj.12227

Miller, D. C., and R. Hajjar. 2020. Forests as pathways to prosperity: empirical insights and conceptual advances. World Development 125:104647. https://doi.org/10.1016/j.worlddev.2019.104647

Miller, D. C., P. Rana, K. Nakamura, S. Irwin, S. H. Cheng, S. Ahlroth, and E. Perge. 2021. A global review of the impact of forest property rights interventions on poverty. Global Environmental Change 66:102218. https://doi.org/10.1016/j.gloenvcha.2020.102218

Molnar, C. 2022. Interpretable machine learning. A guide for making black box models explainable. Second edition. Leanpub, Victoria, British Columbia, Canada.

Molnar, C., G. Casalicchio, and B. Bischl. 2018. iml: an R package for interpretable machine learning. Journal of Open Source Software 3(26):786. https://doi.org/10.21105/joss.00786

Murdoch, W. J., C. Singh, K. Kumbier, R. Abbasi-Asl, and B. Yu. 2019. Definitions, methods, and applications in interpretable machine learning. Proceedings of the National Academy of Sciences 116(44):22071-22080. https://doi.org/10.1073/pnas.1900654116

Murphy, K. P. 2012. Machine learning: a probabilistic perspective. MIT Press, Cambridge, Massachusetts, USA.

Newton, P., J. A. Oldekop, G. Brodnig, B. K. Karna, and A. Agrawal. 2016. Carbon, biodiversity, and livelihoods in forest commons: synergies, trade-offs, and implications for REDD+. Environmental Research Letters 11(4):044017. https://doi.org/10.1088/1748-9326/11/4/044017

Ostrom, E. 1990. Governing the commons: the evolution of the commons for collective action. Cambridge University Press, Cambridge, UK.

Ostrom, E. 2009. A general framework for analyzing sustainability of social-ecological systems. Science 325 (5939):419-422. https://doi.org/10.1126/science.1172133

Ostrom, E., and H. Nagendra. 2006. Insights on linking forests, trees, and people from the air, on the ground, and in the laboratory. Proceedings of the National Academy of Sciences 103 (51):19224-19231. https://doi.org/10.1073/pnas.0607962103

Persha, L., A. Agrawal, and A. Chhatre. 2011. Social and ecological synergy: local rulemaking, forest livelihoods, and biodiversity conservation. Science 331(6024):1606-1608. https://doi.org/10.1126/science.1199343

Pichler, M., M. Bhan, and S. Gingrich. 2021. The social and ecological costs of reforestation. Territorialization and industrialization of land use accompany forest transitions in Southeast Asia. Land Use Policy 101:105180. https://doi.org/10.1016/j.landusepol.2020.105180

Poteete, A. R., and E. Ostrom. 2004. Heterogeneity, group size and collective action: the role of institutions in forest management. Development and Change 35(3):435-461. https://doi.org/10.1111/j.1467-7660.2004.00360.x

Pritchard, R. 2021. Politics, power and planting trees. Nature Sustainability 4:1. https://doi.org/10.1038/s41893-021-00769-5

Ramprasad, V., A. Joglekar, and F. Fleischman. 2020. Plantations and pastoralists: afforestation activities make pastoralists in the Indian Himalaya vulnerable. Ecology and Society 25(4):1. https://doi.org/10.5751/ES-11810-250401

Rana, P., and D. C. Miller. 2019b. Explaining long-term outcome trajectories in social-ecological systems. PLoS ONE 14(4): e0215230. https://doi.org/10.1371/journal.pone.0215230

Rana, P., and D. C. Miller. 2019a. Machine learning to analyze the social-ecological impacts of natural resource policy: insights from community forest management in the Indian Himalaya. Environmental Research Letters 14(2):024008. https://doi.org/10.1088/1748-9326/aafa8f

Rana, P., and D. C. Miller. 2021. Predicting the long-term social and ecological impacts of tree-planting programs: evidence from northern India. World Development 140:105367. https://doi.org/10.1016/j.worlddev.2020.105367

Rana, P., and L. R. Varshney. 2020. Trustworthy predictive algorithms for complex forest system decision-making. Frontiers in Forests and Global Change 3:153. https://doi.org/10.3389/ffgc.2020.587178

Rana, P., and L. R. Varshney. 2023. Exploring limits to tree planting as a natural climate solution. Journal of Cleaner Production 384:135566. https://doi.org/10.1016/j.jclepro.2022.135566

Ridgeway, G. 2024. gbm: generalized boosted regression models. R package version 2.1.9. R Foundation for Statistical Computing, Vienna, Austria. https://cran.r-project.org/web/packages/gbm/gbm.pdf

Roe, S., C. Streck, R. Beach, J. Busch, M. Chapman, V. Daioglou, A. Deppermann, J. Doelman, J. Emmet-Booth, and J. Engelmann. 2021. Land-based measures to mitigate climate change: potential and feasibility by country. Global Change Biology 27(23):6025-6058. https://doi.org/10.1111/gcb.15873

Sarin, M., N. M. Singh, N. Sundar, and R. K. Bhogal. 2003. Devolution as a threat to democratic decision-making in forestry?

Findings from three states in India. Citeseer. https://odi.cdn.ngo/media/documents/2436.pdf

Scheidel, A., and S. Gingrich. 2020. Toward sustainable and just forest recovery: research gaps and potentials for knowledge integration. One Earth 3(6):680-690. https://doi.org/10.1016/j.oneear.2020.11.005

Schultz, B., D. Brockington, E. A. Coleman, I. Djenontin, H. W. Fischer, F. Fleischman, P. Kashwan, K. Marquardt, M. Pfeifer, and R. Pritchard. 2022. Recognizing the equity implications of restoration priority maps. Environmental Research Letters 17 (11):114019. https://doi.org/10.1088/1748-9326/ac9918

Seddon, N., A. Chausson, P. Berry, C. A. Girardin, A. Smith, and B. Turner. 2020. Understanding the value and limits of nature-based solutions to climate change and other global challenges. Philosophical Transactions of the Royal Society B 375 (1794):20190120. https://doi.org/10.1098/rstb.2019.0120

Shyamsundar, P., F. Cohen, T. M. Boucher, T. Kroeger, J. T. Erbaugh, G. Waterfield, C. Clarke, S. C. Cook-Patton, E. Garcia, and K. Juma. 2022. Scaling smallholder tree cover restoration across the tropics. Global Environmental Change 76:102591. https://doi.org/10.1016/j.gloenvcha.2022.102591

Strassburg, B. B., A. Iribarrem, H. L. Beyer, C. L. Cordeiro, R. Crouzeilles, C. C. Jakovac, A. B. Junqueira, E. Lacerda, A. E. Latawiec, and A. Balmford. 2020. Global priority areas for ecosystem restoration. Nature 586(7831):724-729. https://doi.org/10.1038/s41586-020-2784-9

Strobl, C., J. Malley, and G. Tutz. 2009. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. Psychological Methods 14(4):323. https://doi.org/10.1037/a0016973

Zou, M., W.-G. Jiang, Q.-H. Qin, Y.-C. Liu, and M.-L. Li. 2022. Optimized XGBoost model with small dataset for predicting relative density of Ti-6Al-4V parts manufactured by selective laser melting. Materials 15(15):5298. https://doi.org/10.3390/ma15155298

Appendix 1

Supplementary Materials for

Using machine learning to uncover synergies between forest restoration and livelihood support in the Himalayas

This PDF file includes:

Supplementary Text (with Figures A1.1 to A1.4)

Figures A1.5 to A1.12

Tables A1.1 to A1.3

Factor Analysis and Tables A1.4 to A1.6; Figures A1.13 to A1.16

References

Supplementary Text

Study area and data

We collected data during 2018-2019 in Kangra District in the northern Indian state of Himachal Pradesh. India has shown increases in forest cover lately with Himachal Pradesh ranked in the top five states for forest cover increase during last decade(Forest Survey of India 2019). Kangra, one of the 12 districts of Himachal Pradesh, is chosen for this study due to its mountainous context with large scale biophysical, socio-economic and ecological heterogeneity. We explored the livelihood impacts of tree plantations on rural communities in the Kangra district using long-term social-economic, ecological and institutional data collected through surveys and remote sensing (Coleman et al., 2021a, 2021b).

The data set used in this analysis comprises 377 tree plantations, with the size equal to or greater than 5 hectares, planted after 1980 and for which there is no missing data. These plantations are located in 60 panchayats, which were randomly selected from all the panchayats in Kangra. In each panchayat, we interviewed randomly selected 40 households to obtain household-level data about the use and management of plantations, which they access for their livelihood needs as well as broader socioeconomic and political measures. We also interviewed key informants in each panchayat to obtain information about broader panchayat level variables.

In the 60 panchayats we mapped a total of 430 plantations and collected data for all of these plantations(Coleman et al. 2021a). But we removed observations with multiple plantations on same site and those with missing values for our set of variables for this analysis. Out of 430, 40 plantations are multiple year plantations, i.e. planting occurred in the same site on multiple years. After only using the latest year plantation in our study sample, we are left with 390 plantations.

We also removed 13 plantations with missing values for our set of variables, resulting in our study sample of 377 plantations.

The study region has seen massive tree-planting programs to meet the subsistence needs of local agrarian and livestock-based communities while targeting improved forest cover (Rana and Miller 2021). Such high dependency of local communities may reflect the success of these programs and may justify reforestation investments. Substituting requirements for fodder, fuelwood and timber from planted forests might save time, effort and money for local rural communities, as they do not need to buy these products from the market or go far to harvest fuelwood and fodder. Moreover, the people of the study landscape are traditionally dependent on small-scale agriculture and livestock husbandry. Forests play a major role in the livelihood profile of these rural smallholder farmers, and help meet their demands for timber, fodder, grazing, and wild fruits. Recent years have seen a growing proportion of households whose primary sources of income are government service or remittances, yet agriculture and natural resources remain important for most households. In this context, higher dependence of rural people on forest plantations can depict high usefulness of these plantations in terms of contributions to basic livelihood needs by providing resources that would otherwise need to be purchased on the market or simply unavailable to poorer households.

Variables

Forest cover

For each plantation, we calculate an increase in tree cover from the year 2001 (Forest Survey of India 2003) to the year 2017 (Forest Survey of India, 2019) based on national satellite data as an indicator of improvement in tree cover. We measure total forest change for a period between 2001 and 2017 in a particular plantation by calculating a change in the area of three categories of forests as measured by Forest Survey of India - open forests (10-40% forest density), moderately dense forests (40-70%) and dense forests (>70% forest density). If a plantation area gains forests during this period, that plantation is categorized as one with an increase in forest area during this period (dummy =1) and no change/remains the same otherwise (dummy = 0).

Forest Dependence index

Following Chhatre and Agrawal (2009), we measure the extent of rural livelihood benefits through a forest dependence index - which is calculated based on the proportion of fodder, grazing, timber and wild food needs of the local community met through tree plantations (Table A1.4). First, we collected information on the dependence of each studied household for four sources through socio-economic survey. The dependence is measured through a proportion of the total household consumption which is derived from each of the four sources. For each plantation, we then calculate average proportion of household dependence from each of the four sources (averaging all studied households). After that, we extracted the forest dependence index through a factor analysis of the proportions of all four sources for each of the 377 plantations. The factor loadings indicate the relative livelihood benefits in terms of forest dependence rather than the absolute level of benefits from plantations to studied households. Overall, the average proportions lie between 4 to 14%, indicating the modest livelihood support from plantations to households. However, there is considerable variation among households in terms of livelihood benefits derived from plantations.

We use oblique rotation as a transformational system in factor analysis as we expect our latent factors to be correlated. We used Cronbach's alpha to test the reliability and internal consistency of the forest dependence index (average inter-item covariance; scale reliability coefficient: 0.85 with confidence interval: 0.83 and 0.87; a coefficient >0.8 is considered acceptable). We categorized plantations with positive (or zero, or negative) values of forest dependence index as those with high (or low) livelihood benefits (for more details, please refer to supplementary materials). There is only a weak correlation between improvement in tree cover and the forest dependence index, suggesting likely tradeoffs between forest cover improvement and local livelihood needs in many contexts.

The higher values of forest dependence index depict high forest dependence of local communities on plantations. Hence, even when there is a very small proportion of use in our four resource usages, the factor analysis gives small negative value for our livelihood index. We consider these plantations with small negative values (or zeros) as plantations that yield low subsistence use to people or where people have low forest dependence. Although we categorized these forests as "lose" outcomes (or negative livelihood index), care should be taken in interpretation since this variable does not allow us to distinguish whether a low level of forest use is caused by less useful forest resources or alternative livelihood options in the context of agricultural-livestock economy in our study area.

Though, our index yields zero or negative values of index, it does not mean that plantation users lose their existing livelihoods, but only reflect very low dependence of those users on any of the four measured resource usages (wild foods or fish, fodder, grazing and timber).

To examine the effect of these 36 variables on win-win and win-lose outcomes, We use a 4-part classification to categorize the joint distribution of the outcome variables: a) win-win outcomes where there are high livelihood benefits and the forest cover improves, b) livelihood win outcomes where there are high livelihood benefits, c) forest cover win outcomes where there are forest cover gains or forest cover remains the same, and d) lose-lose outcomes where there are low livelihood benefits and the forest cover declines or remains the same. All these outcomes are dichotomous indicators (Tables 1.2-1.3 in supplementary material).

Improvement in tree cover is a dichotomous indicator that records whether a plantation achieves an increase in tree cover between 2001 and 2017 using tree canopy density estimates by Forest Survey of India (Forest Survey of India 2019). Livelihood benefit calculated using forest dependence index is also a dichotomous indicator that indicates whether there are high livelihood benefits from plantation (high forest dependence = 1) or low livelihood benefits (low forest dependence = 0). Plantations where forest cover shows an increase and where there are high livelihood benefits are labeled as win-win outcome plantations (a dichotomous indicator). The win-lose outcome is a dichotomous indicator indicating whether a plantation showed high livelihood benefits with same or reduced forest cover (livelihoods win - forest cover loses) or showed improvement (or stays same) in forest cover with low livelihood benefits (livelihoods lose - forest cover wins) (Tables 1.2-1.3 in supplementary material).

Despite our large set of variables, it is important to note that this may not fully exhaust the full suite of theoretically-relevant social-ecological variables associated with multiple plantation outcomes (Ostrom 2009). For example, though social variables such as leadership and social capital may be important in driving win-win outcomes, we could not include them due to lack of data. Moreover, this study restricts itself to exploring socio-economic variables driving multiple plantation outcomes and does not take into consideration suite of possible biophysical variables into analysis.

Our forest dependence index is a static snapshot of livelihoods (2018-2019). On the other hand, our ecological outcome is calculated based on measuring change in tree cover from 2001 to 2017. This means our win-win outcome reflects positive improvement in tree cover over the years, but high livelihood dependence as reflected on the level of the forest dependence index at one time. Therefore, care should be taken to analyze our results as we expect people to have variable levels of forest dependence on planted enclosures with time. Initially, plantations may yield higher level of grass or even grazing for households due to protection of planted enclosures through fencing.

However, there is a considerable time lag to obtain other livelihood benefits such as fodder, wild fruits or timber and this time lag varies with the type of forests, species planted or naturally regenerated as well as levels of plantation monitoring. For example, timber may take several decades to mature, but tree species such as Robinia (*Robinia pseudoacacia*) or Leucaena (*Leucaena leucocephala*) take only 5-8 years (or even less in highly productive lands) to yield fodder. In addition in each planted enclosure there is pre-existing or post-plantation growth of naturally regenerated seedlings or already existing dense tree growth, which may also determine flow of livelihood benefits to local communities depending upon the location, time and levels of local forest dependence(Coleman et al., 2021; Rana and Miller, 2021). Hence, lack of data on livelihood changes over time is a limitation of our analysis.

Methods

We first use a gradient boosting model to identify key variables that have a higher relative influence in terms of a substantial contribution in predicting joint tree cover and livelihood benefits(Ridgeway 2010; Greenwell et al. 2019) and then, use interpretable machine learning (IML) to find variable effects and interactions between the variables. We use partial dependence plots (PDPs) (Greenwell 2017) and Friedman's H Statistic (Molnar et al. 2018) to explore these variable effects and interactions, respectively.

We have used standard machine learning evaluation methods for controlling for bias or overfitting. These include tuning to balance fit with ROC (Receiver Operating Characteristics) and predictive accuracy to avoid overfitting to determine the optimal set of hyperparameters (number of trees, shrinkage, interaction depth and minimum number of observations in a terminal node with a fixed bag fraction of 0.7) and using tenfold stratified cross-validated accuracy (compares model predictions against withheld portions of data) to assess model performance in each of our 4 plantation models.

i) We first created the train data (all variables with binary outcome as factor) and model data (all variables with binary outcomes as numeric)

ii) We then tune the model using grid functionality in *Caret* R package, for which we need to provide initial values for four parameters: a) number of trees, b) shrinkage, c) number of minimum observations in a node, and d) interaction depth based on guidance from the literature (Epstein et al. 2021). Interaction depth ranging between 1 and 5 seems reasonable for most datasets. We used smaller shrinkage values in a sequence of values (0.001, 0.005, 0.01 and 0.05) to achieve higher predictive value as smaller the shrinkage value, the better the predictive value. We used number of minimum observations in a node in the sequence of 3, 5 and 10 as our data is small. Moreover, we used sequence of trees in the sets of 100, 2000 and 100 mainly because of the small shrinkage values used. Smaller the shrinkage values, higher is the number of trees required and that may entail higher computational cost.

iii) We then train GBM model based on tuning parameters and specific parameters for training control. We used two metrics to evaluate performance of the models:

a. The controls were resampled through the method ‘stratified cross-validation’, in a sequence of X folds. The metric to evaluate performance was kept as ‘ROC’ as it is more reliable in case of unbalanced datasets. We use ROC to select the optimal model using the largest value as ROC is well-suited to imbalanced datasets. The resampling (stratified cross-validation) determined parameters values for the model (shrinkage, interaction depth, number of min observations in a node, number of trees) while maximizing ROC values, which then are used in GBM models to identify key variables and their relationships and interactions.

b. The controls were resampled through the method ‘stratified cross-validation’, with 10 folds. The metric to evaluate performance was kept as ‘ROC’ as the main objective of the model is to select the optimal model using the largest value. The models run for 1-2 hours and for each iteration, strived for improving accuracy. The resampling (stratified cross-validated - with 10 folds) determined parameters values for the model (shrinkage, interaction depth, number of min observations in a node, number of trees) while optimizing for maximizing ROC values. Accuracy values were also mentioned for each plantation outcome models.

iv) Based on the model training, we got final values for our parameters to be used in the GBM model. For example for win-win outcomes, the final values obtained were n.trees = 1000, interaction.depth = 1, shrinkage = 0.05 and n.minobsinnode = 3 and the accompanied maximum ROC value as 0.95% with an accuracy of 89%. We obtained high values for ROC (X to Y) and predictive accuracies (X to Y) for our models, which further increase confidence in our modeling routines. These accuracy values are much higher than other similar modeling efforts (Epstein et al. 2021). We show below the plots showing model training for optimizing all of our four parameters to be used in the GBM modeling:

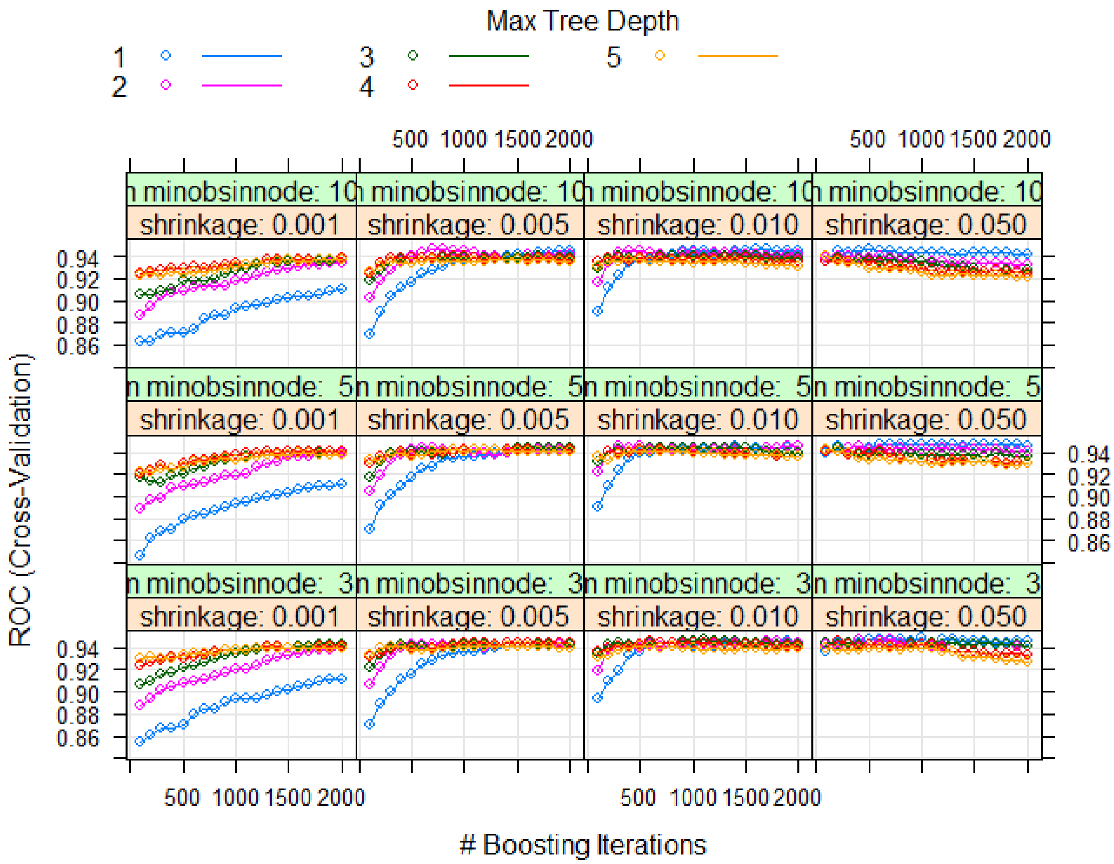


Figure A1.1. Model tuning to determine optimized set of parameters with the objective of improving stratified cross-validated ROC (Receiver Operating Characteristic Curve) for win-win outcomes. Model tuning suggested the following optimal combination of parameters to be used in boosting regression tree: number of trees = 1000, interaction depth = 1, shrinkage = 0.05, minimum number of observations in node = 3 with fixed bag fraction of 0.7. Our model achieved a higher value for ROC of 0.95 and accuracy of 89%. The model performance is assessed using tenfold stratified cross-validated accuracy.

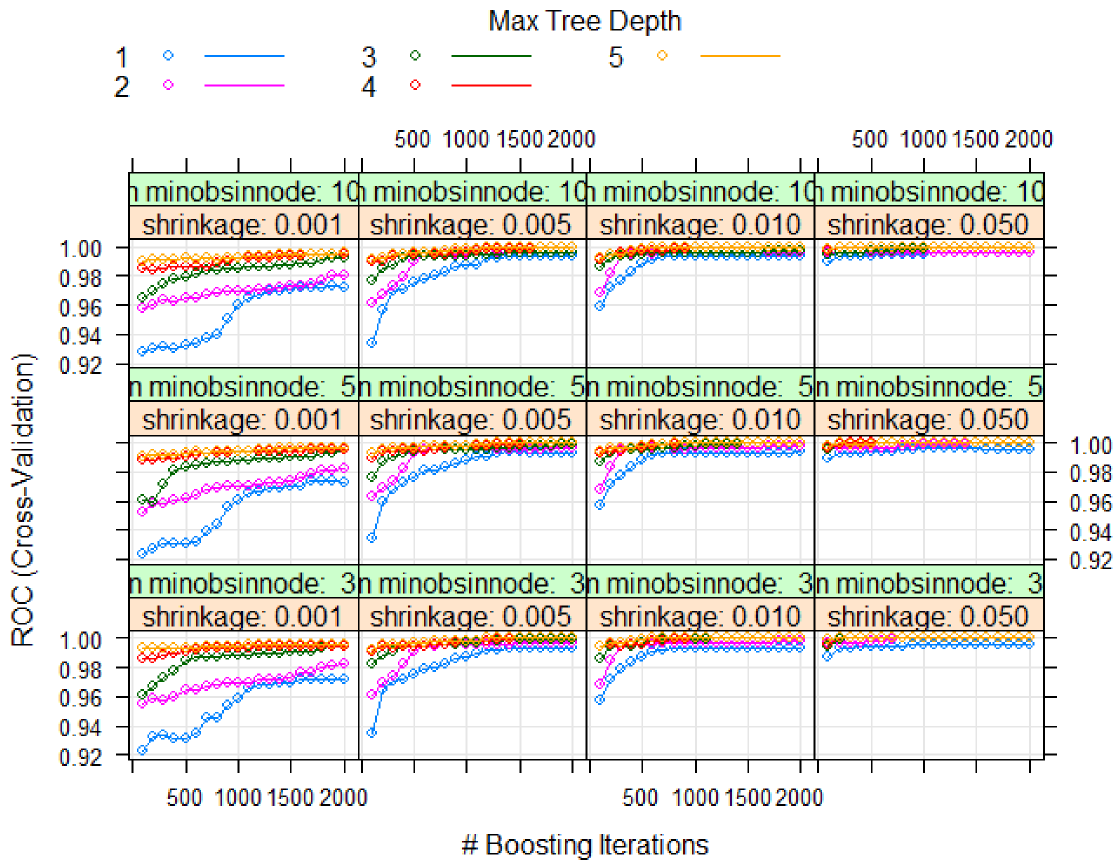


Figure A1.2. Model tuning to determine optimized set of parameters with the objective of improving stratified cross-validated ROC (Receiver Operating Characteristic Curve) for livelihood win outcomes. Model tuning suggested the following optimal combination of parameters to be used in boosting regression tree: number of trees = 200, interaction depth = 4, shrinkage = 0.05, minimum number of observations in node = 3 with fixed bag fraction of 0.7. Our model achieved a higher value for ROC of 1 and accuracy of 99%. The model performance is assessed using tenfold stratified cross-validated accuracy.

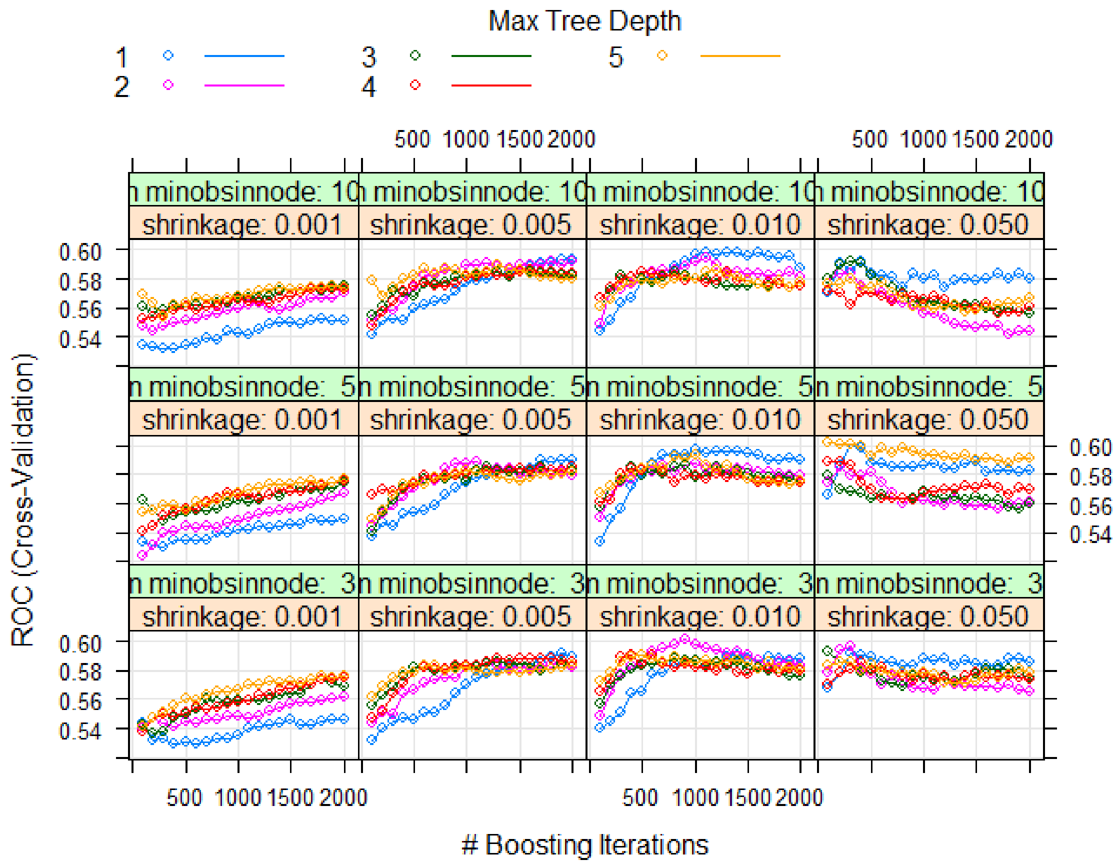


Figure A1.3. Model tuning to determine optimized set of parameters with the objective of improving stratified cross-validated ROC (Receiver Operating Characteristic Curve) for forest cover win outcomes. Model tuning suggested the following optimal combination of parameters to be used in boosting regression tree: number of trees = 100, interaction depth = 5, shrinkage = 0.05, minimum number of observations in node = 5 with fixed bag fraction of 0.7. Our model achieved a higher value for ROC of 0.60 and accuracy of 74%. The model performance is assessed using tenfold stratified cross-validated accuracy.

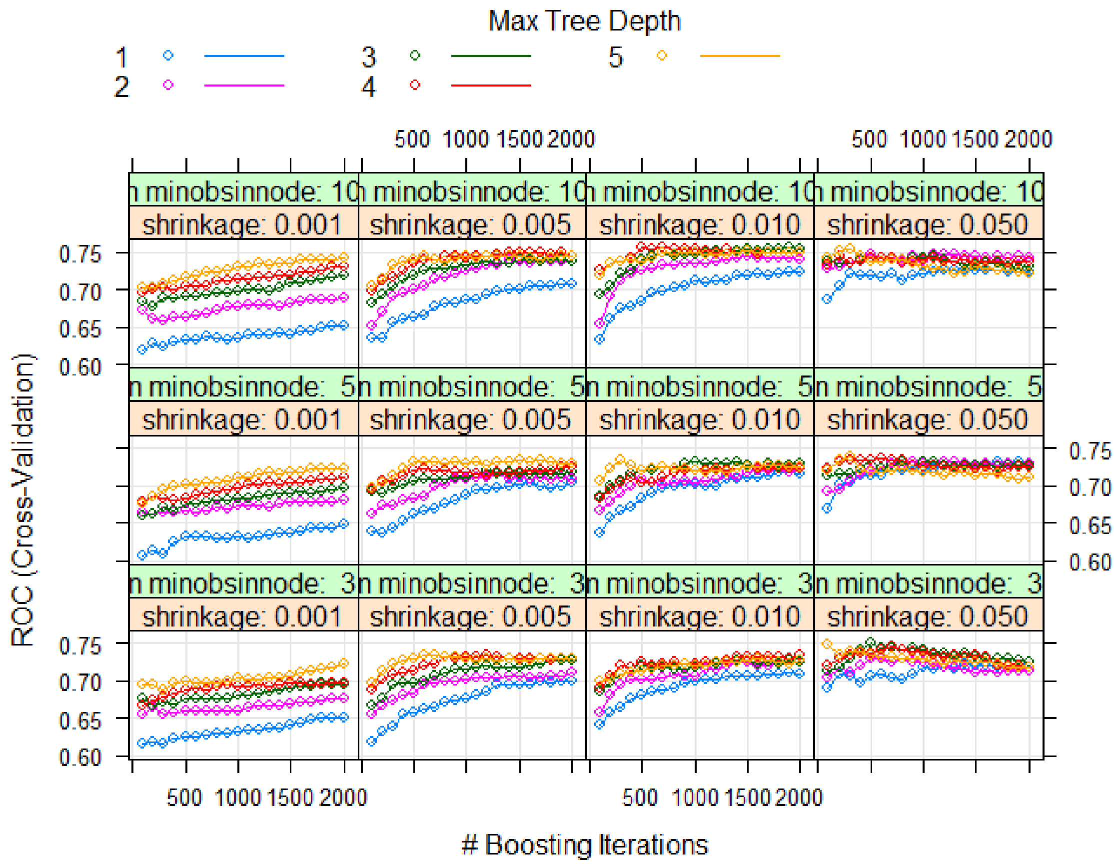


Figure A1.4. Model tuning to determine optimized set of parameters with the objective of improving stratified cross-validated ROC (Receiver Operating Characteristic Curve) for lose-lose outcomes. Model tuning suggested the following optimal combination of parameters to be used in boosting regression tree: number of trees = 500, interaction depth = 4, shrinkage = 0.01, minimum number of observations in node = 10 with fixed bag fraction of 0.7. Our model achieved a higher value for ROC of 0.75 and accuracy of 84%. The model performance is assessed using tenfold stratified cross-validated accuracy.

v) We then ran the GBM model (a gradient boosted model) with Bernoulli loss function, For example for win-win outcomes, we ran this model with 1000 iterations, 36 variables, interaction depth 1, number of min observations in a node as 3, shrinkage as 0.05. Bag fraction (subsampling fraction; fraction of training observations randomly selected for the next tree in the modeling), was fixed at 0.7 due to small sample size. From the model, we then calculated variable importance for the variables, calculated partial dependence plots for our top variables in terms of predictive importance, and also, then calculated interactions (H statistic) based on the selected model. Also, created two-way partial dependence plots (PDPs).

As evident from above (i-v), the standardized procedure was adopted to evaluate the model especially in terms of selecting the optimal model using the largest value of ROC and maximizing accuracy while controlling for the bias/overfitting. The hyperparameters used in the model were not selected randomly or from a set of default parameters, instead they were selected as a result of a systematic model tuning procedure involving 100-1000 combinations of different hyperparameters, and selecting a set of hyperparameters that had the highest 10-fold stratified cross validated ROC.

In gradient boosting, even if the model fails to predict accurately the outcome class for the first time, it gives more weightage to misclassified observations in the next iterations, thereby increasing its ability to predict the class with low cases. The gradient boosting relies on minimizing loss function of the model by adding weak learners using gradient descent. These weak learners are iteratively added in areas where strong/existing learners perform poorly, and the contribution of each of these weak learners to final prediction is decided based on a gradient optimization process. This leads to improved accuracy of the model through minimization of the overall error of the strong learner.

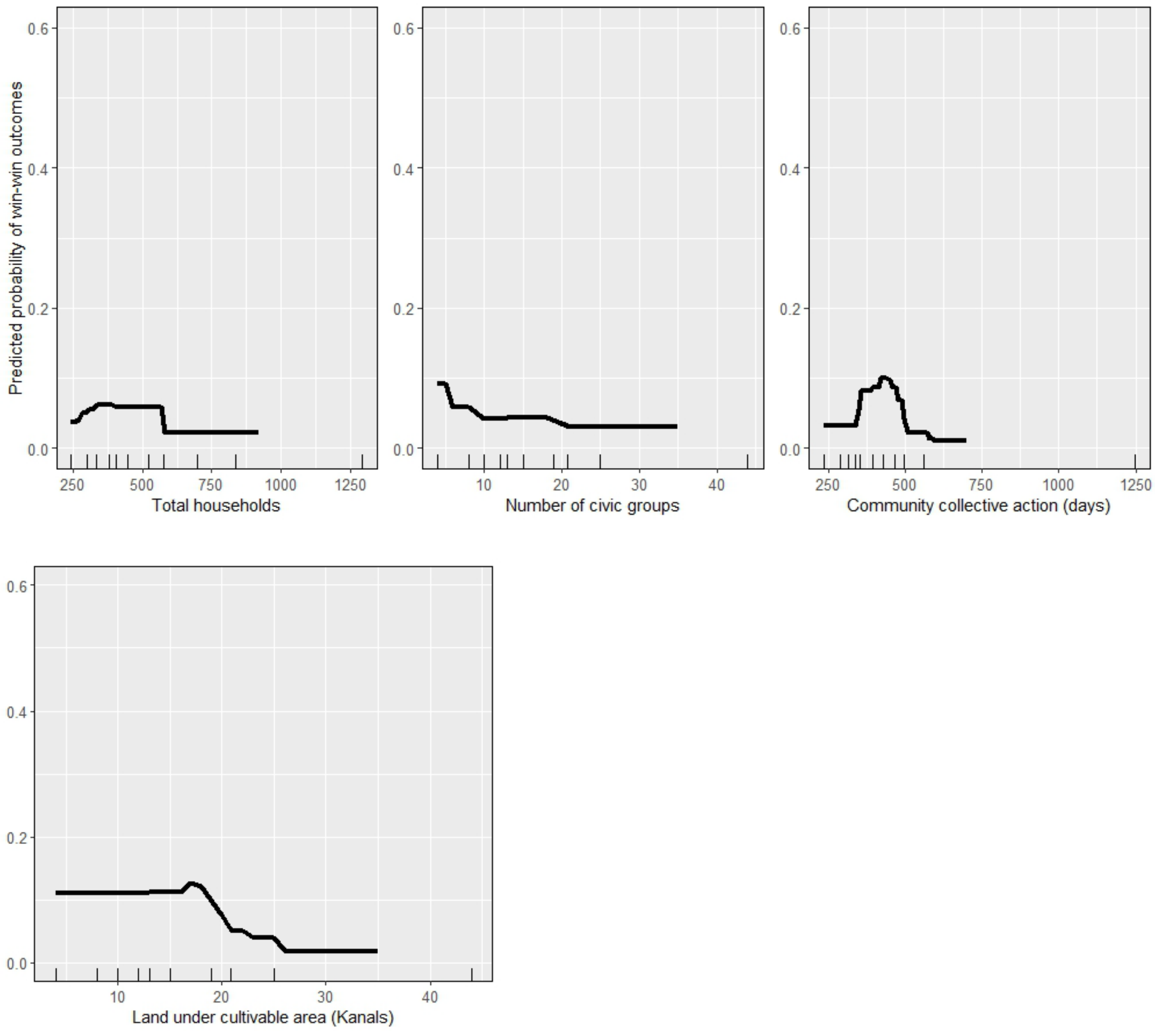


Figure A1.5. Partial dependence plots for win-win (livelihoods win, forest cover wins) outcomes for total households, number of civic groups, community collective action days and land under cultivable area.

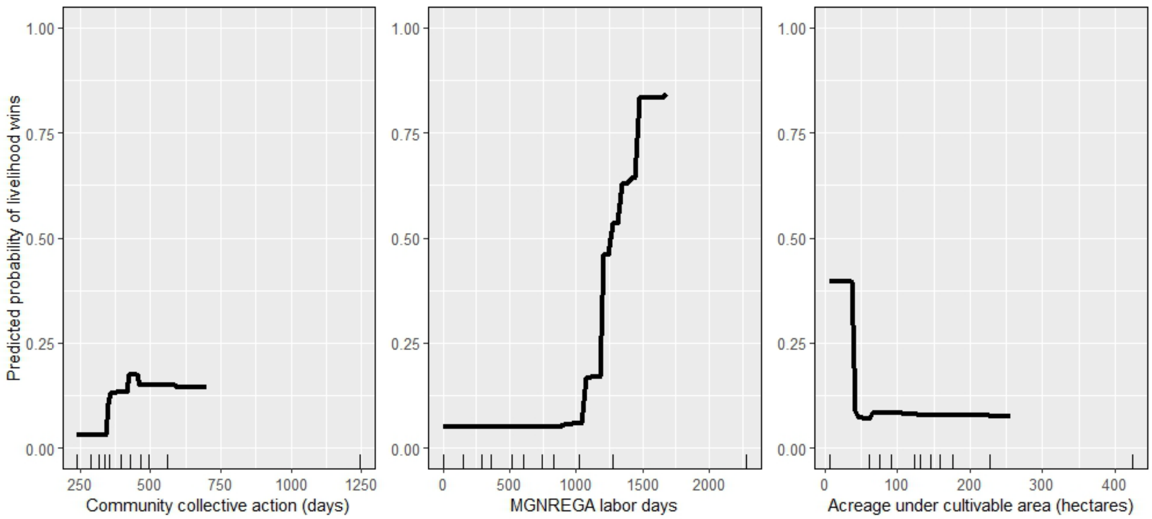


Figure A1.6. Partial dependence plots for livelihood win outcomes for Community collective action, MGNREGA labor days and acreage under cultivable area

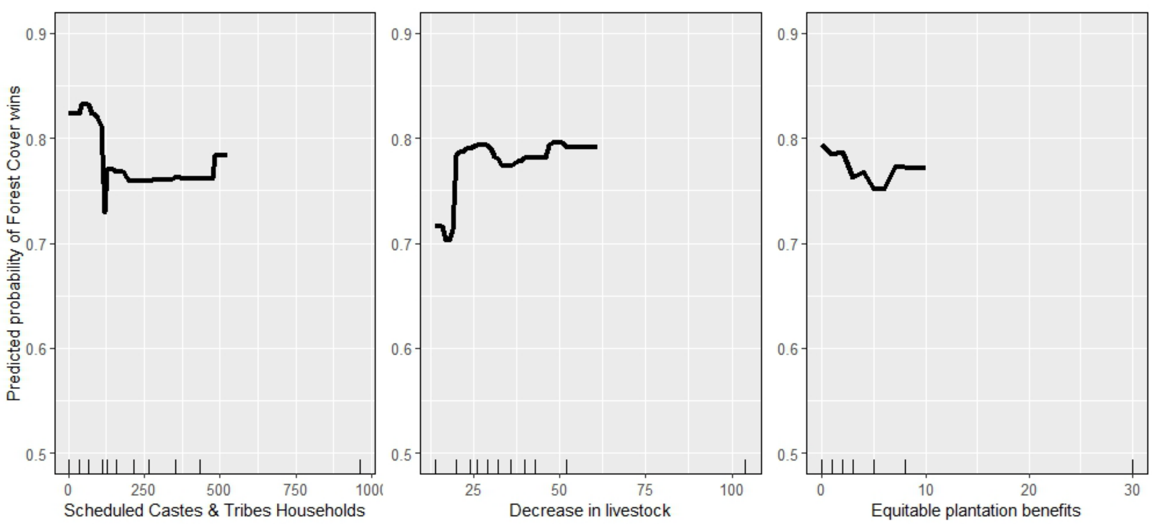


Figure A1.7. Partial dependence plots for forest cover win outcomes for Scheduled Caste and Scheduled Tribe households, decrease in livestock and equitable plantation benefits

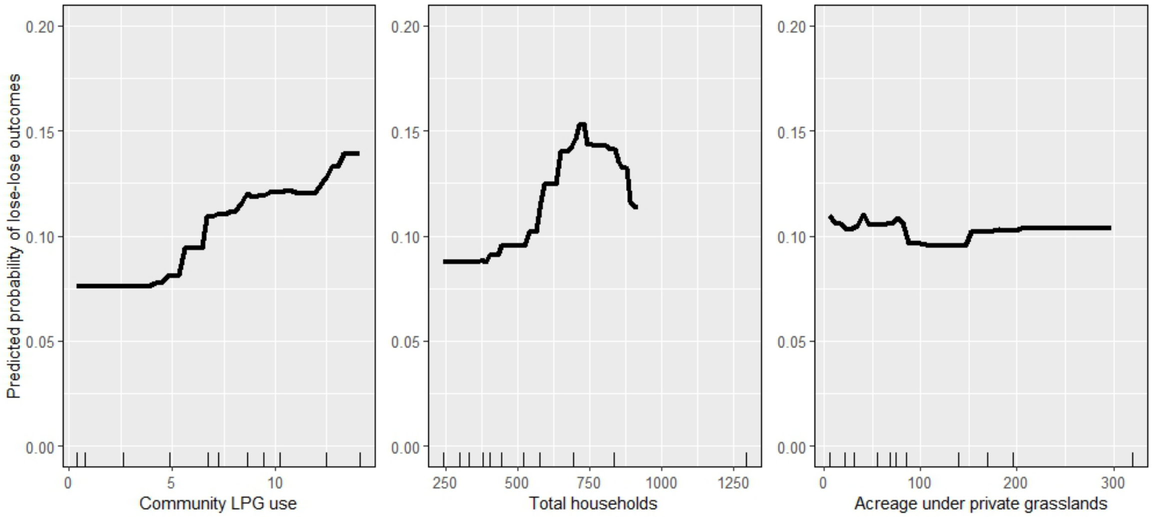
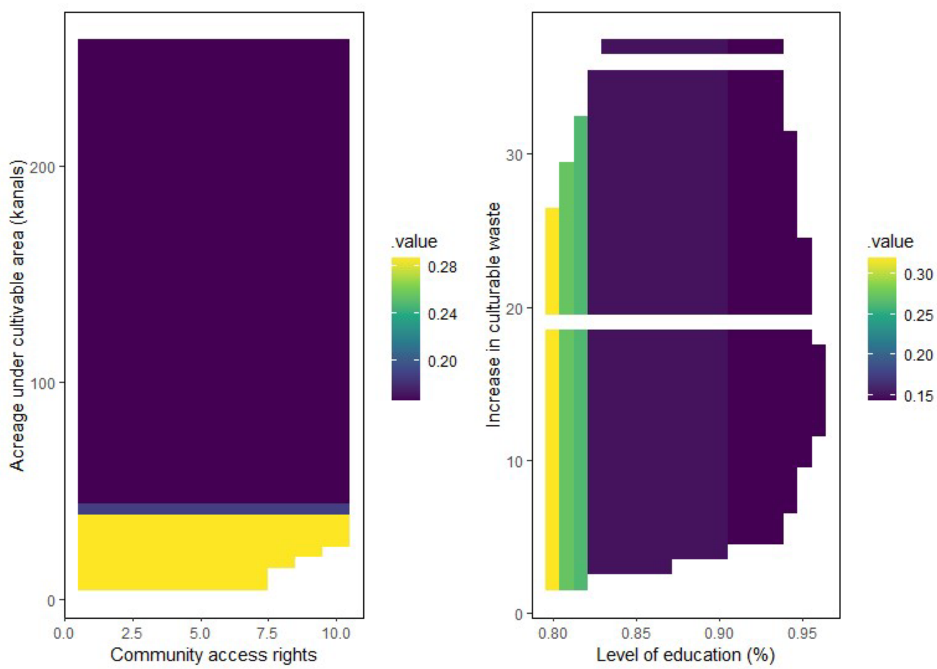


Figure A1.8. Partial dependence plots for lose-lose outcomes (livelihoods lose, forest cover lose) for community LPG use, total households and acreage under private grasslands.



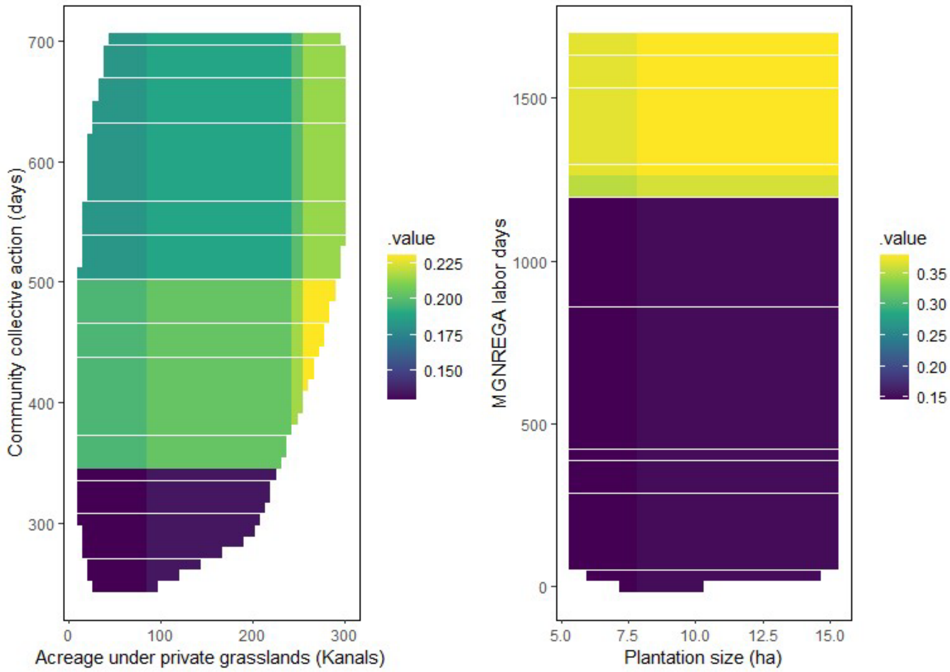
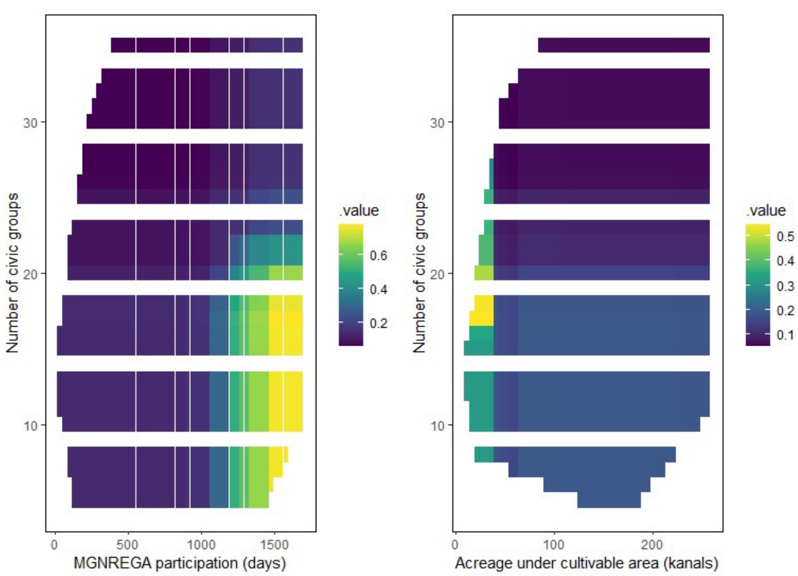


Figure A1.9. Bivariate plots for win-win (livelihoods wins, forest cover wins) outcomes



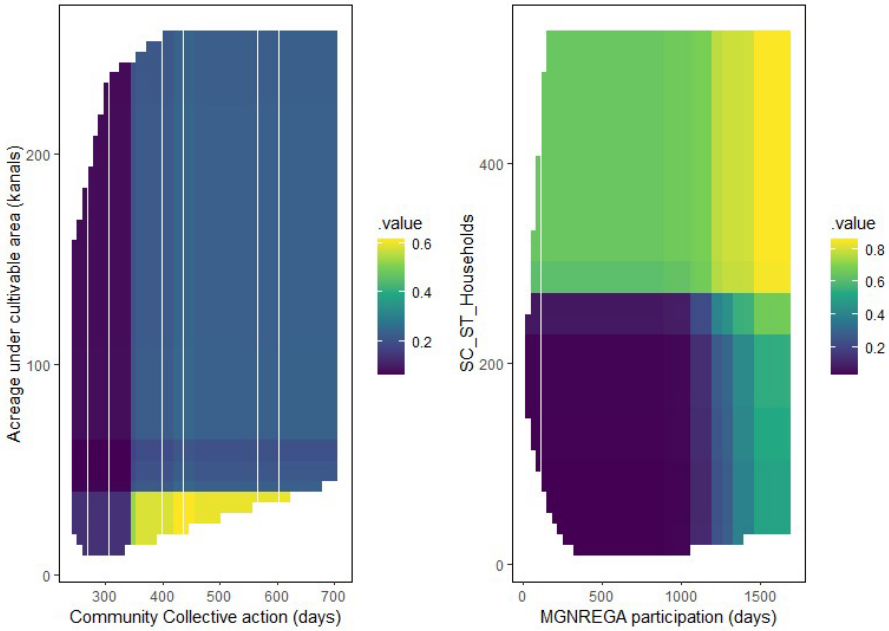
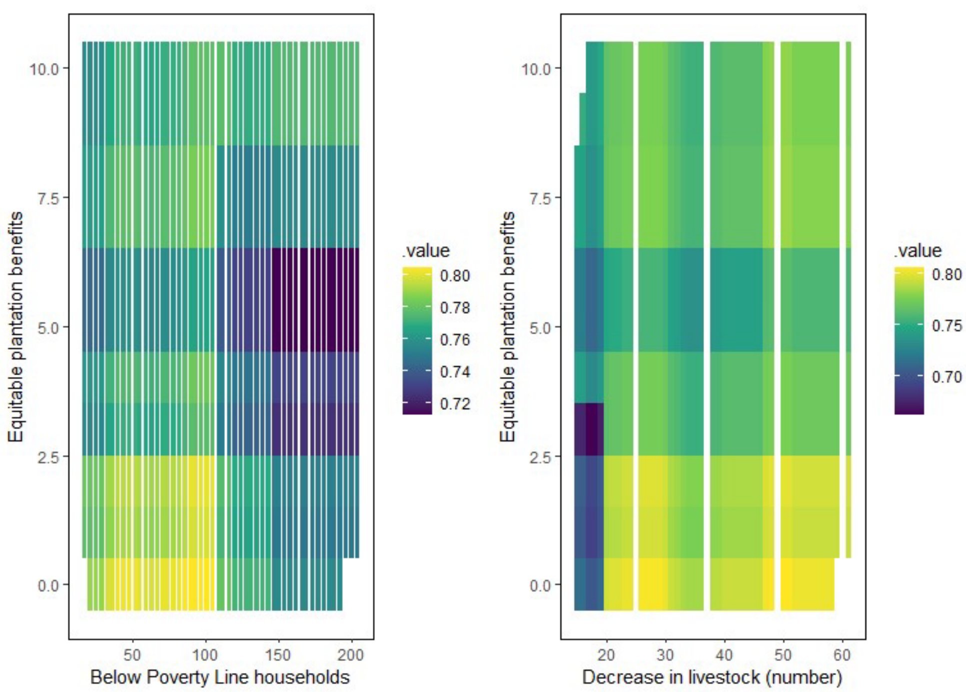


Figure A1.10. Bivariate plots for livelihood win outcomes



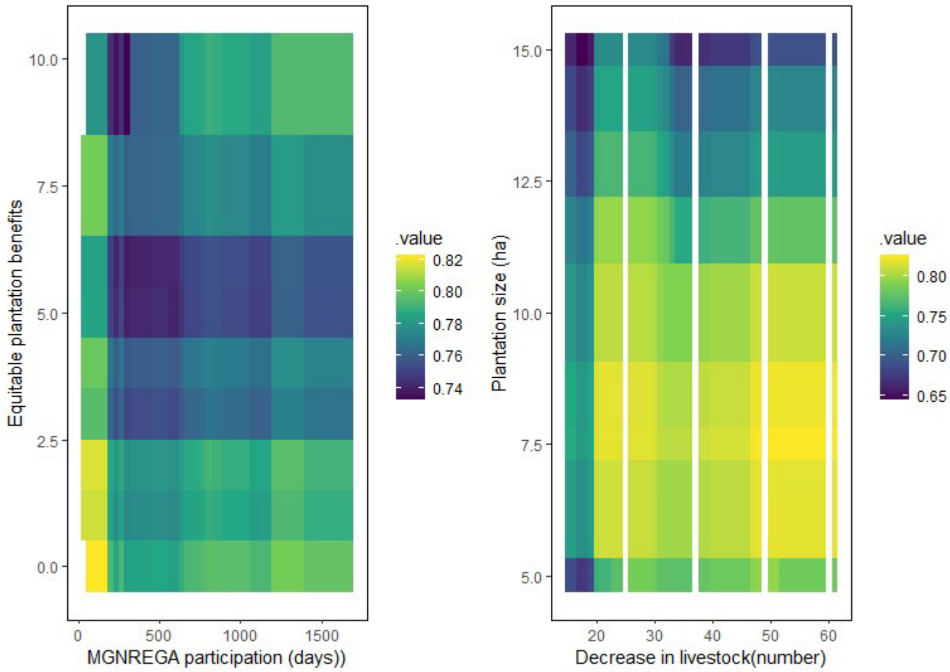


Figure A1.11. Bivariate plots for forest cover win outcomes

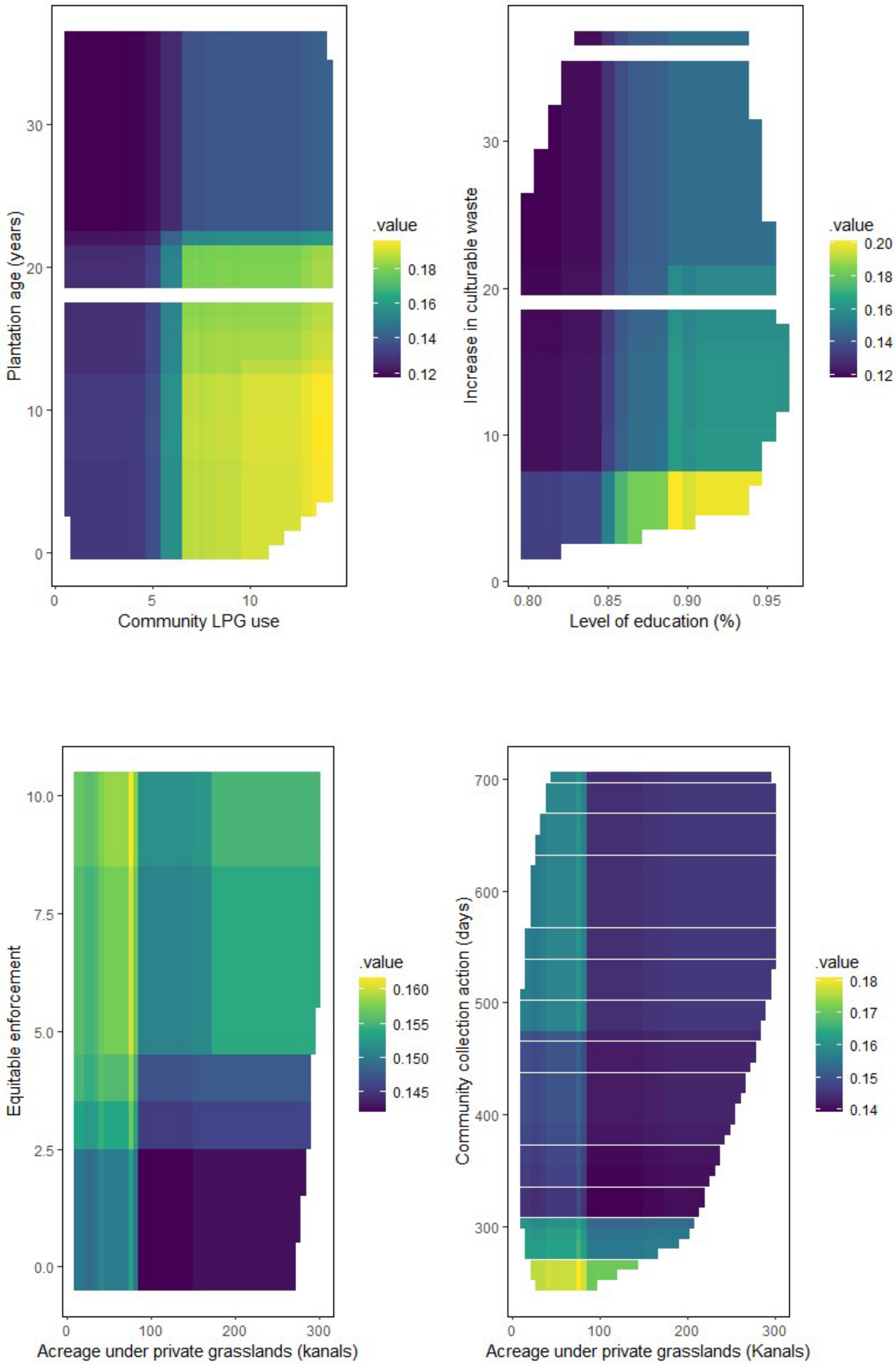


Figure A1.12. Bivariate plots for lose-lose outcomes

Table A1.1. Variables identified as critical in influencing win-win, win-lose and lose-lose outcomes

Social-economic and demographic characteristics	Institutional dynamics of forest governance and plantation activity	Biophysical characteristics
Total households	***Community-led***	Subsistence value
Below Poverty Line households	Community collective action	Plantation age
Scheduled Caste and Tribe households	Number of civic groups/institutions	Plantation size
Level of education	Community resource use decisions	
Community LPG use	Plantation-making participation	
MGNREGA* labor days	Political participation	
Remoteness of community	Site selection – co-management	
	Species selection – co-management	
Resource endowments	Appropriate protection measures	
Decrease in livestock	***State-led***	
Increase in culturable waste	Site selection - Forest Department	
Acreage under common haylands	Species selection - Forest Department	
Acreage under private grasslands	Plantation - official supervision	
Acreage under cultivable area	***Land tenure***	
	Prior forestland	
	Prior grassland	
	Plantation - Protected Forest Tenure	
	Plantation - Unclassed Forest	
	Plantation - Cooperative Forest Society	
	Plantation - Open access regime	

	Plantation access and rule enforcement
	Plantation access rights
	Plantation benefits
	Equitable plantation benefits
	Equitable enforcement

*MGNREGA: Mahatma Gandhi National Rural Employment Guarantee Act. Please refer to Table A1.1 for description and Table A1.3 for descriptive statistics.

Table A1.2. Key SES variables likely to influence long-term plantation outcome trajectories and their description

Variables	Indicator	Description	Unit of measurement	Sources
Socio-economic factors				
Economic backwardness				
	Scheduled Caste and Tribe households	Total SC and ST households in the panchayat	Number	Survey questionnaires
	Below Poverty Line households	Total Below Poverty Line households in the panchayat	Number	Survey questionnaires
	Remoteness of community	Distance of community from the Tehsil (town)	Minutes	Survey questionnaires
	Community LPG use	Supply of energy for household use in cooking and heating from LPG	Proportion	Survey questionnaires
Demographics				
	Total households	Total households in the panchayat	Number	Survey questionnaires
Livelihood diversity				

	Level of education	Total literacy in the panchayat	Proportion	Survey questionnaires
	MNREGA labor days	Number of MNREGA labor days of households last year in the panchayat	Number	Survey questionnaires
Resource endowments				
	Acreage under common haylands	Acreage under common haylands in the panchayat	Hectares	Survey questionnaires
	Acreage under private grasslands	Prior land cover of the planted areas is grasslands	Dummy; Yes =1; No = 0	Survey questionnaires
	Acreage under cultivable area	Total acreage under cultivation in the panchayat	Kanals	Survey questionnaires
Changes in socio-economic profile				
	Decrease in livestock	Decrease in livestock in last 10 years	Number	Survey questionnaires
	Increase in culturable waste	Increase in culturable waste in the panchayat	Number of responses; Dummy; Yes =1; No = 0	Survey questionnaires

Institutional dynamics of forest governance and plantation activity				
Community-led				
	Community collective action	Total person days shared as labor last year for collective action (forest, agriculture, construction, cultural aspects) in the panchayat	Number	Survey questionnaires
	Number of civic groups	Number of civic groups working in the panchayat	Number	Survey questionnaires
	Community resource use decisions	Communities make their own rules (formal and informal) regarding resource use in the panchayat	Dummy; Yes =1; No = 0	Survey questionnaires
	Plantation-making participation	Participation in making the plantations	Number of responses; Dummy; Yes =1; No = 0	Survey questionnaires

	Political participation	Number of times a political representative visited the plantation	Number	Survey questionnaires
	Plantation site selection - comanagement	Involvement of comanagement institutions in selecting plantation species	Dummy; Yes =1; No = 0	Survey questionnaires
	Plantation species selection - comanagement	Involvement of comanagement institutions in selecting plantation species	Dummy; Yes =1; No = 0	Survey questionnaires
	Appropriate protection measures	Conservation measures adopted with regard to plantations are appropriate (right level)	Dummy; Yes =1; No = 0	Survey questionnaires
State-led				
	Plantation site selection - Forest Department	Involvement of Forest Department in selecting plantation sites	Dummy; Yes =1; No = 0	Survey questionnaires
	Plantation species selection - Forest Department	Involvement of Forest Department in selecting plantation species	Dummy; Yes =1; No = 0	Survey questionnaires

	Plantation - official supervision	Number of times a higher forestry official visited the plantation	Number	Survey questionnaires
Land tenure				
	Prior forestland	Prior land cover of the planted areas is forestland	Dummy; Yes =1; No = 0	Survey questionnaires
	Prior grassland	Prior land cover of the planted areas is grasslands	Dummy; Yes =1; No = 0	Survey questionnaires
	Plantation - Demarcated Protected Forest	Official designation of plantation as demarcated protected forest	Dummy; Yes =1; No = 0	Survey questionnaires
	Plantation - Unclassed Forest	Official designation of plantation as unclassed forest	Dummy; Yes =1; No = 0	Survey questionnaires
	Plantation - Cooperative Forest Society	Official designation of plantation as cooperative forest society	Dummy; Yes =1; No = 0	Survey questionnaires
	Open access regime	Open access plantation regime	Dummy; Yes =1; No = 0	Survey questionnaires
Access and enforcement				

	Plantation access rights	Households having access rights in the plantations	Number of responses; Dummy; Yes =1; No = 0	Survey questionnaires
	Plantation benefits	Receipt of any plantation benefits by households in the panchayat	Number of responses; Dummy; Yes =1; No = 0	Survey questionnaires
	Equitable plantation benefits	Equity in distribution of benefits from plantations in the panchayat	Number of responses: Dummy; Yes = 1; No = 0	Survey questionnaires
	Equitable enforcement	Equity in the enforcement rules while accessing plantations in the panchayat	Number of responses; Dummy; Yes =1; No = 0	Survey questionnaires
Biophysical characteristics/ Plantation attributes				
	Plantation subsistence value	High subsistence value of the plantation (above normal and substantially above normal)	Dummy; Yes =1; No = 0	Survey questionnaires

	Plantation age	Age of the plantation	Years	Survey questionnaires
	Plantation size	Size of the plantation	Hectares	Survey questionnaires
Outcomes (O)				
	Tree cover improvement	24 m resolution FC_CHANGE17_01 = FC_2017HA – FC_2001HA	If FC_CHANGE17_01<= 0, Forest cover declines or remains the same = 1, OTHERWISE = 0	Forest Survey of India (2003); Forest Survey of India (2019)(Forest Survey of India 2019)
	Forest dependence Index	Proportion of wild foods, fodder, grazing and timber	Factor analysis	For details, refer to methods and Table A1.3
	Win-win outcomes	Joint production of forest cover and livelihood benefits (forest dependence index) from a plantation in the panchayat	win-win outcome = 1 When forest cover improves (Yes =1; No = 0) and above-average livelihood	Forest Survey of India (2003); Forest Survey of India (2019)(Forest Survey of India 2019) and survey questionnaires

			benefits (Yes = 1; No = 0)	
	Livelihood win outcomes		Positive values for livelihood index = 1 (0= otherwise)	Survey questionnaires
	Forest cover wins outcomes		Positive values for forest cover improvement = 1 (0= otherwise)	Forest Survey of India (2003); Forest Survey of India (2019)(Forest Survey of India 2019)

Table A1.3. Descriptive statistics of the key variables (n = 377)

Variables	Variables	mean	sd	median	min	max	range
Socio-economic and demographic characteristics							
	Scheduled Caste and Tribe households	33.51	17.44	33.00	2.00	67.00	65.00
	Below Poverty Line households	78.04	44.52	63.00	14.00	249.00	235.00
	Remoteness of community	63.46	37.93	60.00	5.00	160.00	155.00

	Community LPG use	7.05	3.98	7.25	0.38	14.12	13.75
	Total households	509.95	200.34	446.00	242.00	1294.00	1052.00
	Level of education	0.88	0.05	0.89	0.73	0.96	0.23
	MNREGA labor days	690.48	461.54	605.00	0.00	2278.00	2278.00
Changes in socio-economic profile							
	Decrease in livestock	34.43	13.60	32.00	14.00	104.00	90.00
	Increase in culturable waste	15.83	8.87	15.00	2.00	59.00	57.00
Resource endowments							
	Acreage under common haylands	13.35	42.66	0.00	0.00	269.00	269.00
	Acreage under private grasslands (in Kanals)	96.78	70.52	74.50	6.00	320.00	314.00
	Acreage under cultivable area	135.11	62.25	131.50	6.50	424.50	418.00
Institutional dynamics of forest governance and plantation activity							
Community-led							
	Community collective action	418.76	131.75	396.00	238.00	1247.00	1009.00

	Number of civic groups	15.01	6.98	13.00	4.00	44.00	40.00
	Community resource use decisions	0.06	0.23	0.00	0.00	1.00	1.00
	Plantation-making participation	0.29	0.74	0.00	0.00	6.00	6.00
	Political participation	0.58	1.13	0.00	0.00	10.00	10.00
	Plantation site selection - comanagement	0.39	0.49	0.00	0.00	1.00	1.00
	Plantation species selection - comanagement	0.21	0.41	0.00	0.00	1.00	1.00
	Appropriate protection measures	0.93	0.25	1.00	0.00	1.00	1.00
State-led							
	Plantation site selection - Forest Department	0.43	0.50	0.00	0.00	1.00	1.00
	Plantation species selection - Forest Department	0.67	0.47	1.00	0.00	1.00	1.00
	Plantation - official supervision	0.63	0.82	0.00	0.00	6.00	6.00

Land tenure							
	Prior forestland	0.64	0.48	1.00	0.00	1.00	1.00
	Prior grassland	0.46	0.50	0.00	0.00	1.00	1.00
	Plantation - Protected Forest Tenure	0.29	0.45	0.00	0.00	1.00	1.00
	Plantation - Unclassed Forest	0.37	0.48	0.00	0.00	1.00	1.00
	Plantation - Cooperative Forest Society	0.12	0.33	0.00	0.00	1.00	1.00
	Open access regime	0.21	0.41	0.00	0.00	1.00	1.00
Access and enforcement							
	Plantation access rights	2.89	4.40	1.00	0.00	30.00	30.00
	Plantation benefits	0.29	0.73	0.00	0.00	6.00	6.00
	Equitable plantation benefits	2.88	4.40	1.00	0.00	30.00	30.00
	Equitable enforcement	2.76	4.28	1.00	0.00	29.00	29.00
Biophysical characteristics/Plantation attributes							
	Plantation subsistence value	0.48	0.50	0.00	0.00	1.00	1.00
	Plantation age	19.50	11.17	20.00	0.00	38.00	38.00

	Plantation size	8.94	5.46	7.50	5.00	40.00	35.00
Outcomes (O)							
	Tree cover improvement change2017_2001Hectare	1.16	3.70	0.81	-50.57	16.24	66.82
	Forest dependence Index	0.01	0.99	-0.34	-0.96	3.03	3.99
	Win-win outcomes	0.24	0.43	0.00	0.00	1.00	1.00

Table A1.4. Forest Dependence index: factor analysis and factor loadings							
					Factor loadings (pattern matrix) and unique variances		
Factor	Eigen value	Difference	Proportion	Cumulative	Variable	Factor 1	Uniqueness
Factor 1	2.92	2.08	0.7318	0.7318	Wild foods	0.85	0.2839
Factor 2	0.84	0.68	0.2105	0.9424	Fodder	0.71	0.4972

Factor 3	0.16	0.09	0.0401	0.9825	Grazing	0.91	0.1721
Factor 4	0.07	_	0.0175	1.0000	Timber	0.73	0.4612

Factor Analysis to construct forest dependence index

Single-factor exploratory factor analysis (EFA) has been used to construct the forest dependence index using the minimum residual (minres) solution. The results of the EFA are: Number of parameters = 4, retained factor =1; rotation: "oblimin"; Bartlett test of sphericity: $chisq 1347.788 with p.value = 0.000; small p value suggests that we can conduct the factor analysis.

Correlation of (regression) scores with factors = 0.98; Multiple R square of scores with factors = 0.96; Minimum correlation of possible factor scores = 0.92.

We also used the Cronbach's alpha to test the reliability and internal consistency of the forest dependence index (average inter-item covariance; scale reliability coefficient: 0.85 with confidence interval: 0.83 and 0.87; a coefficient >0.8 is considered acceptable). Maximum split half reliability (lambda 4) = 0.97; average split half reliability = 0.88; CFI = 0.59; KMO = 0.58. We use dummy variable (increase = 1; low forest dependence = 0) as outcome variable as per the requirements of GBM modeling and to facilitate this analysis given our small dataset.

The factor loadings as given in Table A1.4 are all above 0.5 and therefore, are practically significant.

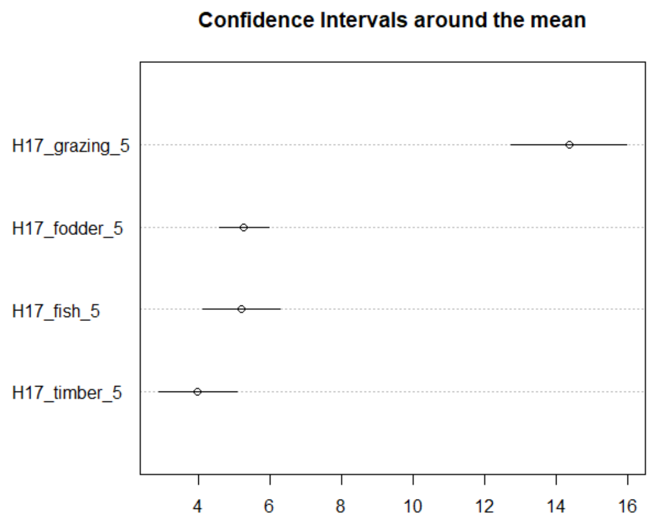


Figure A1.13. Factor Analysis: Confidence intervals around the mean for the proportions of the quantity of (i) wild foods and fish, (ii) fodder, (iii) grazing, and (iv) timber used for domestic consumption that each plantation provides to the local forest users.

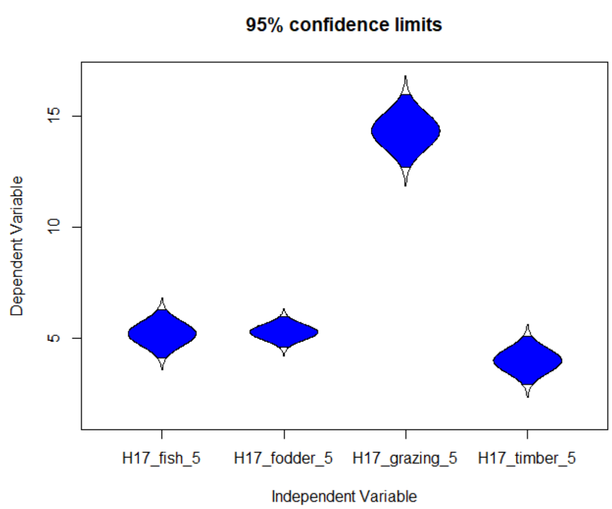


Figure A1.14. Factor Analysis: 95% Confidence intervals around the mean for the proportions of the quantity of (i) wild foods and fish, (ii) fodder, (iii) grazing, and (iv) timber used for domestic consumption that each plantation provides to the local forest users.

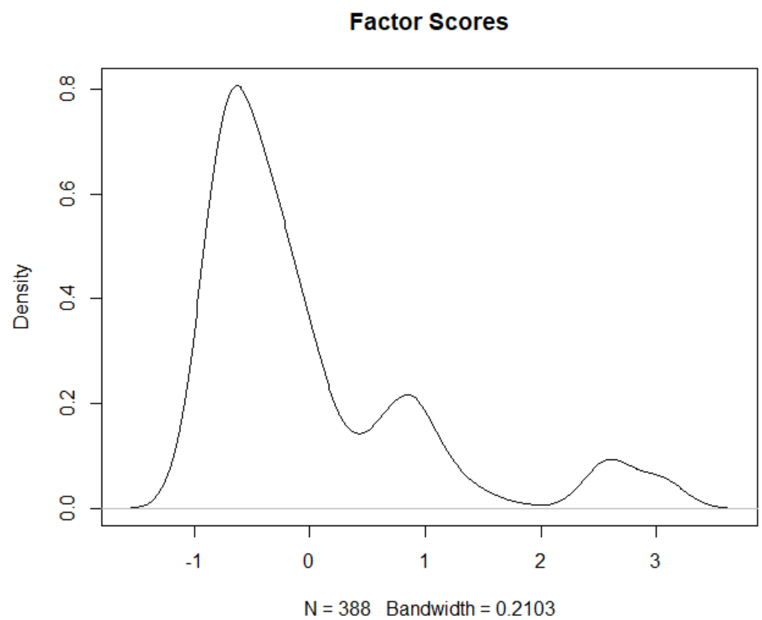


Figure A1.15. Factor Analysis: Density plots for the factor scores for a visual representation.

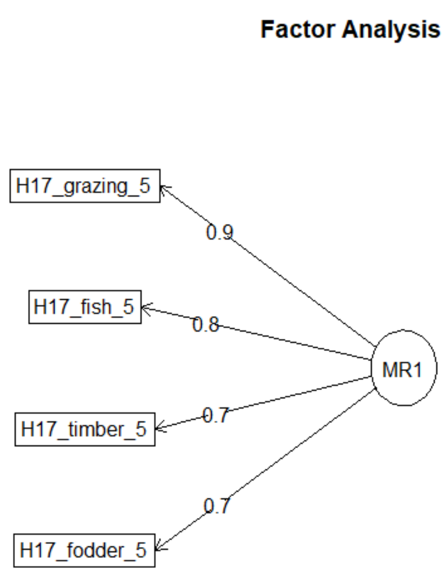


Figure A1.16. Factor Analysis: path diagram for the factor scores

Table A1.5. Testing correlations: lower triangle of the correlation matrix among variables for easy viewing and interpretation

	Wild foods and fish	Fodder	Grazing	Timber
Wild foods and fish	1.00			
Fodder	0.83	1.00		
Grazing	0.65	0.57	1.00	
Timber	0.54	0.35	0.88	1.00

Table A1.6. p-values from ttests of each correlation (p-values are corrected using the Holm Correction by default); significant values mean variables included are meaningfully correlated.

	Wild foods and fish	Fodder	Grazing	Timber
Wild foods and fish	0.00E+00	4.69E-101	1.93E-47	4.05E-30
Fodder	5.21E-102	0.00E+00	1.10E-33	9.29E-12
Grazing	2.42E-48	1.83E-34	0.00E+00	5.31E-125
Timber	8.10E-31	2.32E-12	5.31E-126	0.00E+00

References

Coleman E, Schultz B, Ramprasad V, et al (2021a) Data for Decades of tree planting in Northern India had little effect on forest density and rural livelihoods. Dataset

Coleman EA, Schultz B, Ramprasad V, et al (2021b) Limited effects of tree planting on forest canopy cover and rural livelihoods in Northern India. Nat Sustain 1–8. https://doi.org/10.1038/s41893-021-00761-z

Forest Survey of India (2019) The State of Forest Reports

Greenwell B, Boehmke B, Cunningham J, Developers GBM (2019) gbm: Generalized boosted regression models. R package version 2:

Greenwell BM (2017) pdp: An R Package for Constructing Partial Dependence Plots. R J 9:421

Molnar C, Casalicchio G, Bischl B (2018) iml: An R package for interpretable machine learning. Journal of Open Source Software 3:786

Ostrom E (2009) A general framework for analyzing sustainability of social-ecological systems. Science 325:419–422

Rana P, Miller DC (2021) Predicting the long-term social and ecological impacts of tree-planting programs: Evidence from northern India. World Development 140:105367

Ridgeway G (2010) gbm: generalized boosted regression models. R package version 1.6-3.1