

Supplement S2

Model performance and variable importance

Model performance and variable importance are outcomes of GBM multinomial logistic regression based on classification results (refer to Supplement S1) and risk factors. The difference between different classifications are the malaria transmission indicators used. Models were established based on individual/household surveys in each cluster; model evaluation parameters included overall accuracy, sensitive and specificity, and individual group accuracy for $k > 2$. Individual group sensitivity and specificity were calculated using one group against the rest. Importance was calculated as percentage to total importance of 100%.

1. Classification I

Malaria transmission indicators used: Parasite infection prevalence (%) in June and November 2019, vector density (females/house/night) in July 2019 and January 2020.

Table 1. Model performance and variable importance (top 10 only) of classification I

Number of groups & performance			Top 10 variance	Importance (%)
K= 2 †				
Accuracy	0.990		Variable	Importance
Sensitivity	0.995		Altitude	64.5
Specificity	0.976		Distance to River	3.3
			WT N habitats, dry season ‡	2.0
			Distance to Road	2.0
			Distance to Lake	1.8
			LST Day time, Feb ¶	1.8
			Ground surface roughness	1.7
			Topographic wetness index	1.7
			EVI, 2019 ¶¶	1.6
			NDVI, Sep ¶¶	1.5
K = 3 †				
Accuracy	0.921		Variable	Importance
Sensitivity	0.919		Altitude	53.8
Specificity	0.957		Distance to River	5.5
			WT N habitats, dry season ‡	2.7
			Distance to Lake	2.6
			WT N ponds ‡	2.5
			Distance to Health Facility	2.2
			Distance to Road	2.0
			WT N habitats, wet season ‡	1.9
			Evapotranspiration, 2019 ¶¶	1.7
			EVI, 2019 ¶¶	1.6
K = 5 †				
	Sensitivity	Specificity		
Group 1	0.932	0.913		
Group 2	0.924	0.987		
Group 3	0.902	0.971		

Accuracy	0.800		Variable	Importance
Sensitivity	0.801		Altitude	49.6
Specificity	0.886		Distance to River	3.7
			WT N habitats, dry season ‡	3.7
	Sensitivity	Specificity	Distance to Road	3.5
Group 1	0.742	0.951	WT N ponds ‡	2.7
Group 2	0.780	0.886	Distance to Health Facility	2.4
Group 3	0.825	0.858	Distance to Lake	2.3
Group 4	0.845	0.836	NDVI, Oct ¶	2.2
Group 5	0.814	0.901	WT N habitats, wet season ‡	2.0
			NDVI, May ¶	1.9

† k is the number of classified groups.

‡ Weighted number of habitats or water pond by seasons.

¶ Monthly (e.g., May) or annual (e.g., 2019) mean vegetation index or evapotranspiration.
NDVI: normalized difference vegetation index; EVI: enhanced vegetation index.

2. Classification II

Malaria transmission indicators used: Parasite infection prevalence (%) in June and November 2019, vector density (females/house/night) in July 2019 and January 2020, and malaria incidence (cases/1,000 people/month).

Table 2. Model performance and variable importance (top 10 only) of classification II

Number of groups & performance		Top 10 variance		Importance
K = 2 †				
Accuracy	0.989	Variable		Importance
Sensitivity	0.984	Altitude		47.8
Specificity	0.993	WT N ponds ‡		5.9
		Distance to River		2.8
		WT N habitats, wet season ‡		2.6
		EVI, 2019 ¶		2.5
		Ground surface roughness		2.5
		Valley bottom flatness		2.4
		Distance to Lake		2.0
		Aspect		2.0
		Distance to Road		1.7
K = 3 †				
Accuracy	0.906	Variable		Importance
Sensitivity	0.896	Altitude		44.6
Specificity	0.945	WT N habitats, dry season ‡		4.1

	Sensitivity	Specificity		
Group 1	0.942	0.877	WT N habitats, wet season ‡	3.6
Group 2	0.888	0.978	WT N ponds ‡	3.1
Group 3	0.857	0.980	Distance to Road	3.1
			LST Day time, Feb ¶	2.9
			Distance to River	2.8
			Distance to Lake	2.3
			Ground surface roughness	2.0
			EVI, Feb ¶	2.0

† k is the number of classified groups.

‡ Weighted number of habitats or water pond by seasons.

¶ Monthly (e.g., Feb) or annual (e.g., 2019) mean vegetation index or temperature. NDVI: normalized difference vegetation index; EVI: enhanced vegetation index; LST: land surface temperature.

3. Classification III

Malaria transmission indicators used: Parasite infection prevalence (%) in June and November 2019, and malaria incidence (cases/1,000 people/month).

Table 3. Model performance and variable importance (top 10 only) of classification III

Number of groups & performance		Top 10 variance	Importance (%)
K = 2 †			
Accuracy	0.991	Variable	Importance
Sensitivity	0.985	Altitude	47.7
Specificity	0.994	WT N ponds ‡	6.0
		Distance to River	2.7
		WT N habitats, dry season ‡	2.6
		Ground surface roughness	2.6
		EVI, 2019 ¶	2.5
		Valley bottom flatness	2.4
		Distance to Lake	2.1
		Aspect	2.0
		NDVI, Oct ¶	1.7
K = 3 †			
Accuracy	0.911	Variable	Importance
Sensitivity	0.901	Altitude	44.5
Specificity	0.948	WT N habitats, wet season ‡	4.1
		WT N habitats, dry season ‡	3.6
		WT N ponds ‡	3.1
		Distance to Road	3.1

Group 2	0.892	0.979	Distance River	2.9
Group 3	0.868	0.981	LST Day time, Feb ¶	2.8
			Distance to Lake	2.4
			Ground surface roughness	2.0
			EVI, Feb ¶	2.0

† k is the number of classified groups.

‡ Weighted number of habitats or water pond by seasons.

¶ Monthly (e.g., Oct) or annual (e.g., 2019) mean vegetation index or temperature. NDVI: normalized difference vegetation index; EVI: enhanced vegetation index; LST: land surface temperature.