Institute of Primate Research

STANDARD OPERATING PROCEDURE (SOP) DOCUMENT

Predictive modelling and ensemble modelling

SOP No.	Issue Number	Issue Date	Revision Status	Revision Date
SOP/KIPRE/RPD/DSAS/3.1.76	Version 01	October 2025	-	-

Approvals

	Name	**Signature**	**Date**
Developed by:	_Patrick Waweru Mwaura_	_____	**_6th October; 2025_**
	_____	_____	_____
	_____	_____	_____
Reviewed by:	_____	_____	_____
Approved by:	_____	_____	_____

Table of Contents

1. PURPOSE

To establish a standardized framework for developing, validating, and applying predictive and ensemble models in DS&AS-supported research, ensuring accuracy, transparency, and compliance with ethical and regulatory standards.

2. SCOPE

Covers all DS&AS projects involving statistical or machine learning predictive modelling, including epidemiology, biomedical outcomes, ecological monitoring, and conservation forecasting.

3. PERSONS RESPONSIBLE:

- **Data Scientist / Biostatistician:** Designs and implements models.
- **Computational Biologist (if genomic/proteomic data):** Applies specialized methods.
- **Head of DS&AS:** Reviews and approves modelling frameworks.

4. FREQUENCY

- Mandatory **before deployment** of any predictive model.
- Model re-validation **when new data or methods** become available.

5. MATERIALS

- Statistical software (R, SAS, Python/Scikit-learn, TensorFlow, PyTorch).
- Model validation datasets.
- Documentation templates (model specification, assumptions, and metrics).

6. PROCEDURE

1. **Model Selection:** Identify candidate models (regression, decision trees, random forests, boosting, and deep learning).
2. **Data Preparation:** Clean, partition (train/validation/test), and preprocess datasets.
3. **Model Development:** Train models using cross-validation; tune hyperparameters.
4. **Validation:** Assess performance (AUC, RMSE, calibration plots, sensitivity analyses).
5. **Ensemble Modelling:** Combine models using bagging, boosting, or stacking where appropriate.
6. **Documentation:** Archive all code, parameters, and validation results.
7. **Deployment:** Deploy validated model with monitoring mechanisms.

7. REFERENCES