

LEARNING FROM NATURE TO PREDICT FARMERS' ADOPTION: GENETIC ALGORITHMS IN AGRICULTURAL EXTENSION RESEARCH



In this blog, Dr. J. Paul Mansingh explores how Genetic Algorithms, inspired by natural evolution, can enhance the prediction of farmers' agricultural technology adoption. Using a hypothetical case study, he demonstrates how integrating Genetic Algorithms with Random Forest and Explainable AI can identify informative predictors and support more efficient agricultural extension research.

INTRODUCTION

Agricultural extension research has traditionally focused on understanding why farmers adopt or reject agricultural innovations. Statistical techniques such as correlation, regression, factor analysis and Structural Equation Modelling (SEM) have been widely employed to investigate the influence of knowledge, attitudes, innovativeness, risk orientation, social norms and institutional support on adoption behaviour.



More recently, machine-learning algorithms such as Random Forest (RF), Artificial Neural Networks (ANN) and XGBoost have expanded opportunities for predicting farmers' adoption behaviour. However, an important methodological question remains: Among numerous behavioural, socioeconomic and institutional variables, which combination provides the most useful prediction of agricultural technology adoption? Nature-inspired algorithms (Box 1) offer an interesting approach to addressing this question.

Box 1: What are Nature-Inspired Algorithms?

Nature-Inspired Algorithms (NIAs) are the computational optimization algorithms developed by mimicking the evolution of nature, animal behaviors and natural phenomena. For instance, Particle Swarm Optimization (PSO) simulates the movement of birds and fish, Ant Colony Optimization (ACO) mimics the behavior of ants when they are searching for the most efficient path of movement, and Artificial Bee Colony (ABC) algorithms mimic the food foraging behavior of honeybees.

Other algorithms are Grey Wolf Optimizer, Firefly Algorithm, Whale Optimization Algorithm, Genetic Algorithm etc. The algorithms systematically explore all the different solutions for finding combinations that are promising with respect to a set of objectives and constraints. They can be of great help when there are many potential solutions or when standard optimization techniques have problems.

Nature-inspired algorithms can help with the segmentation of farmers, the selection of predictors, allocation of extension resources, optimization of training programmes and planning for interventions in agricultural extension research. The approach of Genetic Algorithm (GA) is one of the well-established evolutionary optimization algorithms that has great potential for research on predictions of adoption.

GENETIC ALGORITHM: LEARNING FROM BIOLOGICAL EVOLUTION

Genetic Algorithm is an algorithm whose concepts were based on natural selection and natural evolution. It explores and looks for better solutions using three basic operations:

Selection: solutions that are promising are kept because of their performance.

Crossover: Information of selected solutions is mixed in order to create new solutions.

Mutation: Modification is randomly altered to find other possibilities.

This process is repeated from generation to generation until a satisfactory result is achieved or a stopping criterion is met. The distinguishing feature of GA is that it does not estimate the relationships between the explanatory variables and the adoption directly as it does with conventional statistical procedures. Rather, it looks for sets of variables or model parameters that yield the "best" performance for a specified criterion. Thereby, the application of GA to feature selection of the models of agricultural adoption is particularly attractive.

WHY IS FEATURE SELECTION IMPORTANT IN ADOPTION RESEARCH?

Let us take the case of a study on the uptake of Climate-Smart Agriculture (CSA) for farmers. Data can be gathered on many traits such as innovativeness, risk orientation, self-efficacy, social norms, extension contact, institutional support, digital literacy, attitude toward the environment, knowledge of CSA, farm size, level of education, age, access to market, and access to credit and training.

Not every variable necessarily is the most efficient way to go about a predictive model. Predictors could provide some helpful extra information, or some could have redundant information in them.

Feature Selection aims to find a set of variables that are useful and will not reduce the accuracy of the model, but instead will make it simpler. Conventional techniques include stepwise selection, regularization and recursive feature elimination. GA offers an alternative approach to evolution that considers combinations of predictors instead of considering the importance of the predictors alone.

RESEARCH EXAMPLE: PREDICTING CSA ADOPTION AMONG PADDY FARMERS

Imagine that a researcher is carrying out a hypothetical survey of 600 paddy farmers.

CSA adoption is measured as:

- 1 = Adopter
- 0 = Non-adopter

A total of 15 variables in the areas of behaviour, institutions and socioeconomic factors are assumed to be potential predictors.

The research question is:

Does it seem possible that a smaller subset of predictors can be found by using the GA to select them for which the accuracy of predicting the adoption of CSA is the same or better?

With 15 candidate predictors, there are:

$2^{15} = 32,768$ possible predictor subsets, including the empty subset.

If any of these combinations are also used with complex machine-learning models, or if a cross-validation approach is used, it may become computationally intensive to evaluate all of them. GA provides us with a heuristic means for exploring this space.

Step 1: Prepare the Dataset

The number of farmers in the hypothetical data is 600 of which 274 are farmers who adopted and 326 are the non-adopters. Data is split into training set and testing set. About 75% are used for model development and 25% are kept for independent evaluation. This separation is necessary because using the last test data to select predictors can lead to overly optimistic results.

Step 2: Develop a Benchmark Random Forest Model

Initially, a Random Forest classification model is developed using all 15 predictors to predict farmers' adoption of Climate-Smart Agriculture (CSA). The predictive performance is evaluated by accuracy, sensitivity, specificity and Area under the Receiver Operating Characteristic Curve (AUC-ROC). The full-predictor model will serve as a reference model to which the GA-assisted model can be compared.

Step 3: Apply Genetic Algorithm for Predictor Selection

In GA-based feature selection, each predictor is represented by a binary value:

1 = Include the predictor; 0 = Exclude the predictor.

Therefore, each chromosome with 15 binary positions represents a single set of predictors for the adoption. A Random Forest model for each combination of candidates is evaluated and cross validation is used. With a proper fitness function can be written as:

Fitness = Cross-validated ROC-AUC – Penalty for number of predictors

The first part is for the predictive accuracy, and the second part is to prevent the use of unnecessarily complicated models. GA repeatedly applies the selection, crossover and mutation to look for combinations of higher fitness values.

Step 4: Identify the Selected Predictors

In a small computational demonstration using the hypothetical dataset, GA selected nine predictors from the original fifteen:

1. Innovativeness
2. Self-efficacy
3. Social norms
4. Extension contact
5. Institutional support
6. Digital literacy
7. CSA knowledge
8. Credit access
9. Training exposure

This represents a **40% reduction in the number of predictors**.

Importantly, these results are only for the simulated data and computational parameters. These do not reflect the generally accepted empirical evidence that these nine variables are the most important determinants for CSA adoption. Variables not included in GA can be theoretically relevant and/or causally important.

Step 5: Compare Predictive Performance

The demonstration produced the following results:

Performance indicator	Full Random Forest	GA-assisted Random Forest
Predictors	15	9
Test ROC-AUC	0.741	0.741
Test accuracy	66.7%	66.0%

The same number of predictors were used in the GA-assisted model, and both models yielded similar ROC-AUC results. In this particular exercise, the reduced model thus achieved similar variable discrimination but with decreased complexity. The accuracy was slightly lower, however. The results do not mean that GA is superior to the traditional feature selection methods. A decision like this must be reiterated and compared to suitable standards.

WHAT ADDITIONAL INFORMATION DOES GA PROVIDE?

The main strength of GA is the tuning of combinations of predictors, not the identification of new constructs of behaviour. Traditional models of adoption are used to investigate the relationships among variables. The Random Forest model is used to predict outcomes of adoption. GA finds sets of predictors that offer good predictive accuracy while not being too complex.

For instance, if it is possible to identify a subset of informative variables, this can facilitate the creation of shorter survey instruments and could minimize the need for data collection and could enable the construction of more efficient predictive instruments. But, it must be noted that the predictors selected by GA do not automatically deserve to be called causal factors of adoption. A variable can be a good predictor without causing farmers to act in any particular way.

INTEGRATING GA WITH EXPLAINABLE ARTIFICIAL INTELLIGENCE

Once the predictors are selected, Explainable Artificial Intelligence (XAI) techniques can be used to interpret the fitted Random Forest model. Importance of variables can be determined by permutation

and they can contribute significantly to predictive power. SHapley Additive exPlanations (SHAP) can also analyze contribution of individual variables to predictions of specific farmers.

For example, a model may predict that a farmer has a 72% probability of adopting CSA. SHAP may indicate that high self-efficacy and frequent extension contact contribute positively to this prediction, while limited digital literacy contributes negatively. These explanations describe the behaviour of the predictive model, not causal effects.

The combination of GA, Random Forest and SHAP therefore offers a framework connecting predictor optimization, adoption prediction and model interpretation.

SOFTWARE FOR AGRICULTURAL EXTENSION RESEARCHERS

The entire analysis can be performed using **RStudio**, supported by freely available R packages.

R package	Application
GA	Genetic Algorithm and feature selection
ranger	Random Forest classification
caret	Data partitioning and model validation
pROC	ROC-AUC analysis
fastshap	SHAP-based model explanations
ggplot2	Visualization

Python also provides suitable tools for evolutionary optimization and machine learning.

For the training data, feature selection should be done and preferably be based on nested cross validation and checked for stability over multiple runs of this process. Performance should be compared to less complex models like regularized regression and recursive feature elimination.

IMPLICATIONS FOR AGRICULTURAL EXTENSION RESEARCH

GA-assisted feature selection offers several research opportunities.

- **Technology adoption:** Identifying informative combinations of behavioural and institutional predictors of agricultural innovation adoption.
- **Digital extension:** Selecting variables that predict farmers' use of mobile advisory applications and digital platforms.
- **Climate adaptation:** Developing parsimonious models of CSA adoption readiness and adaptation behaviour.
- **Extension programme planning:** Identifying information requirements for adoption-risk screening and differentiated advisory services.
- **Smallholder inclusion:** Examining predictive characteristics associated with low participation in agricultural programmes, while assessing potential bias and exclusion.

Nevertheless, computational optimization cannot replace behavioural theory, sound research design or field validation.

CONCLUSION

The work of agricultural extension research is increasingly shifting from describing the adoption behaviour to forecasting results and providing improved decision-making. In agriculture, a useful evolutionary optimization solution to identifying informative combinations of predictors in adoption studies is the Genetic Algorithm.

GA can be combined with Random Forest and explainable AI to aid the development of more parsimonious, interpretable, and potentially efficient systems for forecasting adoption. Its scientific value, however, must be proven by objective benchmarking, model stability independent validation and verification.

The future of agricultural extension research is not just about the algorithms, but about creating credible evidence to inform coherent, inclusive and context specific agricultural extension decisions.

Suggested references and further reading

- Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems*. University of Michigan Press.
- Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32.
- Kuhn, M., & Johnson, K. (2019). *Feature Engineering and Selection: A Practical Approach for Predictive Models*. CRC Press.
- Yang, X.-S. (2020). [Nature-inspired optimization algorithms: Challenges and open problems. *Journal of Computational Science*, 46, 101104.](#)
- Alotibi, Y. S. (2025). A socio-technical agent-based simulation model for predicting smart agriculture adoption dynamics. *Scientific Reports*, 15, Article 43536. <https://doi.org/10.1038/s41598-025-27523-7>
- Shastry, K. A., & Sanjay, H. A. (2021). A modified genetic algorithm and weighted principal component analysis based feature selection and extraction strategy in agriculture. *Knowledge-Based Systems*, 232, Article 107460. <https://doi.org/10.1016/j.knosys.2021.107460>
- Chatterjee, S., Mondal, S. K., Datta, A., & Gupta, H. K. (2024). Enhancing feature optimization for crop yield prediction models. *Current Agriculture Research Journal*, 12(2). <https://doi.org/10.12944/CARJ.12.2.19>

Note: All numerical results presented in the CSA adoption example are derived from a hypothetical computational demonstration and are intended for methodological illustration.

Dr. J. Paul Mansingh is Professor and Head, Department of Agricultural Extension and Economics, VIT School of Agricultural Innovations and Advanced Learning (VAIAL), Vellore Institute of Technology, Vellore, Tamil Nadu, India. He can be reached at paul.mansingh@vit.ac.in

AESA Secretariat: Centre for Research on Innovation and Science Policy (CRISP)
Road No 10, Banjara Hills, Hyderabad 500034, India

www.aesanetwork.org

Email: aesanetwork@gmail.com