One Health 13 (2021) 100358

Contents lists available at ScienceDirect

One Health

journal homepage: www.elsevier.com/locate/onehlt

Mapping the spatial distribution of the dengue vector *Aedes aegypti* and predicting its abundance in northeastern Thailand using machine-learning approach

M.S. Rahman[a,b,**], Chamsai Pientong[a,c], Sumaira Zafar[d], Tipaya Ekalaksananan[a,c], Richard E. Paul[e], Ubydul Haque[f], Joacim Rocklöv[g], Hans J. Overgaard[a,h,*]

[a] Department of Microbiology, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand
[b] Department of Statistics, Begum Rokeya University, Rangpur, Rangpur-5404, Bangladesh
[c] HPV & EBV and Carcinogenesis Research Group, Khon Kaen University, Khon Kaen, Thailand
[d] Environmental Engineering and Management Program, Asian Institute of Technology, Pathumthani, Thailand
[e] Unité de la Génétique Fonctionnelle des Maladies Infectieuses, Institut Pasteur, CNRS UMR 2000, 75015 Paris, France
[f] Department of Biostatistics and Epidemiology, University of North Texas Health Science Center, Fort Worth, TX 76177, USA
[g] Department of Public Health and Clinical Medicine, Umeå University, 90187 Umeå, Sweden
[h] Faculty of Science and Technology, Norwegian University of Life Sciences, P.O. Box 5003, Ås, Norway

ARTICLE INFO

Keywords:
Supervised learning
Aedes aegypti
Prediction
Dengue
Early warning

ABSTRACT

Background: Mapping the spatial distribution of the dengue vector *Aedes (Ae.) aegypti* and accurately predicting its abundance are crucial for designing effective vector control strategies and early warning tools for dengue epidemic prevention. Socio-ecological and landscape factors influence *Ae. aegypti* abundance. Therefore, we aimed to map the spatial distribution of female adult *Ae. aegypti* and predict its abundance in northeastern Thailand based on socioeconomic, climate change, and dengue knowledge, attitude and practices (KAP) and/or landscape factors using machine learning (ML)-based system.

Method: A total of 1066 females adult *Ae. aegypti* were collected from four villages in northeastern Thailand during January–December 2019. Information on household socioeconomics, KAP regarding climate change and dengue, and satellite-based landscape data were also acquired. Geographic information systems (GIS) were used to map the household-based spatial distribution of female adult *Ae. aegypti* abundance (high/low). Five popular supervised learning models, logistic regression (LR), support vector machine (SVM), k-nearest neighbor (kNN), artificial neural network (ANN), and random forest (RF), were used to predict females adult *Ae. aegypti* abundance (high/low). The predictive accuracy of each modeling technique was calculated and evaluated. Important variables for predicting female adult *Ae. aegypti* abundance were also identified using the best-fitted model.

Results: Urban areas had higher abundance of female adult *Ae. aegypti* compared to rural areas. Overall, study respondents in both urban and rural areas had inadequate KAP regarding climate change and dengue. The average landscape factors per household in urban areas were rice crop (47.4%), natural tree cover (17.8%), built-up area (13.2%), permanent wetlands (21.2%), and rubber plantation (0%), and the corresponding figures for rural areas were 12.1, 2.0, 38.7, 40.1 and 0.1% respectively. Among all assessed models, RF showed the best prediction performance (socioeconomics: area under curve, AUC = 0.93, classification accuracy, CA = 0.86, F1 score = 0.85; KAP: AUC = 0.95, CA = 0.92, F1 = 0.90; landscape: AUC = 0.96, CA = 0.89, F1 = 0.87) for female adult *Ae. aegypti* abundance. The combined influences of all factors further improved the predictive accuracy in RF model (socioeconomics + KAP + landscape: AUC = 0.99, CA = 0.96 and F1 = 0.95). Dengue prevention practices were shown to be the most important predictor in the RF model for female adult *Ae. aegypti* abundance in northeastern Thailand.

Abbreviations: DENV, Dengue virus; GIS, Geographic information systems; ML, Machine learning; KAP, Knowledge, attitude, and practice; SES, Socioeconomic status; PCI, Premise condition index; HCI, Household crowding index; LR, logistic regression; SVM, Support vector machine; kNN, k-nearest neighbor; ANN, Artificial neural network; RF, Random forest; AUC, Area under curve; CA, Classification accuracy..

* Correspondence to: H. J. Overgaard, Faculty of Science and Technology, Norwegian University of Life Sciences, P.O. Box 5003, Ås, Norway.
** Correspondence to: M.S. Rahman, Department of Microbiology, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand.
E-mail addresses: siddikur@brur.ac.bd (M.S. Rahman), hans.overgaard@nmbu.no (H.J. Overgaard).

https://doi.org/10.1016/j.onehlt.2021.100358
Received 10 June 2021; Received in revised form 2 December 2021; Accepted 2 December 2021
Available online 4 December 2021

Conclusion: The RF model is more suitable for the prediction of *Ae. aegypti* abundance in northeastern Thailand. Our study exemplifies that the application of GIS and machine learning systems has significant potential for understanding the spatial distribution of dengue vectors and predicting its abundance. The study findings might help optimize vector control strategies, future mosquito suppression, prediction and control strategies of epidemic arboviral diseases (dengue, chikungunya, and Zika). Such strategies can be incorporated into One Health approaches applying transdisciplinary approaches considering human-vector and agro-environmental interrelationships.

1. Introduction

Dengue is an infectious disease caused by the dengue virus (DENV) and transmitted by *Aedes* mosquitoes. Dengue virus is the most deadly arboviral infection in terms of human morbidity and mortality and threatens around half of the world's population [1]. The global incidence of dengue has increased considerably in the last 40 years, with estimates of 390 million annual cases in at least 128 countries [1]. Dengue has become endemic in most tropical and subtropical regions globally; Southeast Asia and the Western Pacific have been particularly hard hit [2]. Thailand is a South-East Asian country where dengue is endemic in urban and rural regions, with all four dengue serotypes present [3,4]. As a result, Thailand's frequent dengue outbreaks endanger public health, causing hospitalization and periodically death in the country [3].

Aedes (Ae.) aegypti, the primary dengue vector, is widely distributed globally, including Thailand. Identifying the spatial distribution of *Ae. aegypti* abundance requires a well-designed entomological surveillance system that provides robust entomological data on local mosquito population densities. Another challenge lies in implementing and sustaining such programs in many dengue-endemic countries [5]. Adult *Ae. aegypti* mosquitoes are also challenging to capture and are rarely collected [5–7]. The abundance of *Ae. aegypti* has become an increasing problem and is affected by several socio-ecological factors such as local socioeconomic conditions, education status, land use/cover changes, climate change, urban growth patterns, and dengue prevention practices [8–12]. In this study, the complex nonlinear relationships between socioeconomic, climate change and dengue knowledge, attitude, and practices (KAP), landscape factors, and female adult *Ae. aegypti* abundance poses a serious challenge to analyze using more commonly employed hypothesis-driven parametric regression method (e.g., logistic regression model) [13]. Machine learning (ML)-based modeling techniques have gained popularity in the recent decade in analyzing and handling complex databases (e.g., outliers, nonlinearity, multicollinearity) in infectious disease studies without depending on a priori hypotheses [11,14–16]. The robust estimation and better predictive accuracy of ML models have been extensively recognized [11]; however, such models have not been widely used to predict dengue vector abundance [11,15,17,18].

Due to the lack of efficient dengue treatment alternatives and vaccine limitations [3,19], dengue prevention is currently limited to vector control. Mapping and prediction form an essential part of vector surveillance systems, specifically in early warning systems (EWS) of disease outbreaks and transmission [20]. In its global plan for dengue prevention and control, WHO highlighted the need for developing predictive models to assess the risk of dengue outbreaks [5]. Understanding the spatial distribution of *Ae. aegypti* can also help prevent dengue epidemics, including chikungunya and Zika [21]. To formulate strategies for public health planning and vector control, this study i) mapped the spatial distribution of female adult *Ae. aegypti* abundance (encoded as high/low) among households in four study sites in northeastern Thailand, ii) developed an operational and robust ML model to predict female adult *Ae. aegypti* abundance (encoded as high/low) in households using socioeconomic, KAP, and landscape factors in households of northeastern Thailand, a dengue-endemic area, and the country's third-largest region in terms of people and land area. We built four commonly-used ML learning models, including support vector machines (SVM), k-nearest neighbor (kNN), artificial neural network (ANN), and random

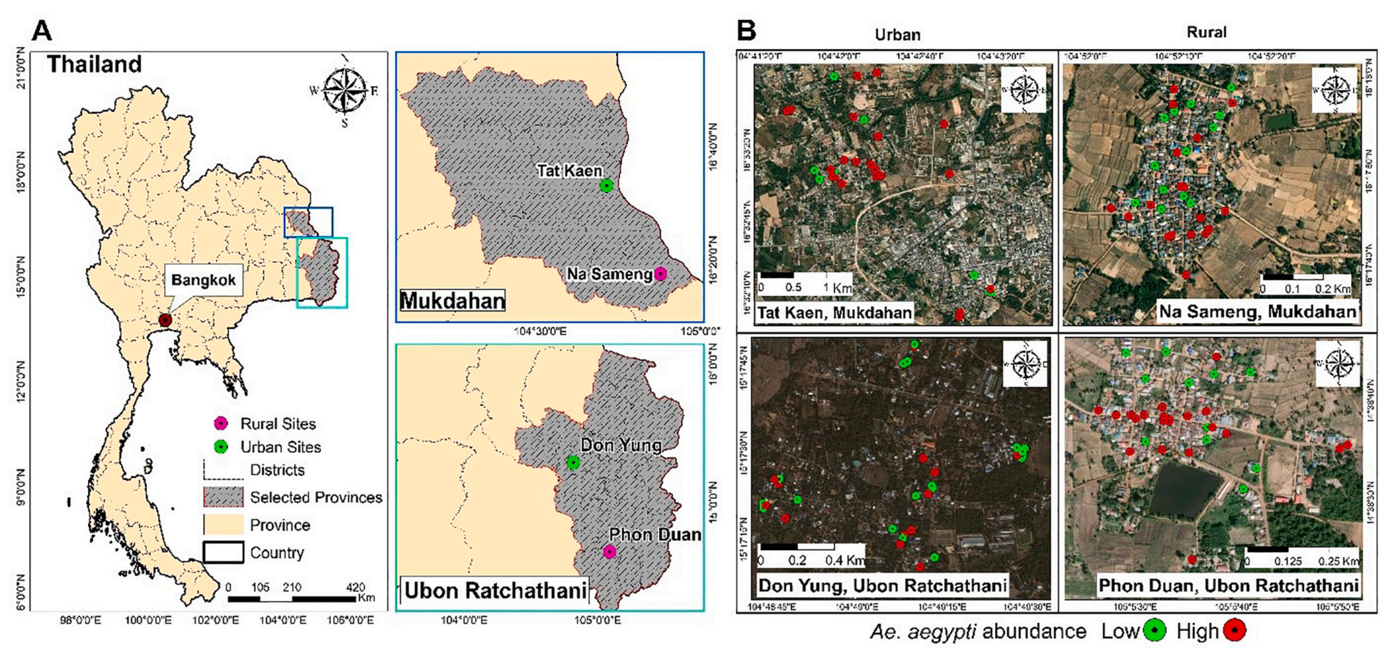

Fig. 1. (A) Locations of the four *Aedes* data collection sites, B) Spatial distribution of the dengue vector, female adult *Ae. aegypti* abundance (high vs. low) calculated based on median values above and below respectively in 128 households of northeastern Thailand during Janu-ary-December 2019.

Table 1
Description of climate change and dengue knowledge, attitude and practice (KAP) scores, premise condition index (PCI), household crowding index (HCI) and socioeconomic status (SES).

Index	Variables	Description	Classification score
Climate change knowledge scores	Climate change knowledge	Beliefs, understanding & awareness about climate change and its connection to dengue, local & global climate change problem	1-correct,0.5-moderately good, or good,0-incorrect, poor, or "do not know" answers
Climate change attitude scores	Climate change attitude	Concern and seriousness about the climate change	1 for "positive," 0.5 for "moderate," and 0 for "negative"/ "not sure"
Climate change practice scores	Climate change practice	Climate change adaptation and mitigation practices	1-correct,0.5-moderately good, or good,0-incorrect, poor, or "do not know" answers
Dengue knowledge scores	Dengue knowledge	Transmission of dengue, symptoms, and signs of dengue, vector morphology (identification), vector breeding (places)	1-correct,0.5-moderately good, or good,0-incorrect, poor, or "do not know" answers
Dengue attitude scores	Dengue attitude	Strategies of mosquito control and dengue prevention	1 for "positive," 0.5 for "moderate," and 0 for "negative"/ "not sure"
Dengue practice scores	Dengue practice	Dengue prevention practices, bite prevention practices, destroying *Aedes* breeding sites	-correct,0.5-moderately good, or good,0-incorrect, poor, or "do not know" answers
Premise condition index (PCI)	House condition	Good (well-maintained, e.g., newly painted or new house)	1
		Intermediate (moderately well-maintained house)	2
		Bad (not well-maintained house, e.g., paint peeling, broken items visible, dilapidated old house)	3
	Yard condition	Good (tidy yard)	1
		Intermediate (moderately tidy yard)	2
		Bad (untidy yard)	3
	Shade condition	Not shaded (very little or no shade)	1
		Intermediate (some shade: > 25% but <50%)	2
		Shady (plenty of shade: > 50%)	3
	Water supply and storage	Piped water	1
		Ground water/well water supply	2
		Rainwater and/or open water source: river/stream/lake/ mountain water/river water	3
Household crowding index (HCI)	Co-residents	Monthly number of co-residents per household	–
	Number of rooms	Number of rooms per household	–
Socioeconomic status (SES)	House roof material	Ceramic/Wood/ Metal	–
	House walls material	Plastered/Cement/ Bricks/Wood	–
	Ownership of durable assets	Television/VCD/ Refrigerator/Washing machine/Mobile/ Smartphone/ Computer/Oven/ Microwave/ Airconditioner/Car/ Pickup/ Motorcycle	–
	Ownership of toilet facility	Yes/No	–
	Toilet /bathroom floor material	Tiles/Cement/Earth	–
	Ownership of flush toilet/ squat toilet	Yes/No	–

forest (RF), to predict female adult *Ae. aegypti* abundance based on socioeconomics, KAP, and landscape as well as for all factors combined. The performance of the ML models was compared to that of a more often used traditional logistic regression (LR) model. The findings of this study will help to better identify mosquito-infested locations and high risks of mosquito-borne infectious diseases, such as dengue, chikungunya, and Zika.

2. Materials and methods

2.1. Mosquito collection and handling

Female adult *Ae. aegypti* mosquito collections were carried out indoors and outdoors of 128 households in four study sites (two urban and two rural) in two provinces in northeastern Thailand from January to December 2019 (Fig. 1). Mosquitoes were collected using battery-powered Prokopack aspirators for 10 min indoors (in main rooms of activity, e.g., living rooms and bedrooms) and 10 min outdoors (among man-made articles, cars, motorcycles, vegetation, tree holes, roof gutters, etc.) in each household [22]. Collected mosquitoes were stored in the aspirator collection cups in styrofoam boxes and brought back to the laboratory. Mosquitoes from these samples were killed by freezing and then morphologically identified to *Ae. aegypti* using a stereomicroscope. Mosquitoes were individually preserved at −20 °C in 1.5 mL microcentrifuge tubes until further analysis. A detailed description of the four study sites, mosquito collection, and handling procedures were reported in our previous study [23].

2.2. Socioeconomic and KAP

Household-based socioeconomic and KAP factors associated with female adult *Ae. aegypti* abundance and potential dengue risk were included in this study: urban-rural residence, education status, household income, socioeconomic status (SES), household crowding index (HCI), premise condition index (PCI), climate change and dengue knowledge, attitude, and practices (KAP) scores. These factors were chosen from a broader collection of household and KAP surveys reported in our previous study [23]. A detailed description of these factors is presented in Table 1.

2.3. Landscape data

Potential land use/land cover factors that are considered to impact human-mosquito interactions were included: built-up area, permanent wetlands, natural tree cover, rubber plantation, rice crop, all assessed as percentage in the 30 m radius area of the sampling households. Supervised maximum likelihood classification was used to acquire and map land use/land cover factors from Sentinel 2 satellite images at 10-m

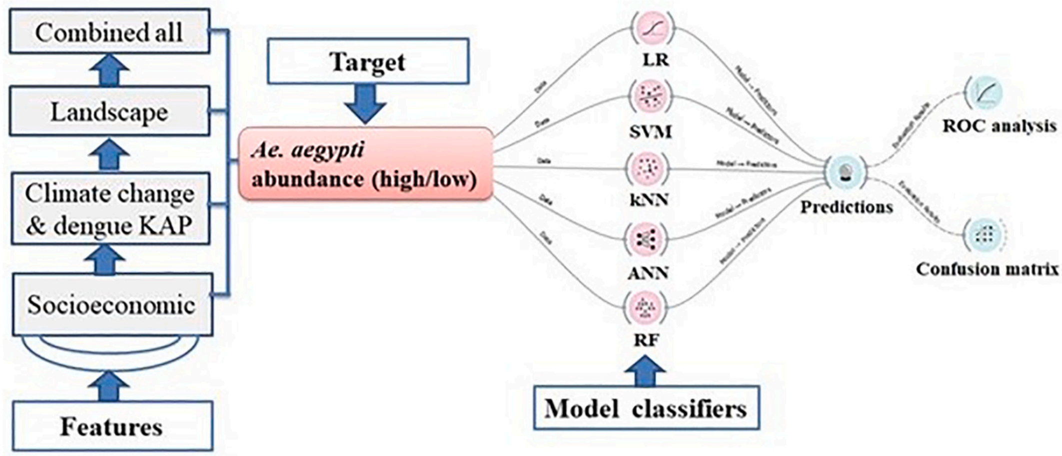

Fig. 2. The pipeline of ML model workflow. The left side (features) shows the input *fac*-tors/predictors; the right side (models classifiers and predictive measurements) produces the full dataset model output and overall predictive performances of each classifier to predict female adult *Ae. aegypti* abundance (high/low).

resolution [24]. House density within 30 m radius (houses per km^2) was also calculated by mapping the neighbouring households using high-resolution google earth images.

2.4. Creating the dengue vector abundance variable

The target variable (female adult *Ae. aegypti* abundance) was created (1 = high, 0 = low) in comparison to the median value above and below, respectively. Thus, there are 75 households for *Ae. aegypti* abundance = high and 53 households for *Ae. aegypti* abundance = Low. This will provide public health officers and the general public with an operational criterion for estimating relative dengue or other infectious disease vector abundance in specific houses or locations [11].

2.5. Model building, prediction, and performance evaluation

Predictive modeling was conducted using Orange, an open-source machine learning and data mining toolkit [25]. Five common and relatively more popular supervised classification models: Logistic regression (LR), support vector machine (SVM), k-nearest neighbors (kNN), artificial neural network (ANN), and random forest (RF) were used to predict female adult *Ae. aegypti* abundance based on socioeconomic, KAP and/or landscape factors. These five classification models were picked from among many others recently developed classification models because of their popularity and applicability [11]. Four sets of factors were utilized as inputs for each type of model: socioeconomic factors only, KAP factors, landscape factors, and all factors together. A cross-validation technique with ten folds was applied to evaluate the classifiers because of the relatively small sample size [11]. Statistical analyses were performed using RStudio [54].

Predictive efficiency refers to the capacity of the classifier to predict the *Ae. aegypti* abundance. A classifier is effective if it has good classification performance, assessed by AUC (Area Under the ROC Curve), CA (Classification Accuracy), and F1 Score. Receiver operating characteristics (ROC) is one of the most important evaluation measures for assessing the predictive performance of any classification model. The ROC curve is created by plotting sensitivity versus '1 – specificity'. The AUC, which is calculated from the ROC curve, gives an overall assessment of model fit and, in comparison, is more efficient. The model fit is excellent when the AUC is greater than 0.9. If the area is between 0.8 and 0.9, it is considered a good fit. A fit that is between 0.7 and 0.8 is considered acceptable. A fit of 0.5 to 0.7 suggests a poor fit, whereas a fit of 0.5 implies no fit [26]. CA is the most intuitive performance measure, and it is simply a ratio of correctly predicted *Ae. aegypti* abundance (high/low) to the observed *Ae. aegypti* abundance. Precision is the percentage (or proportion) of *Ae. aegypti* abundance cases classified correctly with respect to the observed *Ae. aegypti* abundance (high). Recall is the percentage (or proportion) of *Ae. aegypti* abundance (high) correctly classified by the model. F1 Score is the weighted average of Precision and Recall. A detailed conceptual framework of the model building and evaluation process is outlined in Fig. 2.

LR is a supervised learning model, and the predictors are continuous/discrete, and the response variable is dichotomous (high *Ae. aegypti* abundance vs. low). The logit of the response variable (Y) is the linear combination of the predictors (X), which can be written as follows:

$$logit(P_j) = log_e\left(\frac{P_j}{1-P_j}\right) = \sum_{i=0}^{K} \beta_i X_i$$

where P_j is defined as the probability for Y = 1 (high abundance) and 1 − P_j (low abundance) is defined when Y = 0. β_i's (i = 0, 1, …,K) are the unknown regression coefficients, K is the total number of predictors (18 factors) and X_i's (i = 1, …, K) are the predictors [27].

SVM attempts to distinguish between the two classes (high and low *Ae. aegypti* abundance). SVM's classification task can be implemented using a variety of kernel functions (e.g., radial, linear, sigmoid, and polynomial) [28–30]. However, Sachindra et al. (2018) and Smits, et al. (2002) found that SVM is outperformed by polynomial kernel function [29,31]. Hence, the polynomial kernel was employed in this study. The polynomial kernel is defined as follows:

$$K(x_i, x_j) = (x_i.x_j + c)^d$$

where x_i and x_j are the predictors, the predicted data, d is the degree of the polynomial, and c is a constant that allows for a trade-off between the higher and lower order parts' influence.

kNN is a supervised learning technique that uses surrounding data points to classify a new data point into the target class. For example, the predicted value of *Ae. aegypti* abundance (classifier) for kNN was based on its surrounding "neighbors" whose "identity" (i.e., a 0 or 1 value corresponding to low and high levels of female adult *Ae. aegypti* abundance) was already known.

An ANN can contain multiple layers and neurons. In general, the ANN architecture has three layers: an input layer where predictor data is entered, a hidden layer with nonlinear activation functions that transfer inputs to outputs, and an output layer where simulations/predictions are stored) [11]. ANN maps input "neurons" (socioeconomic, KAP, and/or landscape factors) to output "neurons" (female adult *Ae. aegypti* abundance) using hidden layers.

RF is a commonly used ML-based classifier that functions by constructing decision trees [32]. RF was used to predict and assess the

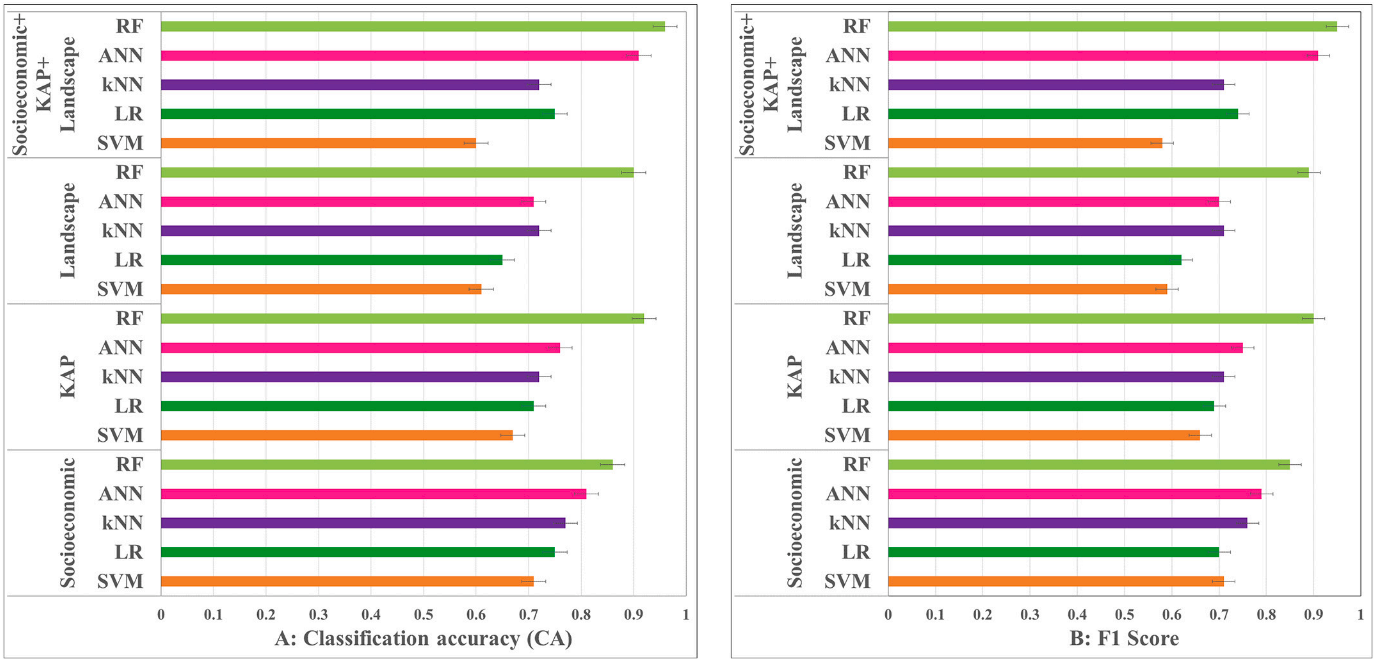

Fig. 3. Predictive performance evaluation parameters of five models to predict *Ae. aegypti* abundance (high/low). SVM: Support vector machine, LR: logistic regression, kNN: k-nearest neighbor, ANN: Artificial neural network, and RF: Random forest.

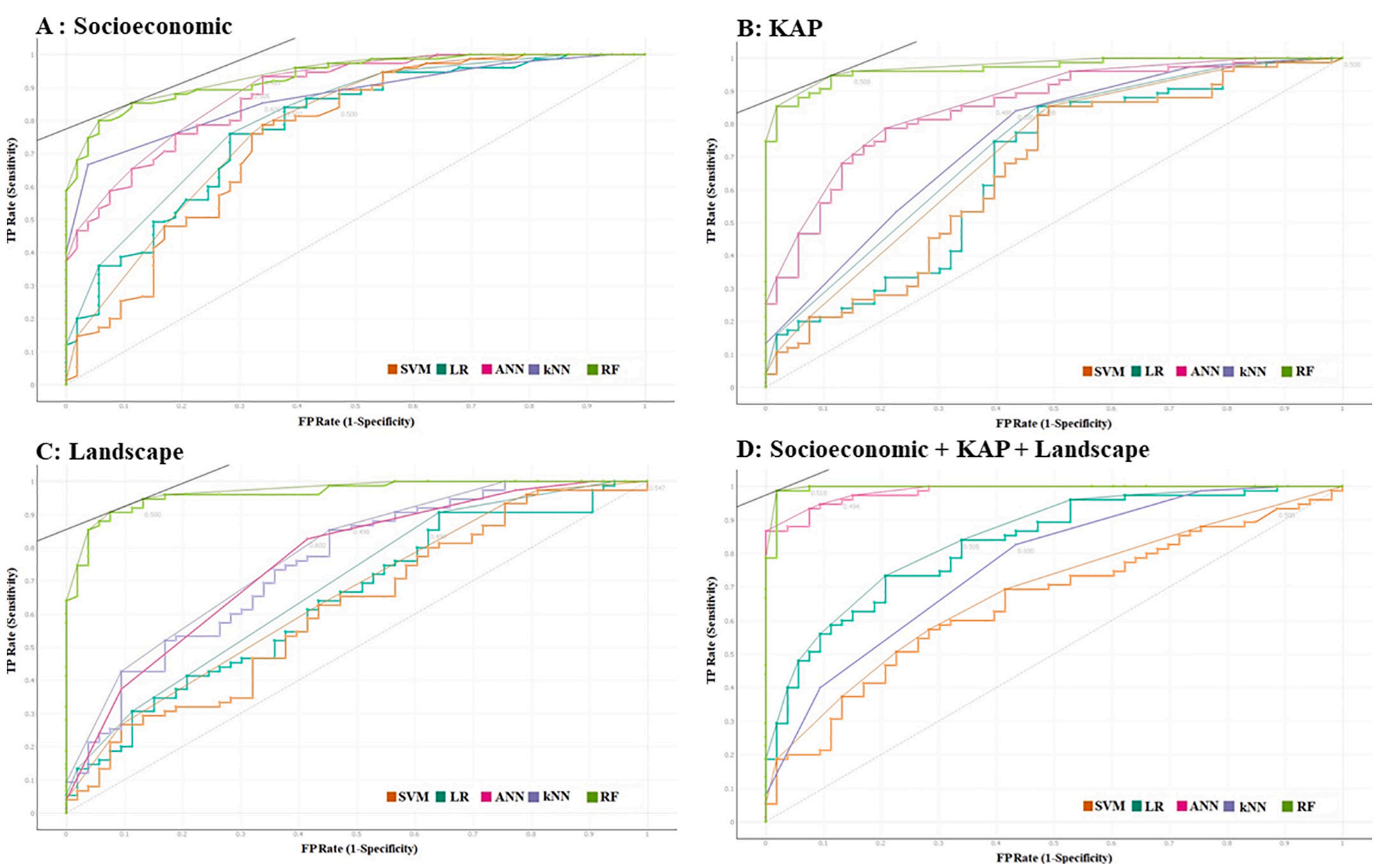

Fig. 4. Receiver operating characteristics (ROC) curves predicting female *Ae. aegypti* abundance (high/low) in 128 households of northeastern Thailand during January–December 2019.

relative importance of predictors on female adult *Ae. aegypti* abundance. The predictor's relative importance was determined using the mean decrease of the Gini coefficient [8]. The higher the coefficient, the greater the predictors's contribution to female *Ae. aegypti* abundance; all contributions should add up to one.

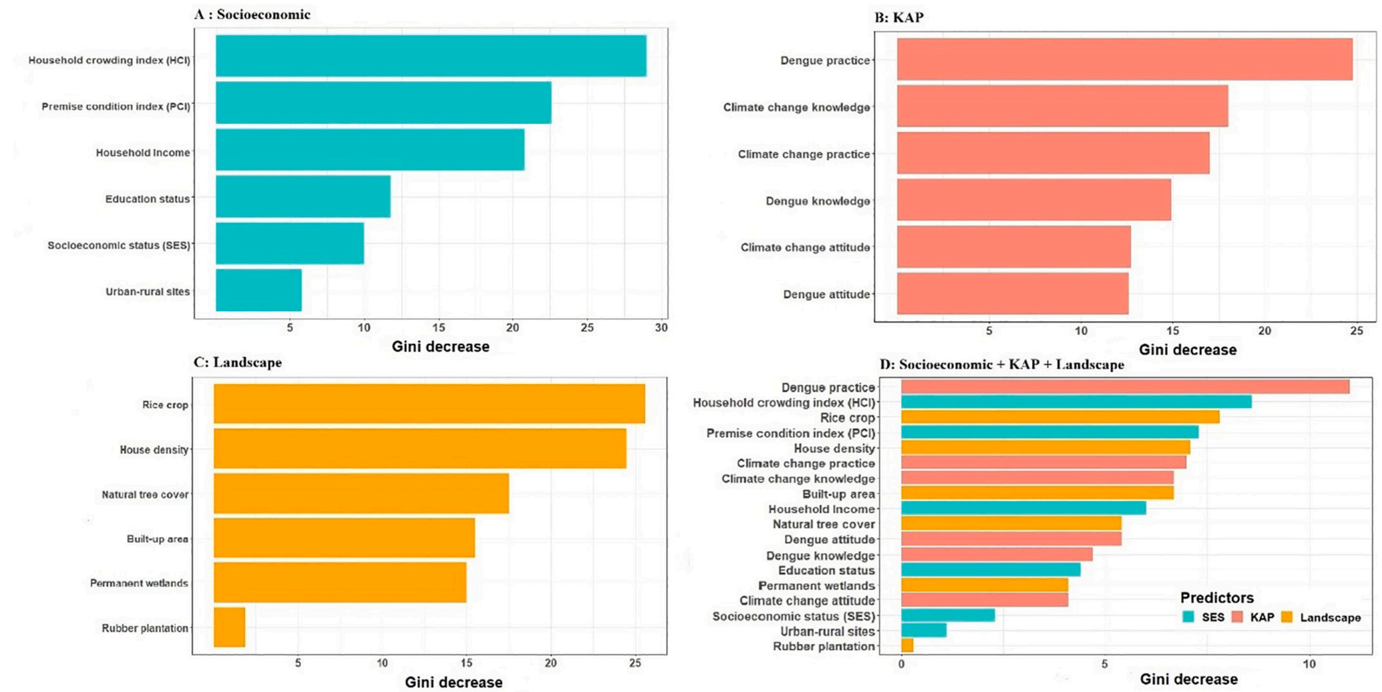

Fig. 5. Mean decrease in Gini of random forest important predictors for female adult *Ae. aegypti* abundance (high/low) in 128 households of northeastern Thailand during January–December 2019.

3. Results

3.1. Aedes mosquito collections, socioeconomic, landscape and KAP factors

A total of 1066 females adult *Ae. aegypti* mosquitoes were collected in both urban (551) and rural (515) sites; the numbers collected during the wet season (May–October) were 345 and 284, respectively. Out of 128 households, 59% had a high abundance of female adult *Ae. aegypti* mosquitoes (Fig. 1). Residents in urban areas had higher KAP scores on climate change and dengue than rural areas, but there were no significant differences regarding KAP levels between the two areas (Tables S1–S6). Result on socioeconomic and KAP factors can be found in our previous study [23]. Average landscape factors per household in urban areas were rice crop (47.4%), natural tree cover (17.8%), built-up area (13.2%), permanent wetlands (21.2%), and rubber plantation (0%), and the corresponding figures for rural areas were 12.1, 2.0, 38.7, 40.1 and 0.1% respectively.

3.2. Performance analysis of machine learning system

RF, which used socioeconomic, KAP, and landscape factors as inputs, had the highest prediction model accuracy, followed by ANN, kNN, LR, and SVM (Figs. 3 and 4). Among all assessed models, RF exhibited the best prediction performance (socioeconomic: CA = 0.86 and F1 = 0.85; KAP: CA = 0.92 and F1 = 0.90, and landscape: CA = 0.89 and F1 = 0.87). In RF, combining all types of covariates increased model accuracy (socioeconomic + KAP+ landscape: CA = 0.96 and F1 = 0.95). In general, including KAP factors alone performed better than socioeconomic and landscape factors alone across all models (Fig. 3).

Fig. 4 shows the ROC curves for all models for better efficiency compared to accuracy. The AUC was also found to be highest in RF (socioeconomic: 0.93, KAP: 0.95; landscape: 0.96, and socioeconomic + KAP+ landscape: = 0.99) compared to ANN (socioeconomic: 0.87, KAP: 0.84, landscape: 0.75, and socioeconomic + KAP+ landscape: 0.98); kNN (socioeconomic: 0.87, KAP: 0.75, landscape: 0.75, and socioeconomic + KAP+ landscape: 0.76); LR (socioeconomic: 0.78, KAP: 0.66, landscape: 0.64, and socioeconomic + KAP+ landscape: 0.82); and SVM (socioeconomic: 0.75, KAP: 0.65, landscape: 0.61, and socioeconomic + KAP+ landscape: 0.65). The range exhibited in RF for all combinations of the socioeconomic, KAP and landscape factors is acceptable to excellent model fit.

3.3. Important socioeconomic, KAP, and landscape factors in predicting female adult Ae. aegypti abundance

The relative importance of predictors of *Ae. aegypti* abundance was identified by the mean decrease in the Gini coefficient of the best fitted RF model (Fig. 5). In general, KAP factors were more important than socioeconomic and landscape factors for predicting female adult *Ae. aegypti* abundance. KAP factors contributed 40% to the prediction of *Ae. aegypti* abundance, while landscape and socioeconomic factors only contributed 31% and 29%, respectively. Besides, dengue prevention practices were the single most important predictor (11%) for mosquito abundance, followed by HCI (8.6%) and rice crop (7.8%).

4. Discussion

Here we present the spatial distribution of female adult *Ae. aegypti* and prediction of its abundance (classified into high and low) in northeastern Thailand using an ML-based system. An LR-based model was adopted for the same purpose. The study demonstrated that overall climate change and dengue-related KAP factors and landscape factors were more important in explaining mosquito abundance than socioeconomic factors in northeastern Thailand based on the proposed RF ML model. This finding is in line with a recent study in the United States, which indicated that landscape factors were more relevant than socioeconomic ones in predicting vector abundance [11]. KAPs were also found to be related to dengue transmission and vector control in several countries [33–35], including Thailand [36,37]. Therefore, the assessment of KAP might be a vital tool for planning social mobilization and community integration in dengue and vector control [38]. In addition, dengue prevention practices, rice crops, and house crowding index (HCI) were the most influential factors within the KAP, landscape, and

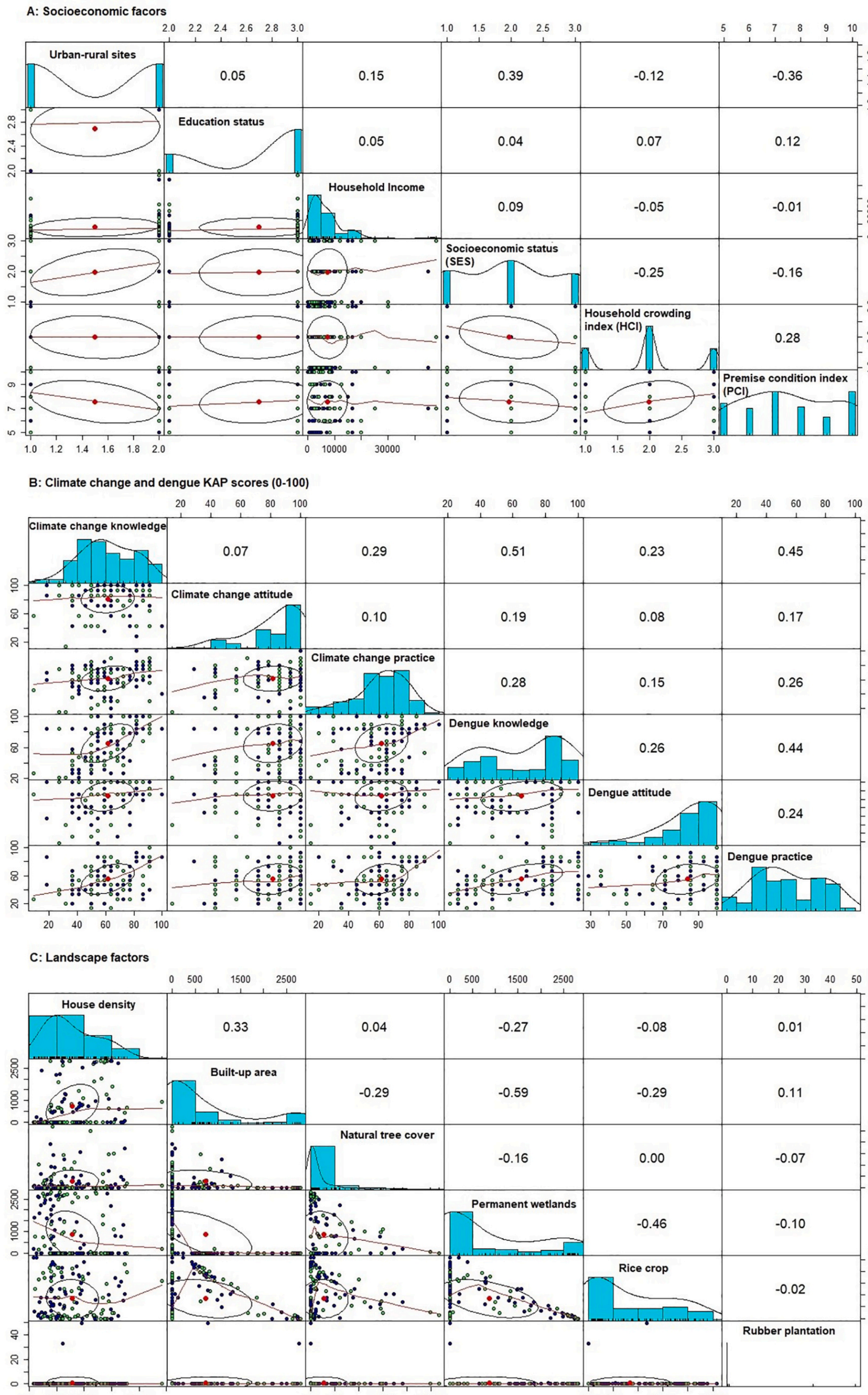

Fig. 6. Scatter diagram (lower left), histogram (diagonal), and correlation coefficients (upper right) of relationships between socioeconomic factors (A), KAP scores (B), landscape factors (C), and female adult *Ae. aegypti* abundance. "Blue", "green" color represents high and low female adult *Ae. aegypti* abundance respectively in scatter diagram. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

socioeconomic domains in predicting mosquito abundance. Methods to prevent vector breeding and human-vector contact can reduce *Ae. aegypti* abundance. The elimination of larval habitats (i.e., source reduction) in the domestic and peridomestic environments is frequently advocated as a simple and effective technique of managing dengue vector populations [39]. Increased rice production area due to forest conversion is a major land-use change in northern Thailand [40,41]. Association between rice crop area and mosquito abundance shows the complexity of features underlying mosquito abundance in northeastern Thailand. Therefore, it might be essential to include other public sectors, such as agriculture, and not only the public health sector, in controlling dengue vectors [42]. A previous study conducted in Mali also reported that rice cultivation has dramatic consequences for human health and might sharply increase mosquito abundance and vector-borne diseases, including malaria [43]. Another important socioeconomic predictor of female adult *Ae. aegypti* abundance is household population density, represented by HCI. This human-vector proximity factor has been reported in several studies where human population density positively affected the number of *Ae. aegypti* females and dengue transmission [44–46]. Higher HCI provides a suitable environment for human-vector contact, potentially increasing transmission of dengue in Thailand and similar dengue-endemic countries. To control vector-borne diseases in entire communities or regions, it is essential to plan, implement, and integrate multiple initiatives, such as social mobilization, campaigns, climate change adaptation and mitigation methods, and vector control. More importantly, effective communication and coordination between different institutions working on climate change and dengue and local people are essential to identify key constraints and practical challenges in dengue management [3]. In recognition of the deleterious effects of pesticides on biodiversity, the generation of resistance with potential impact on the agricultural sector and for impact on human health, the adoption of greener solutions for mosquito control within a One Health perspective has been called for [47–50]. The development of a global strategy for effective dengue control within which the one health approach is embedded will, however, be a challenge. Current examples tend to rely heavily on community engagement where improved KAP can reduce the burden of disease [51]. The extent to which education and community participation alone can sufficiently reduce the disease burden is, however, debatable [52]. Strategies combining greener solutions with community engagement should be encouraged and developed.

The aforementioned important predictors for female adult *Ae. aegypti* abundance could be detected and quantified by an ML-based system, mainly supervised learning models. These models provide powerful tools for detecting potential hidden relationships within complex datasets that are typically undetectable by traditional methods [11]. For example, there exists no significant correlation and linear relationship between socioeconomic, KAP, and landscape factors, and *Ae. aegypti* abundance (Fig. 6). The more often employed hypothesis-driven parametric models struggle with this type of complex nonlinear relationship. Recently, data-driven ML techniques such as ANN, SVM, and RF have shown promising results in predictive analytics in mosquito-borne disease [53]. ML models, especially RF, were found to be appropriate to identify hidden interactions between different factors and provide high accuracy in this study. Furthermore, our suggested model included socioeconomic, climate change and dengue KAP, and micro-scale landscape characteristics as features/inputs, which are less dynamic than other typically utilized climatic variables like temperature, rainfall, and humidity, lowering monitoring and data collection expenses [11]. Thus, ML models can estimate and predict *Ae. aegypti* abundance accurately across different socioeconomic, KAP and landscape factors. This ML model output can provide integrated evidence for public health officials and the general public to know where there will be high mosquito abundance (i.e., a "hot-spot"). While the model has used 128 sampling households in northeastern Thailand, predictions of *Ae. aegypti* abundance could be made using this method in other places as long as their socio-ecological parameters are measures. More information on which other factor combinations would predict high *Ae. aegypti* abundance could be gained, which could aid in the development of more effective dengue vector surveillance and control programs in northeastern Thailand. The proposed ML models will be applied throughout several years and in numerous locations in the future for this type of research.

5. Limitations

Although RF achieves the best model performance with socioeconomic, KAP, and landscape variables inputs in this study, other models with different input sets may yield greater forecasting from RF based on the new observed dataset. A limitation of our study, therefore, is that, for example, fine scale climate data such as temperature, rainfall, and humidity variables were not incorporated but which are known to impact mosquito densities [11]. As mentioned above, this will be explored progressively with increasing data.

6. Conclusions

ML models are useful for classifying, predicting, and identifying the important factors associated with dengue vector abundance. The performances of LR, SVM, kNN, ANN, and RF classifiers to predict female adult *Ae. aegypti* abundance (high/low) were evaluated. The RF model predicted female adult *Ae. aegypti* abundance efficiently based on socioeconomic, KAP, and landscape factors in 128 households of northeastern Thailand. Combined effects of all factors could improve the prediction accuracy. Moreover, dengue prevention practices were found to be the most important predictor of *Ae. aegypti* abundance. We conjecture that the proposed RF model might be used for the large-scale prediction of mosquito abundance in other locations and other vector-borne diseases, including dengue. The use of the ML models for the prediction of *Ae. aegypti* abundance can provide important information to healthcare authorities to design improved vector surveillance and better prepare for dengue fever outbreaks.

Ethics approval and consent

Ethical statements for human and animal subjects were approved by the Khon Kaen University Ethics Committee (Ref. No. HE611228, 02/08/2018 and HE631077, 24/03/2020), and the Regional Committees for Medical and Health Research Ethics in Norway (2018/1085/REK sør-øst C, 27/06/2018). The household head and study participants' written consent was obtained before interviews of all participating households.

Funding

This study received funding from the Research Council of Norway [DENCLIM project, grant number 281077] and Khon Kaen University Faculty of Medicine (grant number IN63312).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors sincerely thank all field staff, village health volunteers, anonymous government officials, and local authorities in Thailand for their assistance and support. Particular thanks to Dr. Petchaboon Poolphol and staff at the Office of Disease Prevention and Control 10, Ubon Ratchathani, Thailand; Tanyee Sukanda, Sirinart Aromseree, Supranee Phanthanawiboon, Dyna Doum, and Panwad Tongchai at Khon Kaen University, Thailand.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.onehlt.2021.100358.

References

[1] S. Bhatt, P.W. Gething, O.J. Brady, J.P. Messina, A.W. Farlow, C.L. Moyes, et al., The global distribution and burden of dengue, Nature 496 (7446) (2013) 504–507, https://doi.org/10.1038/nature12060.
[2] N.E. Murray, M.B. Quam, A. Wilder-Smith, Epidemiology of dengue: past, present and future prospects, Clin. Epidemiol. 5 (2013) 299–309, https://doi.org/10.2147/clep.S34440.
[3] M.S. Rahman, H.J. Overgaard, C. Pientong, M. Mayxay, T. Ekalaksananan, S. Aromseree, et al., Knowledge, attitudes, and practices on climate change and dengue in Lao People's Democratic Republic and Thailand, Environ. Res. 193 (2021), 110509.
[4] N. Vannavong, R. Seidu, T.A. Stenström, N. Dada, H.J. Overgaard, Effects of socio-demographic characteristics and household water management on Aedes aegypti production in suburban and rural villages in Laos and Thailand, Parasit. Vectors 10 (2017) 1:170, https://doi.org/10.1186/s13071-017-2107-7.
[5] S. Maneerat, E. Daudé, A spatial agent-based simulation model of the dengue vector Aedes aegypti to explore its population dynamics in urban areas, Ecol. Model. 333 (2016) 66–78.
[6] D.A. Focks, A Review of Entomological Sampling Methods and Indicators for Dengue Vectors, 2004.
[7] T.W. Scott, A.C. Morrison, Vector dynamics and transmission of dengue virus: implications for dengue surveillance and prevention strategies, Dengue Virus (2010) 115–128.
[8] N.L. Achee, J.P. Grieco, H. Vatandoost, G. Seixas, J. Pinto, L. Ching-Ng, et al., Alternative strategies for mosquito-borne arbovirus control, PLoS Negl. Trop. Dis. 13 (1) (2019) e0006822, https://doi.org/10.1371/journal.pntd.0006822.
[9] S. Koyadun, P. Butraporn, P. Kittayapong, Ecologic and sociodemographic risk determinants for dengue transmission in urban areas in Thailand, Interdiscip. Perspect. Infect. Dis. 2012 (2012).
[10] N. Vannavong, R. Seidu, T.-A. Stenström, N. Dada, H.J. Overgaard, Effects of socio-demographic characteristics and household water management on Aedes aegypti production in suburban and rural villages in Laos and Thailand, Parasit. Vectors 10 (1) (2017) 1–14.
[11] S. Chen, A. Whiteman, A. Li, T. Rapp, E. Delmelle, G. Chen, et al., An operational machine learning approach to predict mosquito abundance based on socioeconomic and landscape patterns, Landsc. Ecol. 34 (6) (2019) 1295–1311, https://doi.org/10.1007/s10980-019-00839-2. 10.1007/s10980-019-00839-2.
[12] C. Lorenz, M.C. Castro, P.M. Trindade, M.L. Nogueira, Lage M. de Oliveira, J. A. Quintanilha, et al., Predicting Aedes aegypti infestation using landscape and thermal features, Sci. Rep. 10 (1) (2020) 1–11.
[13] F. Jopp, B. Breckling, H. Reuter, Modelling Complex Ecological Dynamics, Springer, London, 2010.
[14] A. Fielding, Machine Learning Methods for Ecological Applications, Springer Science & Business Media, 1999.
[15] G.R. Humphries, D.R. Magness, F. Huettmann, Machine Learning for Ecology and Sustainable Natural Resource Management, Springer, 2018.
[16] S.V. Burger, Introduction to Machine Learning With R: Rigorous Mathematical Analysis, O'Reilly Media, Inc, 2018.
[17] D. Cianci, N. Hartemink, A. Ibáñez-Justicia, Modelling the potential spatial distribution of mosquito species using three different techniques, Int. J. Health Geogr. 14 (1) (2015) 1–10.
[18] B. Young, J. Yarie, D. Verbyla, F. Huettmann, K. Herrick, F.S. Chapin, Modeling and mapping forest diversity in the boreal forest of interior Alaska, Landsc. Ecol. 32 (2) (2017) 397–413.
[19] M.S. Rahman, A. Karamehic-Muratovic, M. Baghbanzadeh, M. Amrin, S. Zafar, N. N. Rahman, et al., Climate change and dengue fever knowledge, attitudes and practices in Bangladesh: a social media–based cross-sectional survey, Trans. R. Soc. Trop. Med. Hyg. 115 (1) (2020) 85–93, https://doi.org/10.1093/trstmh/traa093.
[20] R. Jain, S. Sontisirikit, S. Iamsirithaworn, H. Prendinger, Prediction of dengue outbreaks based on disease surveillance, meteorological and socio-economic data, BMC Infect. Dis. 19 (1) (2019) 272, https://doi.org/10.1186/s12879-019-3874-x.
[21] J. Santos, B.M. Meneses, An integrated approach for the assessment of the Aedes aegypti and Aedes albopictus global spatial distribution, and determination of the zones susceptible to the development of Zika virus, Acta Trop. 168 (2017) 80–90, https://doi.org/10.1016/j.actatropica.2017.01.015.
[22] L.M. Rueda, Pictorial Keys for the Identification of Mosquitoes (Diptera: Culicidae) Associated With Dengue Virus Transmission, Walter Reed Army Inst Of Research Washington Dc Department Of Entomology, 2004.
[23] M.S. Rahman, T. Ekalaksananan, S. Zafar, P. Poolphol, O. Shipin, U. Haque, et al., Ecological, social, and other environmental determinants of dengue vector abundance in urban and rural areas of northeastern Thailand, Int. J. Environ. Res. Public Health 18 (11) (2021) 5971. https://www.mdpi.com/1660-4601/18/11/5971.
[24] D. Phiri, J. Morgenroth, Developments in Landsat land cover classification methods: a review, Remote Sens. 9 (9) (2017) 967.
[25] J. Demšar, B. Zupan, G. Leban, T. Curk, Orange: from experimental machine learning to interactive data mining, in: Knowledge Discovery in Databases: PKDD 2004, Springer Berlin Heidelberg, Berlin, Heidelberg, 2004, pp. 537–539.
[26] L. Chandrakantha, Risk prediction model for dengue transmission based on climate data: logistic regression approach, Stats 2 (2) (2019) 272–283.
[27] M. Maniruzzaman, M.J. Rahman, B. Ahammed, M.M. Abedin, Classification and prediction of diabetes disease using machine learning paradigm, Health Inform. Sci. Syst. 8 (1) (2020) 7, https://doi.org/10.1007/s13755-019-0095-z.
[28] P. Ganguli, M.J. Reddy, Ensemble prediction of regional droughts using climate inputs and the SVM–copula approach, Hydrol. Process. 28 (19) (2014) 4989–5009.
[29] G.F. Smits, E.M. Jordaan, Improved SVM regression using mixtures of kernels, in: Proceedings of the 2002 International Joint Conference on Neural Networks IJCNN'02 (Cat No 02CH37290), IEEE, 2002, pp. 2785–2790.
[30] S. Tripathi, V. Srinivas, R.S. Nanjundiah, Downscaling of precipitation for climate change scenarios: a support vector machine approach, J. Hydrol. 330 (3–4) (2006) 621–640.
[31] D. Sachindra, K. Ahmed, M.M. Rashid, S. Shahid, B. Perera, Statistical downscaling of precipitation using machine learning techniques, Atmos. Res. 212 (2018) 240–258.
[32] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32.
[33] F. Hairi, C.H. Ong, A. Suhaimi, T.W. Tsung, M.A. Bin Anis Ahmad, C. Sundaraj, et al., A knowledge, attitude and practices (KAP) study on dengue among selected rural communities in the Kuala Kangsar district, Asia Pac. J. Public Health 15 (1) (2003) 37–43, https://doi.org/10.1177/101053950301500107.
[34] S. Selvarajoo, J.W.K. Liew, W. Tan, X.Y. Lim, W.F. Refai, R.A. Zaki, et al., Knowledge, attitude and practice on dengue prevention and dengue seroprevalence in a dengue hotspot in Malaysia: a cross-sectional study, Sci. Rep. 10 (2020) 1: 9534, https://doi.org/10.1038/s41598-020-66212-5.
[35] H. Harapan, Y. Rajamoorthy, S. Anwar, A. Bustamam, A. Radiansyah, P. Angraini, et al., Knowledge, attitude, and practice regarding dengue virus infection among inhabitants of Aceh, Indonesia: a cross-sectional study, BMC Infect. Dis. 18 (2018) 1:96, https://doi.org/10.1186/s12879-018-3006-z.
[36] S. Boonchutima, K. Kachentawa, M. Limpavithayakul, A. Prachansri, Longitudinal study of Thai people media exposure, knowledge, and behavior on dengue fever prevention and control, J. Infect. Public Health 10 (6) (2017) 836–841, https://doi.org/10.1016/j.jiph.2017.01.016.
[37] C.J. Koenraadt, W. Tuiten, R. Sithiprasasna, U. Kijchalao, J.W. Jones, T.W. Scott, Dengue knowledge and practices and their impact on Aedes aegypti populations in Kamphaeng Phet, Thailand, Am. J. Trop. Med. Hygiene 74 (4) (2006) 692–700.
[38] F.A. Diaz-Quijano, R.A. Martínez-Vega, A.J. Rodriguez-Morales, R.A. Rojas-Calero, M.L. Luna-González, R.G. Díaz-Quijano, Association between the level of education and knowledge, attitudes and practices regarding dengue in the Caribbean region of Colombia, BMC Public Health 18 (1) (2018) 143, https://doi.org/10.1186/s12889-018-5055-z.
[39] P. Dhar-Chowdhury, C.E. Haque, R. Lindsay, S. Hossain, Socioeconomic and ecological factors influencing Aedes aegypti prevalence, abundance, and distribution in Dhaka, Bangladesh, Am. J. Trop. Med. Hygiene 94 (6) (2016) 1223–1233, https://doi.org/10.4269/ajtmh.15-0639.
[40] S. Kasem, G.B. Thapa, Crop diversification in Thailand: status, determinants, and effects on income and use of inputs, Land Use Policy 28 (3) (2011) 618–628.
[41] S. Arunyawat, R.P. Shrestha, Assessing land use change and its impact on ecosystem services in Northern Thailand, Sustainability 8 (8) (2016) 768.
[42] H.J. Overgaard, B. Ekbom, W. Suwonkerd, M. Takagi, Effect of landscape structure on anopheline mosquito density and diversity in northern Thailand: implications for malaria transmission and control, Landsc. Ecol. 18 (6) (2003) 605–619.
[43] M.A. Diuk-Wasser, M.B. Toure, G. Dolo, M. Bagayoko, N. Sogoba, S.F. Traore, et al., Vector abundance and malaria transmission in rice-growing villages in Mali, Am. J. Trop. Med. Hygiene 72 (6) (2005) 725–731.
[44] M. Rodrigues Mde, G.R. Marques, L.L. Serpa, B. Arduino Mde, J.C. Voltolini, G. L. Barbosa, et al., Density of Aedes aegypti and Aedes albopictus and its association with number of residents and meteorological variables in the home environment of dengue endemic area, São Paulo, Brazil, Parasites Vectors 8 (2015) 115, https://doi.org/10.1186/s13071-015-0703-y.
[45] Y.-R. Chen, J.-S. Hwang, Y.-J. Guo, Ecology and control of dengue vector mosquitoes in Taiwan, Kaohsiung J. Med. Sci. 10 (1994) S78–S87.
[46] C.M. Romero-Vivas, A.K. Falconar, Investigation of relationships between Aedes aegypti egg, larvae, pupae, and adult density indices where their main breeding sites were located indoors, J. Am. Mosq. Control Assoc. 21 (1) (2005) 15–21.
[47] G. Benelli, F. Maggi, R. Pavela, K. Murugan, M. Govindarajan, B. Vaseeharan, et al., Mosquito control with green nanopesticides: towards the One Health approach? A review of non-target effects, Environ. Sci. Pollut. Res. Int. 25 (11) (2018) 10184–10206, https://doi.org/10.1007/s11356-017-9752-4.
[48] P. Nicolopoulou-Stamati, S. Maipas, C. Kotampasi, P. Stamatis, L. Hens, Chemical pesticides and human health: the urgent need for a new concept in agriculture, Front. Public Health 4 (2016) 148, https://doi.org/10.3389/fpubh.2016.00148.
[49] H.J. Overgaard, W. Suwonkerd, J. Hii, The Malaria Landscape: Mosquitoes, Transmission, Landscape, Insecticide Resistance, and Integrated Control in Thailand. Socio-ecological Dimensions of Infectious Diseases in Southeast Asia, Springer, 2015, pp. 123–153.
[50] S. Morand, J.-P. Dujardin, R. Lefait-Robin, C. Apiwathnasorn, Socio-ecological Dimensions of Infectious Diseases in Southeast Asia, Springer, 2015.
[51] M. Saputra, H. Oktaviannoor, One health approach to dengue haemorrhagic fever control in Indonesia: a systematic review, KnE Life Sci. [Internet] 4 (1) (2018) 201.
[52] S. Sulistyawati, F. Dwi Astuti, S. Rahmah Umniyati, T.B. Tunggul Satoto, L. Lazuardi, M. Nilsson, et al., Dengue vector control through community empowerment: lessons learned from a community-based study in Yogyakarta,

Indonesia, Int. J. Environ. Res. Public Health 16 (2019) 6, https://doi.org/10.3390/ijerph16061013.
[53] N.A.M. Salim, Y.B. Wah, C. Reeves, M. Smith, W.F.W. Yaacob, R.N. Mudin, et al., Prediction of dengue outbreak in Selangor Malaysia using machine learning techniques, Sci. Rep. 11 (1) (2021) 939, https://doi.org/10.1038/s41598-020-79193-2.
[54] RStudio Team, RStudio: Integrated Development Environment for R, RStudio, PBC, Boston, MA, 2021. http://www.rstudio.com/.