IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 30, 2020, accepted July 20, 2020, date of publication August 3, 2020, date of current version October 28, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3013945

Improving Data Quality for Better Control of *Aedes*-Borne Disease Risk

NAIYANA SAHAVECHAPHAN[1], JUKKRAPONG PONHARN[1], ASAMAPORN CHATRATTIKORN[1], PONGSAKORN SADAKORN[2], AND SOPON IAMSIRITHAWORN[2]

[1]National Electronics and Computer Technology Center, Pathumthani 12120, Thailand
[2]Department of Disease Control, Nonthaburi 11000, Thailand

Corresponding author: Naiyana Sahavechaphan (naiyana.sahavechaphan@nectec.or.th)

This work was supported in part by the Department of Disease Control, Ministry of Public Health under Grant P1851568 and in part by the National Electronics and Computer Technology Center (NECTEC), National Science Technology and Development Center (NSTDA), Ministry of High Education, Science, Research and Innovation (MHESI).

ABSTRACT Visual larval survey of container habitats is conducted as a routine mission of the Department of Disease Control (DDC), Thailand. To facilitate this, DDC has deployed a mobile application, namely `TanRabad SURVEY`, throughout the country since 2016. Here, each inspected place with its intrinsic place type and building names must be initially input via natural language to proceed the larval survey data collection. Upon a survey completion, the appropriate larval indices (e.g. House Index (HI), Container Index (CI) and Breteau Index (BI)) are automatically calculated. HI and BI are for villages, while CI is for other inspected places. These larval indices are then applied as factors for the vector control management. However, about 21% of inspected places stored in `TanRabad` database are found with inappropriate place types. These poor place types result in the procurement of inapplicable larval indices and hence ineffective vector control management. Ideally, the quality of place types can potentially be improved once their poorness is notified to users. This paper has thus proposed a novel and comprehensive place type quality assessment technique, namely `pAssessor`, with respect to buildings textually and variously defined for places. Specifically, `pAssessor` is driven by the building-place ontology, building semantic selection, boosted features, learned building-place relations and probability values of all place types. The experimental results showed that the efficiency of `pAssessor` in assessing the quality of place types is greater than 87.5%.

INDEX TERMS *Aedes*-borne diseases, *Aedes* mosquito, Breteau Index, Container Index, data quality, House Index, larval indices, machine learning, ontology, place type, TanRabad, TanRabad SURVEY, visual larval survey.

I. INTRODUCTION

Aedes mosquito is an important vector of arboviruses such as dengue, zika, and chikungunya. It is breeding in a variety of water-holding, artificial, containers that are typically sit on different places including villages,[1] schools, temples, hospitals, hotels and factories. In Thailand, over 50,000 people [1], [2] infected with *Aedes*-borne diseases are annually reported throughout the country. The strategy for prevention and control of *Aedes*-borne diseases relies on timely elimination of *Aedes* mosquito breeding habitats. Visual larval survey of container habitats [3] has thus been conducted as a routine mission of the Department of Disease Control (DDC). This facilitates DDC to detect any increase in vector density. Upon a survey completion, larval indices [4] (House Index (HI), Container Index (CI) and Breteau Index (BI)) are calculated and applied as factors for the vector control management. Depending on the place types, different larval indices are used. HI and BI are for the villages, while CI is applicable for other places. This is mainly due to the fact that each village basically covers many houses in a large spatial area and have various residences with different ages and careers. Conversely, the others are public and sharing places with less number of buildings in a small area and has specific residences.

To facilitate the larval survey, DDC has deployed a mobile application, namely `TanRabad SURVEY` [2], [5],

[1]A village here is a collection of only houses.

The associate editor coordinating the review of this manuscript and approving it for publication was Khalid Aamir.

throughout the country since 2016. It has been implemented based on the visual larval survey guided by WHO [3] and DDC. Essentially, it supports (i) the real-time collection of larval survey data for inspected places; and (ii) the real-time processing of larval indices along with key container habitats. An inspected place with its intrinsic place type and building names (or buildings for short) must be initially input via natural language to proceed the larval survey data collection. Currently, there is an approximation of 11,000 inspected places and 170,000 buildings in `TanRabad` database. Here, according to building semantics, about 21% of places are found with inappropriate place types. This is mainly because users have (i) used the default place type; or (ii) not followed the larval survey guideline. These poor place types result in the procurement of inapplicable larval indices which could in turn increase *Aedes*-borne disease transmission due to an ineffective vector control management and risk communication [6].

Ideally, the quality of place types can potentially be improved once their poorness is notified to users. This study has thus focused on the assessment of place type quality. Today, several approaches [7]–[9] related to place recognition have been proposed in the literature. These approaches are for developing robots to perform simple chores (e.g. cleaning and fetching objects) and hence to improve the quality of life. Technically, they rely on objects detected from vision and/or laser. The detection scores as well as learned object-place relations are then used to perform place classification. These approaches are clearly applicable for an identification of rooms sitting in houses based on well-defined objects and frequently occurring objects. Conversely, in this study, places are villages, schools, temples, hospitals, hotels and factories that encompasses various buildings in a large scale; objects are buildings textually and variously defined; and some rare buildings are helpful for identifying the poorness of place types.

To achieve the place type quality assessment, this study has recognized the significance of natural language processing – "What is the appropriate place type given a set of building names?". This paper has thus proposed a novel and comprehensive place type quality assessment technique, namely `pAssessor`, with respect to building semantics. Specially, `pAssessor` has (i) detected the semantics from building names wherein each consists of one or more word(s); (ii) extracted features from a collection of buildings defined for places; (iii) learnt the building-place relations with respect to features extracted from qualified places in `TanRabad` database; (iv) classified an evaluated place using the learned building-place relations against its extracted feature; and finally (v) assessing the place type quality based on probability values of all place types. In particular, `pAssessor` makes the following contributions:

- Building-Place Semantic Ontology. A well-defined ontology with potential relations is created to enable the semantic detection. The relations include `relates-to`, `is-a`, `is-in`, `is-acronym-of`, `is-typo-of` and `is-eng-of`.
- Building Semantic Selection Algorithm. A comprehensive algorithm is implemented to enable the selection of the most applicable building semantic among different potential semantics as per a single building.
- Feature Extraction. A boosted feature algorithm is developed to promote the visibility of places as buildings along with the number of occurrences of building semantics.

The experimental results showed that the efficiency of `pAssessor` in assessing the quality of place types is greater than 87.5%. The completeness of ontology must be gradually improved to reflect the new conceptual knowledge input into `TanRabad SURVEY`. This enables better building classification, efficient features and finally effective place type quality assessment.

Roadmap: The rest of this paper is organized as followed: Section II gives the basic background. Related works are described in Section III. Section IV presents the `pAssessor`. Experimental evaluations are given in Section V. Section VI gives the discussion and Section VII concludes the paper.

II. BACKGROUND

A. VISUAL LARVAL SURVEY GUIDELINE

DDC has assigned public health officials to regularly conduct the larval survey for places classified as: *village*, *school*, *temple*, *hospital*, *hotel* and *factory*. For an individual village, a random of 40% of its underlying houses must be examined. Conversely, in other places, all buildings must be surveyed. In each individual house or building, the number of indoor and outdoor container habitats with and without larvae must be reported. These container habitats are classified into 12 categories: water tank, water drinking jar, vase, anti-ant bowl, plant saucer, lotus basin, plant leaf, pet bowl, water dispenser, old tire, other used container and unused container.

Once the larval survey of each inspected place is completed, the appropriate larval indices [4] are calculated along with the key container habitats. These indices are (i) House Index (`HI`) – a percentage of houses infested with larvae and/or pupae; (ii) Container Index (`CI`) – a percentage of water-holding containers infested with larvae or pupae; and (iii) Breteau Index (`BI`) – number of positive containers per 100 houses inspected. HI and BI are for the villages, while CI is applicable for other places.

B. VECTOR CONTROL MANAGEMENT

For vector control management, DDC has required that the visual larval survey in all districts at risk must be performed. It should be noted that a district at risk is defined when a number of patients in the past 4 weeks is greater than an average number of patients in such 4 weeks of the previous 5 years. Essentially, at least one place in each particular place type as per a district at risk must be randomly examined. Table 1 and 2 illustrate the vector control policy for villages

TABLE 1. Vector control policy for villages.

Container Index (CI)	Vector Control Policy
HI >50	every week
10 <HI <= 50	every two weeks
HI <= 0	every month

TABLE 2. Vector control policy for other places.

Container Index (CI)	Vector Control Policy
CI >5	every week
0 <CI <= 5	every two weeks
CI = 0	every month

and other places, respectively. Here, if an inspected village is reported with HI over 50, such village must be repeatedly examined on a weekly basis. The examination period would be extended to every two weeks if HI is in between 10 and 50, and every month if HI is below 10. On the other hand, if an inspected school, temple, hospital, hotel or factory is reported with CI over 5, it must be repeatedly examined on a weekly basis. The examination period would be extended to every two weeks if CI is in between 0 and 5, and every month if CI is 0.

In addition, the guidelines and campaigns for vector control are appropriately created with respect to the larval survey results. Essentially, key place types and container habitats are reported to people at risk. Methods for *Aedes* mosquito elimination as per each container habitat is also communicated.

C. TANRAbAD SURVEY

TanRabad SURVEY [2], [5] (see Figure 1) is a mobile application that supports the real-time collection of larval survey data for inspected places; and the real-time processing of larval indices along with key container habitats. Specifically, it stores the place data p and building data b as master data and larval survey data s for each building b under a specified place p as transactional data. The place data p, the building data b and the survey data s are formally defined as following:

Definition 1: A place data p is a 6-tuple $(\iota, \eta, \tau, \phi, \psi, B)$ where ι is a place identification, η a place name, τ a place type, ϕ a place location, ψ a geographic coordinate and B a collection of buildings b.

Definition 2: A building data b is a 3-tuple (ι, η, ψ) where ι is a building identification, η a building name and ψ a geographic coordinate.

Definition 3: A survey data s is a 5-tuple (ι, p, b, δ, C) where ι is a survey identification, p a place, b a building, δ a surveyed date and C a collection of surveyed container habitats c as per each container type.

Definition 4: A surveyed container habitat c is a 5-tuple $(\tau, \alpha_t, \alpha_f, \gamma_t, \gamma_f)$ where τ a container type, α_t the total number of indoor containers inspected, α_f the total number of indoor containers infested with larvae or pupae, γ_t the total number of outdoor containers inspected and γ_f the total number of outdoor containers infested with larvae or pupae.

In addition, a particular building can be either infested or not infested with larvae or pupae. It is formally defined as following:

Definition 5: A building b is said to be infested with larvae or pupae only if there is at least one of its surveyed container habitats infested with larvae or pupae.

Definition 6: A building b is said to be not infested with larvae or pupae only if all surveyed container habitats are not infested with larvae or pupae.

The appropriate larval indices are calculated with respect to a place type $(p.\tau)$ and formally defined as per Equations 1- 3.

$$HI = \frac{|B_f|}{|B|} \times 100 \tag{1}$$

$$CI = \frac{\sum_{b\epsilon B} \sum_{h\epsilon H} (\alpha_f^h + \gamma_f^h)}{\sum_{b\epsilon B} \sum_{h\epsilon H} (\alpha_t^h + \gamma_t^h)} \times 100 \tag{2}$$

$$BI = \frac{\sum_{b\epsilon B} \sum_{h\epsilon H} (\alpha_f^h + \gamma_f^h)}{|B|} \times 100 \tag{3}$$

where $|B|$ is the number of buildings inspected, $|B_f|$ the number of buildings infested, b a building in B, H a collection of container habitat types, h a container type in H, α_t^h and γ_t^h the total number of indoor and outdoor containers with type h inspected as well as α_f^h and γ_f^h the number of indoor and outdoor containers with type h infested.

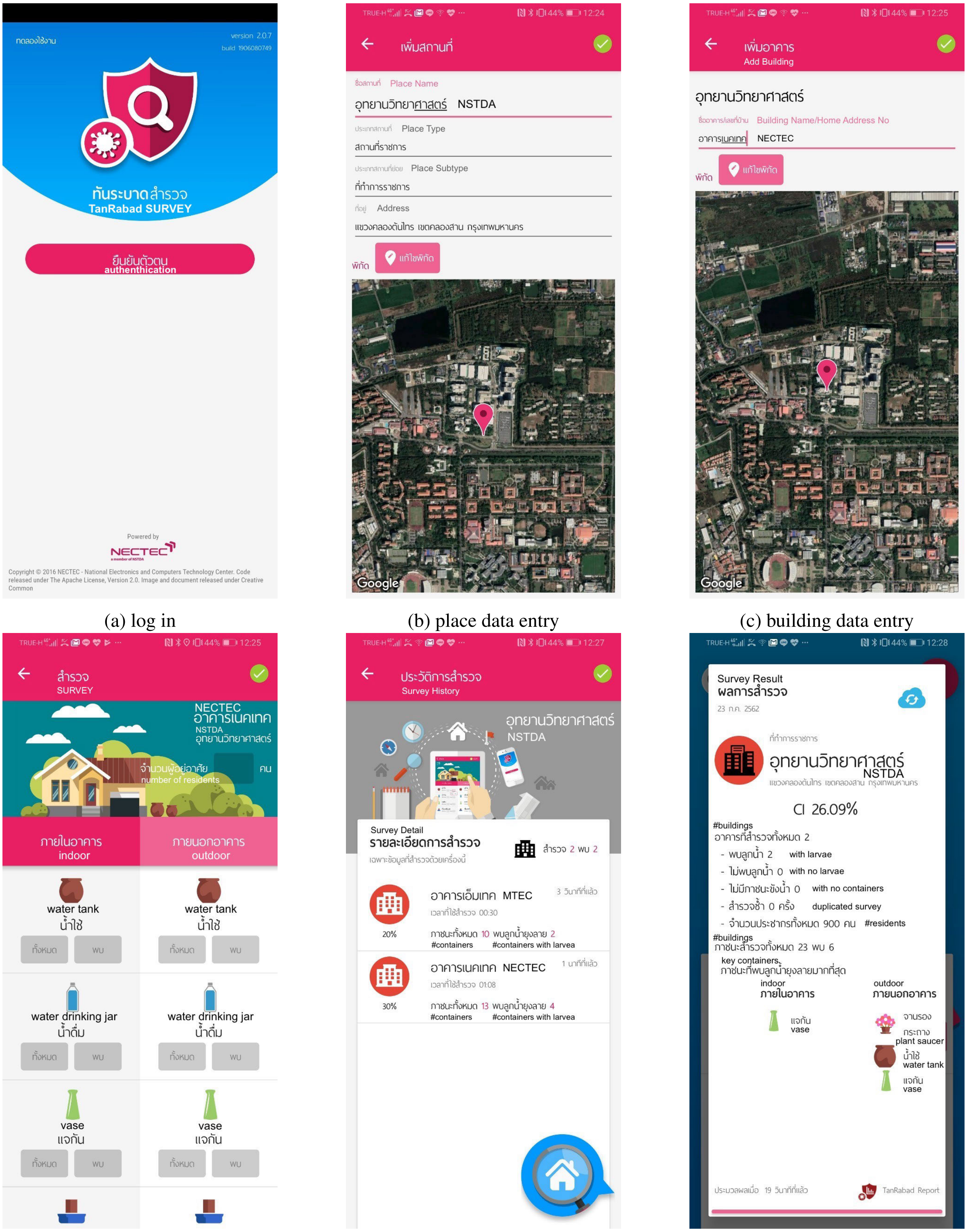
According to the previously defined place types (see Section II-A), the underlying building names of a place can potentially be used to identify the place type. As examples, a village has various houses specified via their address numbers or owner names, while a temple has cubicles and sermon halls.

D. ADVERSE EFFECT OF POOR DATA

"Garbage in – Garbage out" is a colloquial recognition of poor quality data entry that leads to unreliable data output [10]. In `TanRabad SURVEY`, about 21% of places are specified with inappropriate place types. This is mainly because users have (i) used the default place type. As an example, a place "Wat Sri Sudaram" with 2 buildings "cubicle" and "sermon hall" is defined as village by default. In fact, a temple type must be selected; or (ii) not followed the larval survey guideline defined in Section II-A. As an example, a place "Moo 3" with village type consists of "Wat Bang Pra", "Wat Bang Pra School", and several address numbers which represent temple, school and houses, respectively. In fact, this ambiguous place must be partitioned into 3 places: "Wat Bang Pra", "Wat Bang Pra School" and "Moo 3 village" with temple, school and village types, respectively. Essentially, these poor place types result in the procurement of inapplicable larval indices and hence ineffective vector control management.

III. RELATED WORKS

Several approaches [7], [8] with respect to place recognition techniques have been proposed in the literature. Most are for



(a) log in (b) place data entry (c) building data entry

(d) container data entry (e) container index (f) overall report

FIGURE 1. Screenshot of TanRabad SURVEY. Source of figures: TanRabad software suite https://www.tanrabad.org.

developing robots to perform simple chores (e.g. cleaning and fetching objects) and hence to improve the quality of life. Specifically, these approaches rely on objects detected from vision alone or both vision and laser. The detection scores as well as learned object-place relations are then used to perform place classification.

Viswanathan *et al.* [7] have proposed a system that integrates object detection and place classification. Specifically, they performed an automated learning of object-place relations from images in the `LabelMe` database, a free online data source that provides a large and growing amount of human-labeled visual data. They then trained object detectors on some of the most frequently occurring objects. Finally, they used detection scores as well as learned object-place relations to perform place classification. Their future work could involve language processing to eliminate ambiguous labels as well as combine synonymous labels together.

Rottmann *et al.* [8] and Mozos [9] has addressed the problem of semantic classification of the environment using range finder and vision features. Their approach has used the AdaBoost algorithm to boost simple features extracted from laser and vision data, which on their own are insufficient for a reliable categorization of places, to a strong classification. To reduce the number of outliers during the classification, Hidden Markov Model (HMM) is applied to filter the current classification result based on previously calculated labels.

The above approaches are clearly application for an identification of rooms sitting in houses based on well-defined objects and frequently occurring objects. Conversely, in this study, places are villages, schools, temples, hospitals, hotels and factories that encompasses various buildings in a large scale; objects are buildings textually and variously defined; and some rare buildings are helpful for identifying the poorness of place types.

IV. PLACE TYPE QUALITY ASSESSMENT

This section presents the place type quality assessment technique, namely `pAssessor`, which relies on ontology-based place recognition. In particular, `pAssessor` takes a set of building names specified for a place as input and produces as output the place type quality assessment. Figure 2 illustrates an architectural overview of `pAssessor` and highlights its four key components: *Building Semantic Detection*, *Feature Extraction*, *Place Recognition* and *Place Assessment*. Their details are given in the following sections.

A. BUILDING SEMANTIC DETECTION

Recall that a place can be described by the buildings it encompasses. The exploration of all building names input via `TanRabad SURVEY` [2], [5] is thus needed. Here, various names are textually specified. Some may use different words or abbreviations (with correct spelling and typographical error) to represent the same meaning. In this study, for the 6 place types (see Section II-A), their building semantics can be classified as followings:

- Address Semantic – to represent address numbers. Examples include 121 and 121/29.
- CoSpace Semantic – to represent spaces typically existing in any places. Examples are restroom, canteen and parking lots.
- Education Semantic – to represent education-related building. Examples are school building, library, museum, cultural centre and nursery.

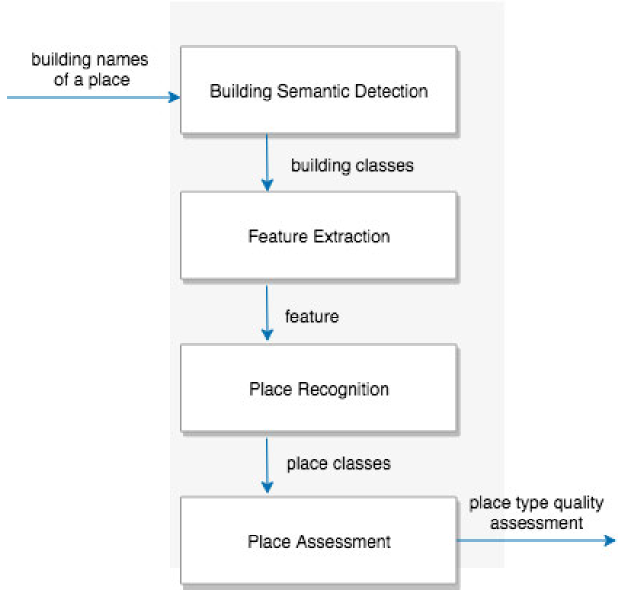


FIGURE 2. **The** `pAssessor` **architecture.**

```
[0-9๐-๙]+/*[0-9๐-๙]*(ม)*(ท)*(ข)*(ต)*(อ)*(\.)*[0-9๐-๙]
```

FIGURE 3. **Regular expressi to detect address number pattern.**

- Factory Semantic – to represent manufacturing-related building. Examples are publisher, printing house and garage.
- Habitat Semantic – to represent residences either for a small number of people or long stay. Examples are house, room, dormitory, apartment and even pet house.
- Health Semantic – to represent health- and medical-related building. Example are patients, medical operations, x-ray, emergency ward, drugs, ICU, CCU and physical therapy.
- Hotel Semantic – to represent residences either for a lot of people or short stay. Examples are hostel, hotel, inn, guesthouse, boutique and resort.
- Person Semantic – to represent person names. Examples are Mr.Robert and Ms.Anne.
- Religious Semantic – to represent religion-related building. Examples are cubicles, sermon halls, church, sanctuary and crematorium.
- OtherWorkSpace Semantic – to represent other working spaces. Examples are store, market, minimart, office and company.

The address semantic can simply be detected using regular expression [11], a sequence of characters that define a search pattern. The regular expression for validating the address number pattern such as "121/29" and "address no 12" is thus defined as depicted in Figure 3. The rest semantics, on the other hand, require ontology [12] as it is capable of capturing semantic knowledge via the conceptualization structure. However, existing ontologies are typically developed for the domains of interest such as herbal medicine [13], flora [14] and rice disease [15]. None of

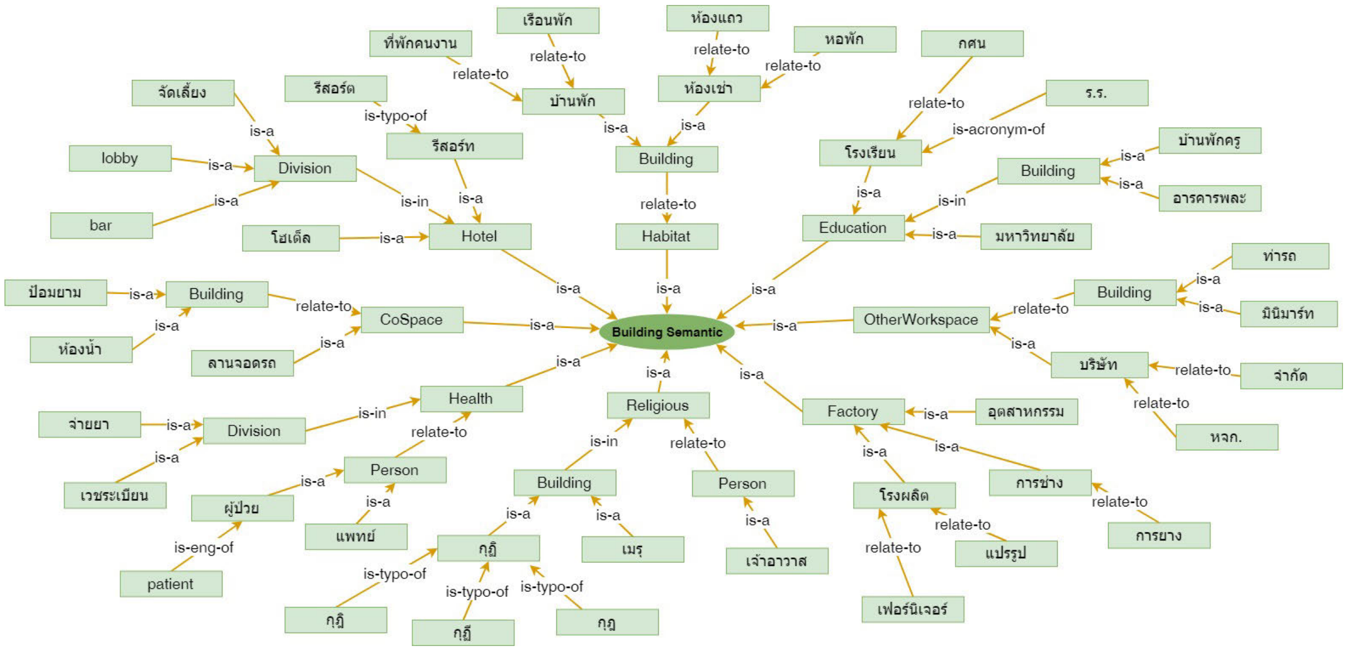


FIGURE 4. **The partial building-place semantic ontology.**

ontologies clearly exists for the building semantics previously defined as above. Such ontology is thus developed as partially illustrated in Figure 4. Specifically, the root `BuildingSemantics` node has the previously defined semantics (except the address semantic) with the relation *is-a*. Each individual semantic node has its conceptualization structure with relations: `relates-to`, `is-a`, `is-in`, `is-acronym-of`, `is-typo-of` and `is-eng-of`.

Based on the developed regular expression and ontology, the overall process for detecting building semantics is depicted in Figure 5. In particular, it takes a building name *bname* as input and returns as output the building semantic *bsemantic*. As per a building name *bname*, the first step is to remove stop words (e.g. around, front and back) which are meaningless and to perform word correction, resulting in $bname'$. The second step is to detect the address pattern based on the regular expression. If the address pattern is detected, the "address" semantic is reported and the process then stops. Otherwise, the ontology-based semantic matching has come into play at the third step. Here, there are 9 semantic matching submodules. Each of which corresponds to an individual building semantic as discussed above. Essentially, each module performs the matching of its knowledge represented in form of ontology (see Figure 4) against the text in $bname'$. Its output can be either the matching result or none if there is a match or no match, respectively. As per a building name $bname'$, different matches can be identified by one or more matching submodules. A collection of matching results, denoted as R, is thus considered as the output of the ontology-based semantic matching. A matching result $r \in R$ is formally defined as below.

Definition 7: A matching result r is a 3-tuple (w, i, s) where w is a substring of a building name matching with the

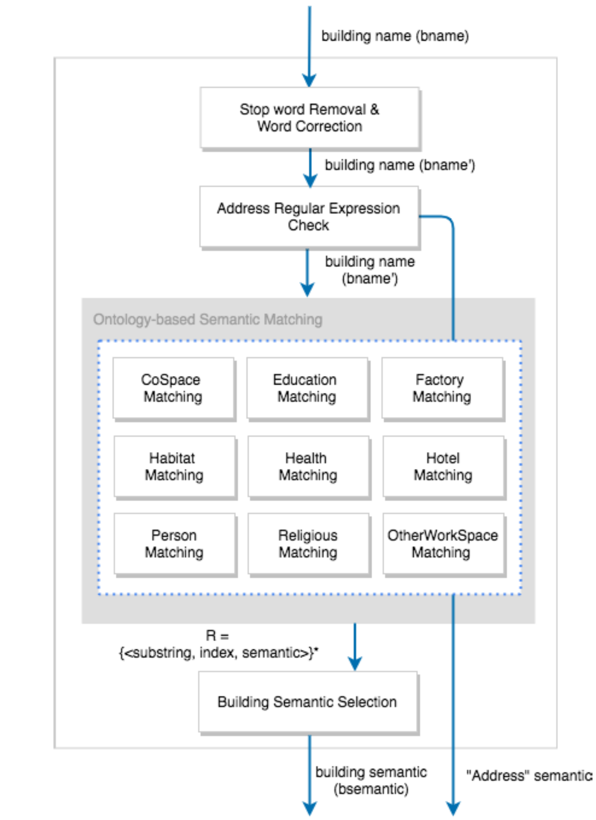


FIGURE 5. **Building semantic detection process.**

knowledge, i an index of such substring and c a representative semantic of the knowledge.

Finally, the building semantic selection next takes a collection of matching results R and a building name $bname'$ as

```
procedure BUILDINGSEMANTICSELECTION(R, bname')
    if |R| = 0 then return "Others";
    if |R| = 1 then return r.semantic where r ∈ R;
    if there exists r_i, r_j ∈ R where r_i ≠ r_j and r_i.index = r_j.index then
        lm_i ← longest_match(r_i.substring, bname');
        lm_j ← longest_match(r_j.substring, bname');
        if lm_i > lm_j then
            R = R − r_j
        else
            R = R − r_i
    if |R| = 1 then return r.semantic where r ∈ R;
    if there exists r_i ∈ R where r_i.semantic = 'Person' then
        if r_i.index ≠ 0 then
            R = R − r_i
    if |R| = 1 then return r.semantic where r ∈ R;
    if there exists r_i, r_j ∈ R where
     r_i.semantic ∉ { "CoSpace", "OtherWorkSpace" } and
     r_j.semantic ∈ { "CoSpace", "OtherWorkSpace" } then
        R = R − r_j
    if |R| = 1 then return r.semantic where r ∈ R;
    r ← find_the_first_index(R);
    return r.semantic
```

FIGURE 6. BuildingSemanticSelection algorithm.

input and returns as output the appropriate building semantic. Its algorithm is depicted in Figure 6. Here, the "other" semantic is returned if R is an empty set. If R contains only one member $r \in R$, the semantic of r is then returned. Otherwise, all members $r \in R$ must be examined. The inapplicable members $r \in R$ are gradually filtered out along the process. In particular, if any two members r_i, $r_j \in R$ have an identical index, the one having its longest common subsequence [16] with $bname'$ the most will be selected. Next, a member $r_i \in R$ whose semantic is person and index is not at the 0^{th} position is pruned. Finally, if there exists a member $r_i \in R$ whose semantic is neither `CoSpace` nor `OtherWorkSpace` but any members $r_j \in R$ with either `CoSpace` or `OtherWorkSpace` semantic, a member r_i will be chosen. This pruning process may result in two or more candidate members. The one with the least index is finally chosen and hence its semantic is returned as output. Samples are shown in Figure 7.

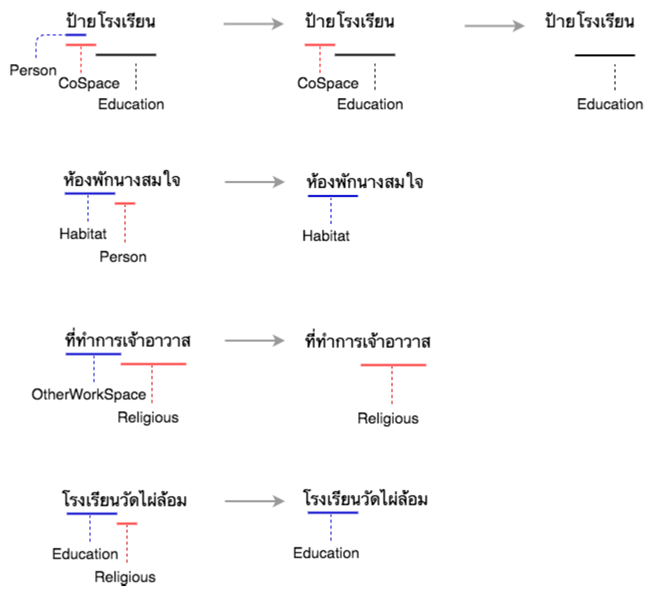


FIGURE 7. Building semantic detection samples.

B. FEATURE EXTRACTION

Figure 8 shows the normalized proportions of building semantics detected for roughly 7,100 qualified places classified by place types. Here, each place is likely to have building names with several semantics in which their most semantics correspond to its intrinsic place type. For the reliable place recognition, it is desirable to extract appropriate feature space based on the number of occurrences of building semantics detected as per a place p. However, places are likely to be defined at the building level, not conforming to the definition (see Section II-C). Such places should thus be boosted up to reflect their visibility. The feature as per a place p, denoted as F, is formally defined as:

Definition 8: The place feature F is a 10-tuple (η_a, η_c, η_e, η_f, η_{ha}, η_{he}, η_{ho}, η_p, η_r, η_o) where η_a is the number of (boosted) occurrences of Address semantic detected for a

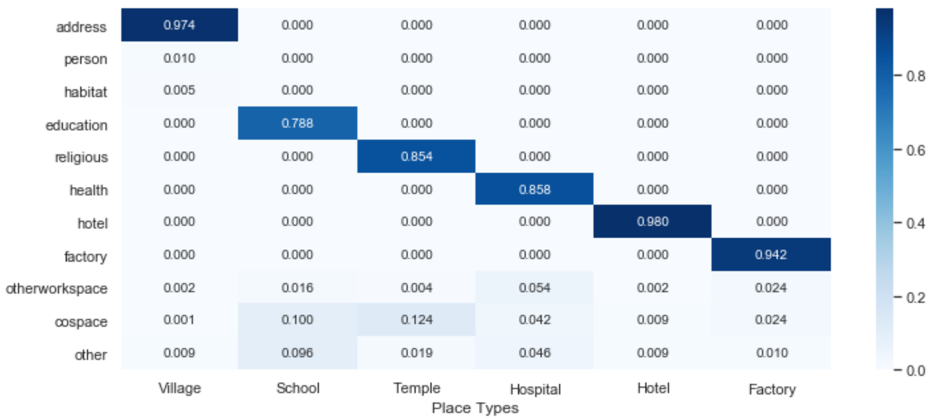


FIGURE 8. **The normalized proportions of building semantics for 6 place types.**

place p. Similarly, η_c, η_e, η_f, η_{ha}, η_{he}, η_{ho}, η_p, η_r and η_o are for CoSpace, Education, Factory, Habitat, Health, Hotel, Person, Religious and OtherWorkSpace semantics, respectively.

Essentially, the number of (boosted) occurrences of a specified building semantic χ, denoted as η_χ, is computed as per Equation (4) - (5).

$$\eta_\chi = \sum_{b \in B} \mathbb{M}(b.n, \chi) \tag{4}$$

$$\mathbb{M}(b.n, \chi) = \begin{cases} 1, & \text{if } \mathbb{SC}(b.n) = \chi \text{ and } \mathbb{T}(b.n) = \text{``building''} \\ \omega, & \text{if } \mathbb{SC}(b.n) = \chi \text{ and } \mathbb{T}(b.n) = \text{``place''} \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

where B is a set of buildings defined for a place p, b a building in B, $b.n$ the name of a building b, χ the preferable building semantic, $\mathbb{M}$ a mapping function defined in Equation (5), $\mathbb{SC}$ the building semantic detector described in Section IV-A, $\mathbb{T}$ the tier of $b.n$ either a place or a building and ω the boosted value (> 1). The mapping function $\mathbb{M}$ is to evaluate the correspondence between the preferable building semantic χ and the semantic of $b.n$ returning from the building semantic detector $\mathbb{SC}$ along with the tier of $b.n$.

As an example, Table 3 is the result of feature extraction from building names specified for 14 places illustrated in Figure 9.

C. PLACE RECOGNITION

Various classification algorithms have been developed. Examples include logistic regression [17], random forest [18], support vector machine [19] and naive bayes [20]. To select the best applicable algorithm, their capabilities in learning the defined features must be evaluated. Based on the binary classification evaluation (see Section V-B), all classification algorithms equivalently performed. In this paper, the logistic regression is chosen as it is simple and easy to implement. Given a place type t, its logistic model based on the feature defined in Section IV-B is shown in Equations 6 - 7.

$$p^t = \frac{1}{1 + b^{-\beta^t}} \tag{6}$$

$$\beta^t = \beta_0^t + \sum_{f \in F} \beta_f^t \eta_f \tag{7}$$

where p^t is the probability of a place to be the type t, β_0^t the intercept of place type t, β_f^t the coefficient of a predictor η_f for the place type t, F a set of building semantics – CoSpace, Education, Factory, Habitat, Health, Hotel, Person, Religious and OtherWorkSpace semantics and η_f the number of occurrences of building semantic f.

D. PLACE ASSESSMENT

In this study, the place type quality is classified as (i) *low* – in that none of building semantics is helpful for place classification. As an example, a place consists of buildings "male restrooms" and "female restrooms" which are common for all place types; (ii) *ambiguous* – in that two or more candidate place types are found. As an example, a place consists of "Wat Bang Pra" and "Wat Bang Pra School" which are temple and school; and (iii) *high* – in that an exact place type is detected. As an example, a place with 2 buildings "cubicle" and "sermon hall" is predicted as temple alone. It is thus necessary to assess the applicability of all place types as per an evaluated place. In particular, the quality of place p to be type τ, denoted as $\mathbb{Q}(p, \tau)$, is formally defined as per Equations (8) - (9).

$$\mathbb{Q}(p, \tau) = \begin{cases} high, & \text{if } \sum_{t \in PT} \mathbb{T}(p^t(p, \tau)) = 1 \\ low, & \text{if } \sum_{t \in PT} \mathbb{T}(p^t(p, \tau)) = 0 \\ ambiguous, & \text{otherwise} \end{cases} \tag{8}$$

$$\mathbb{T}(p^t(p, \tau)) = \begin{cases} 1, & \text{if } p^t(p, \tau) > \lambda \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

where PT is a set of place types, $\mathbb{T}$ the function transforming a probability of a place being a type $t \in PT$ to either 1 or 0, p^t

Place Id	Place Names	Building Names
1	ชุมชนหมู่3	2/1, 15, 31/2, 3, 14/2, 1/8, 8/1, 42, 30, 33/8, 37, 1/19, 6, 1/16, 1/18, 15/1, 9, 14/1, 30/3, 2/2, 14, บ้านเลขที่2/1, บ้านเลขที่2/4, บ้านเลขที่1/11, บ้านเลขที่1/17, 25, วัดบางพระ, 26/1, 1/14, 10, 3/1, 22/1, รร.วัดบางพระ
2	บ้านห้วยเอียน	89, 84, 4, 58, 12, 17, 53/1, ผช.รัตน์, 18, 15, นายรัตน์, 25, 64, นายไชยรัตน์, 75, 81, นางตู, 41, นางพจมาน, 72, 45 ,นายผ่อง, 53, 13, 7, 80 ,ผช.ก็อต, 1, 54, 56, 59, 67, นายแดง, 100, 49, 50, 55, 74, 92, 86, 114
3	บ้านหนองแฟบ ม.8	26/1, 58, ประเทือง, 29/1, 85ก, นางผาด, 157, บ้านสวนยูคา2, 42/1, บ้านร้าง, 53, ตรงข้ามเจริญร้านค้า, 100, 52, บ้านสวนยูคา, 67, 85, 82, นางสมร, 11/4, เจริญร้านค้า, บ้านหลังร้านค้า, 26/4, 27, 70, ไพวิล
4	วัดจันตะคาม ม 7	กุฏิ1, กุฏิ2, กุฏิ3, กุฏิ4, กุฏิ5, กุฏิ6, กุฏิ7, กุฏิ8, ศาลา, โรงครัว
5	วัดพะเนียงแตก	หอฉัน, ศาลาฉัน, อาคาร108ปี, หอประชุมพระ, ศาลาเมรุ, กุฏิพระ2, กุฏิพระ3, กุฏิพระ1, ศาลาอเนกประสงค์, อาคารปฏิบัติธรรมจารุโภ
6	สำนักแม่ชีไทย	กุฏิแม่ชี1, กุฏิแม่ชี2, กุฏิแม่ชี3, กุฏิแม่ชี4, กุฏิแม่ชี5, กุฏิแม่ชี6, กุฏิแม่ชี7, กุฏิแม่ชี8, กุฏิแม่ชี9, กุฏิแม่ชี10, กุฏอแม่ชี11, กุฏิแม่ชี12, กุฏิแม่ชี13, อาคารศรัทธานุสรณ์สนั่น จานุวรรโน, กุฏิแม่ชีด้านหน้า, สมเด็จพระญาณวโรดม, อาคารผู้ปฏิบัติธรรมด้านหน้า, อาคารพักผู้มาปฏิบัติธรรมด้านหลัง
7	โรงเรียนบ้านฉาง	อาคารอำนวยการ, ผักไร้ดิน, ตึก ป.6, อาคาร2, ห้องน้ำ ป.5, ห้องน้ำ ป.4, อาคารเรียนไม้, ห้องน้ำ ป.6, ห้องน้ำอาคารอเนกประสงค์, อาคารป.1 ป.2, อาคารอนุบาล
8	โรงเรียนวัดไผ่หูช้าง	อาคารเรียน3, อาคาร 1, อาคาร 2, อาคาร ICT, บ้านครูฝน, โรงอาหาร, บ้านพักครูธุรการ, บ้านพักครูน้ำ, บ้านพักครูต้อม, บ้สนพักครูกฒติ
9	โรงเรียนวัดบางหลวง	หอประชุม, อาคาร2, อาคาร3, อาคาร4, อาคาร 1 พิพิธภัณฑ์, อาคาร 2, อาคาร 3, ศูนย์การเรียนรู้งานอาชีพ, อาคารเฉลิมพระเกียรติ, อาคารเตโชทัย, อาคารศูนย์เรียน, อาคารอำนวยการ, อาคาเอนกประสงค์
10	โรงพยาบาลเมืองพัทยา	สำนักงานแพทย์, อาคาร 1, แผนกรวมเด็ก, ศูนย์แลปฯ, check up, ศาลพระภูมิ, กายภาพ, ศูนย์ไตเทียม, ward พิเศษชั้น4, จุดรวมพล, wardรวมชาย, wardรวมหญิง, แผนกแม่และเด็ก, ทันตกรรม, ศูนย์อาหาร, OPD, แพทย์แผนไทย, อาคารจอดรถ
11	โรงพยาบาลจิตเวช	ตึกเฟื่องฟ้า, OPD, ตึกศรีตรัง, ราชพฤกษ์, ฝ้ายคำ, ชงโค, ศูนย์บริการรักษาด้วยไฟฟ้า, ตัดสินใจมา, พัสดุ, เวชภัณฑ์, การแพทย์ทางเลือก, ศูนย์ศึกษาและวิจัย
12	โรงพยาบาล บ้านม่วง	โรงพยาบาล บ้านม่วง, อาคารโรงบาลบ้านม่วง, ห้องประชุม, ตึกจ่ายกลาง, บ้านพัก, ห้องประชุมลีลาวดี, ตึกสงฆ์อาพาธ, หมวดยานพาหนะ, ห้องผู้ป่วยหญิง, บ้านพัก1, บ้านพัก2
13	โรงงานคาสเตย์คอร์โปเรชั่นจำกัด	อาคาร รง.1, อาคาร รง.2, อาคาร รง.3, อาคาร รง.4, อาคาร รง.6, อาคาร รง.7, อาคาร รง.8, อาคาร รง.9, อาคาร รง.10, ป้อมตำรวจหน้า รง., สำนักงานoffice
14	โรงงาน NC coconut	อาคารแปรรูป, อาคารแคนทีน
15	โรงแรมพาราโซ	จุดทิ้งขยะ, เพาะชำ, lobby, ร้านค้า, bar, ศาลพระภูมิ
16	โรสรีสอร์ท	อาคารสำนักงาน, อาคาพักหลังที่1, ห้องพักที่2, ห้องพักที่3, ห้องพักที่4, ห้องพุกที่5, ห้องพักหลังที่, ห้องพักหลังที่6, ห้องพักหลังที่7, ห้องพักหลังที่8, ห้องพักหลังที่10, บ้านพักหลังที่11, ห้องพักหลังที่ 12

FIGURE 9. The place samples.

TABLE 3. Features of places in figure 9.

Building Semantic	Place Id															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Address	31	33	16													
CoApace				2	2		3	4	10	5	2	2		1	2	
Education	1						7	6	3		1					
Factory													11	1		
Habitat			4									3				
Health										13	4	5				
Hotel															2	12
Person		8	2													
Religious	1			8	8	18										
OtherWorkSpace			2								1	1			2	1
Others			2				1				4					

the probability of a place p to be the type t and λ the specified threshold.

V. EXPERIMENTAL EVALUATION

The `pAssessor` has been developed via Python script language. It is driven by the larval survey data and building-place ontology privately stored in PostgreSQL database. To evaluate the potential benefits of `pAssessor`, a series of experiments were conducted to test the following:

- The effectiveness of building semantic detection
- The capability of classification algorithms
- The efficiency of place type quality assessment
- The applicability of boosted features

A. EFFECTIVENESS OF BUILDING SEMANTIC DETECTION

The first set of experiments measured the effectiveness of building semantic detection. Here, a collection of building names, both with correct spelling and typographical error,

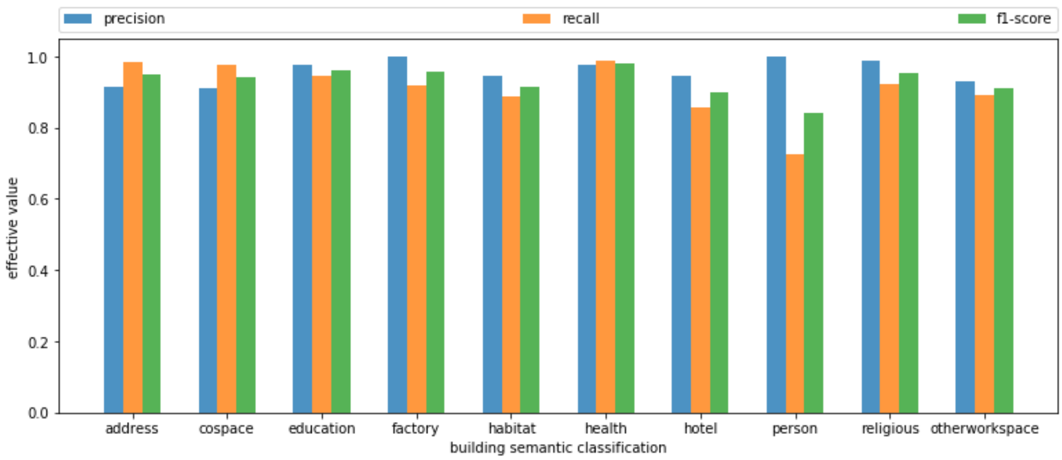


FIGURE 10. The effectiveness in detecting semantics of building names with correct spelling.

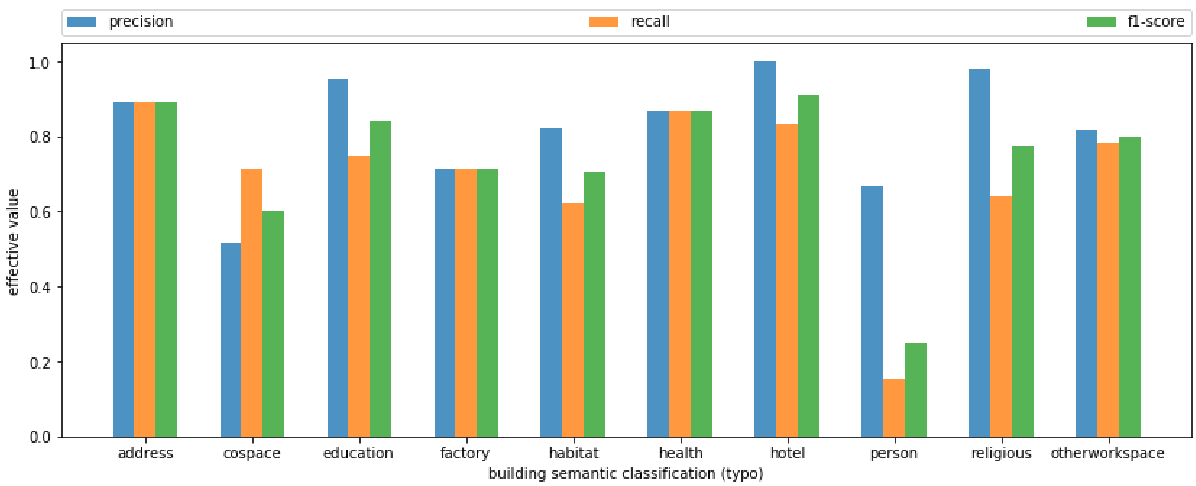


FIGURE 11. The effectiveness in detecting semantics of building names with typographical error.

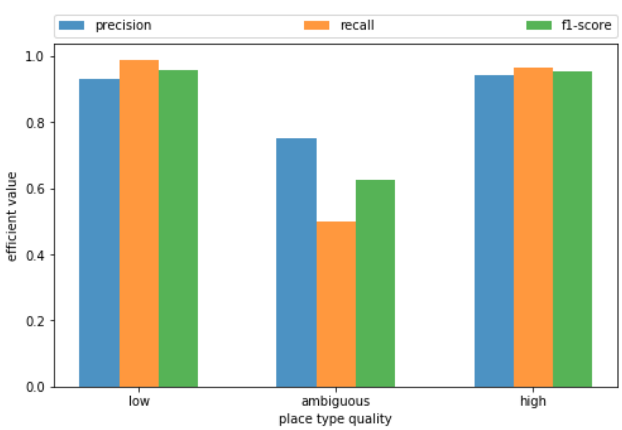


FIGURE 12. The efficiency of data quality assessment with boosted features.

TABLE 4. The number of buildings manually classified and validated by experts.

Building Semantics	Building Names	
	Correct Spelling	Typographical Error
address	66	28
cospace	42	21
education	91	28
factory	49	7
habitat	79	37
health	84	23
hotel	21	12
person	58	13
religious	92	81
otherworkspace	73	23
total	655	273

was randomly selected from `TanRabad` database. Their semantics were then labeled by experts. The number of buildings manually classified and validated by experts is depicted in Table 4. Next, the building semantic detection module defined in `pAssessor` took this data set as input and

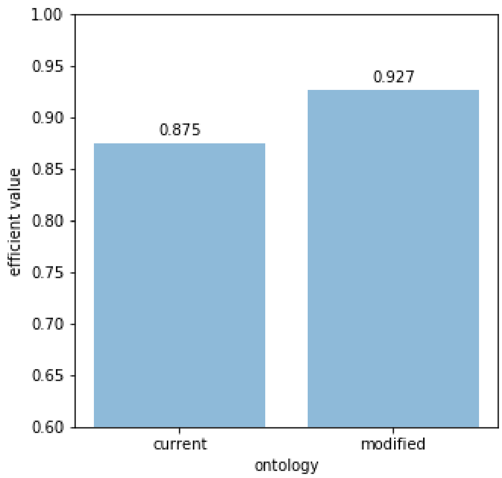


FIGURE 13. The overall efficiency of data quality assessment.

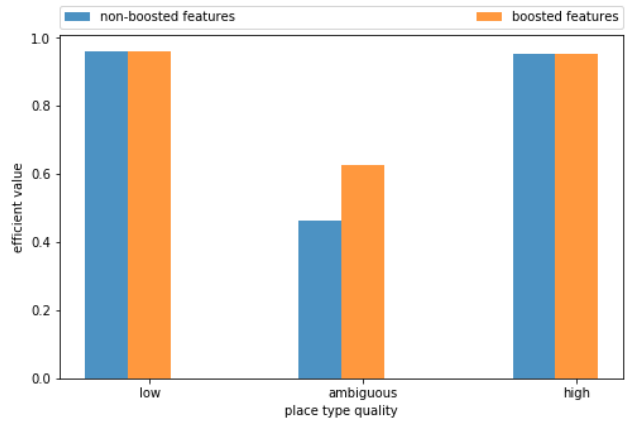


FIGURE 14. The efficiency of data quality assessment with boosted against non-boosted features.

returned as output their corresponding semantics. Finally, the results of the detection module was compared against what experts had labeled.

Figure 10 shows the effectiveness for detecting semantics of building names with correct spelling. The x-axis has building semantic classification and the y-axis is the effective value with 1.0 representing 100%. Here, all semantics except the person semantic were well detected with f-score ranging from 0.90 to 0.98. The person semantic was poorly detected with f-score of 0.84. This is mainly due to the fact that some person names have no prefixation and hence no clue for identifying the person names themselves.

Figure 11 shows effectiveness for detecting semantics of building names with typographical errors. The x-axis has building semantic classification and the y-axis is the effective value with 1.0 representing 100%. Here, the overall detection of building semantics was not well performed. This is because there are various typographical errors wherein some of them (i) have not yet been defined in the ontology; or (ii) could not be applicable for word correction.

B. CAPABILITY OF CLASSIFICATION ALGORITHMS

The second set of experiments evaluated the capability of existing classification algorithms in learning the proposed features. Logistic regression [17], random forest [18], support vector machine [19] and naive bayes [20] were chosen for such evaluation. Here, a set of place data from `TanRabad` database was selected. The poor data set was next filtered out, resulting in a cleaned data set. The experts were then asked to validate this cleaned data set and manually label their place types. The feature extraction module next took each place data in the cleaned data set as input and produced as output its corresponding features. Finally, the selected

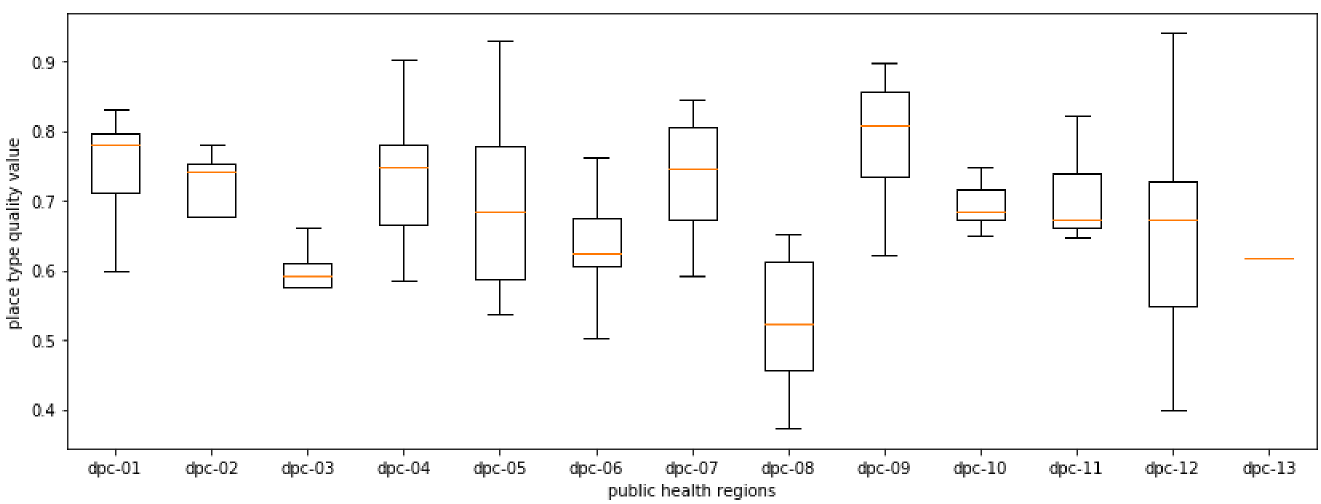


FIGURE 15. The place type quality by public health regions.

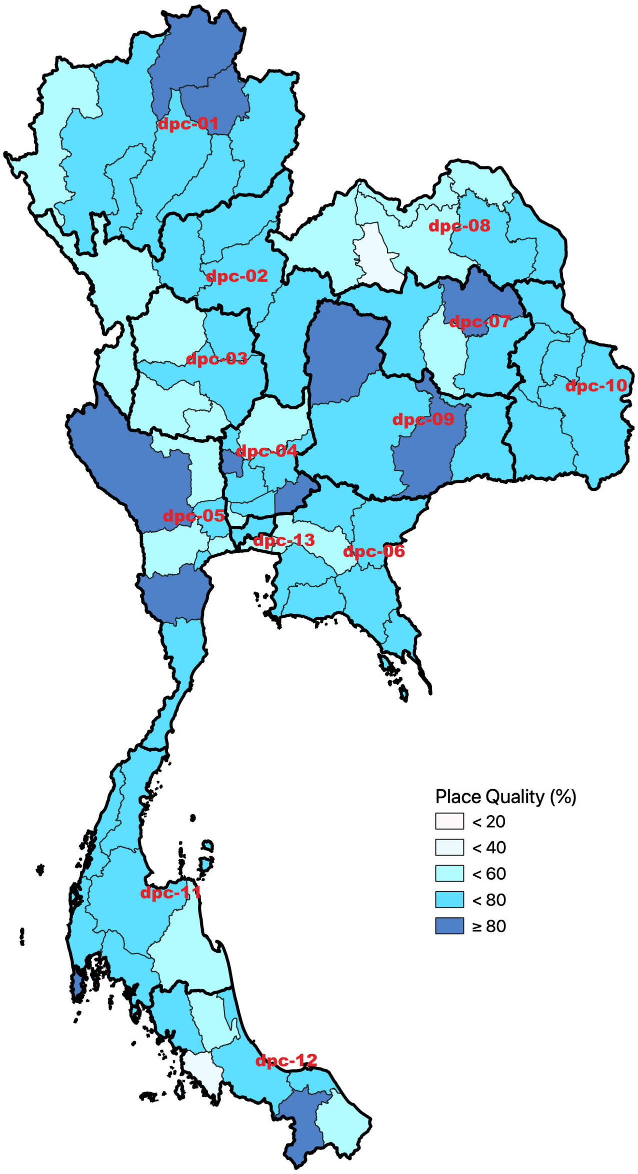


FIGURE 16. The place type quality at provincial level.

classification algorithms were trained upon features extracted from the cleaned data set. In particular, the binary classification with the balanced data set and cross validation of 10 were performed. Interestingly, all classification algorithms equivalently performed.

C. EFFICIENCY OF PLACE TYPE QUALITY ASSESSMENT

The third set of experiments assessed the quality of unseen data. Here, the unseen data set (287 place data records) from `TanRabad` database was randomly selected. The experts were then asked to label the quality of place types: *low*, *ambiguous* and *high*. The `pAssessor` system next ran this unseen data set using the boosted value ω of 5 (see Equation (5)) and the threshold λ of 0.75 (see Equation (9)). Finally, the results of `pAssessor` was compared against what experts had labeled.

Figure 12 depicts the efficiency of place type quality assessment via boosted features. The x-axis has the place type quality and the y-axis is the efficient value with 1.0 representing 100%. Here, the assessment of high quality performed the best with 90% accuracy. The low quality assessment was next performed with 85% accuracy. Unfortunately, the ambiguous quality was poorly evaluated. This is mainly due to the fact that two more more building semantics with too small number of their occurrences were detected as per a place.

In addition, the current and modified versions of ontologies were accompanied – the more coverage ontology, the more efficiency of overall assessment. Figure 13 depicts the overall efficiency of place type quality assessment. The x-axis has the versions of ontology and the y-axis is efficient value with 1.0 representing 100%. Here, the efficient value had an improvement from 87.5% to 92.7% when accompanying the modified version of ontology. Expectedly, this is because the modified version has captured better conceptual knowledge.

D. APPLICABILITY OF BOOSTED FEATURES

The forth set of experiments evaluated the applicability of boosted features. Here, this experiment relied on the previous unseen data set labeled by experts. The `pAssessor` system separately ran this unseen data set using boosted and non-boosted features. Essentially, the non-boosted features reflect the pure number of occurrences of objects employed by existing place recognitions. The results via boosted features were compared against the return using non-boosted features.

Figure 14 depicts the efficiency of place type quality assessment via boosted features against non-boosted features. The x-axis has the place type quality and the y-axis is the efficient value with 1.0 representing 100%. Here, the assessment of high and low quality using boosted and non-boosted features equivalently performed. The ambiguous quality was poorly evaluated. However, the boosted features was capable of improving the ambiguous assessment performance, from 0.46% to 0.63%.

VI. DISCUSSION

There are 13 public health regions (or dpc) under DDC. Each of which is responsible for the disease prevention and control in the assigned provinces. As for the *Aedes*-borne diseases, each region has employed `TanRabad SURVEY` for the visual larval survey data collection. Beside this, each region has conducted a local training of `TanRabad SURVEY` to provincial public health offices underneath for the procurement of rich larval indices. Due to the engagement of various users and the poor results rendered via `TanRabad` system, DDC and these regions have now recognized the significance of data quality.

Accordingly, the `pAssessor` was run with all place data set from `TanRabad` database. Places whose types were assessed with high quality were selected. Their place types were then compared against the user-defined place types. Their matches were finally measured to reflect the quality of user-defined place types. Figure 15 illustrates the place type quality classified by 13 public health regions. The x-axis has the public health regions and the y-axis is place type quality value with 1.0 representing 100%. Here, the region 9 performed the best with an average above 80%, while the rest regions poorly performed with an average ranging from 52% to 79%. Clearly, all regions must regularly review the larval survey guideline and conduct an effective training to improve the place type quality.

Figure 16 depicts the place type quality classified by public health regions at provincial level. Here, only 11 provinces out of 77 have good place type quality with an average above 80%, while the rest provinces must be seriously taken into account.

VII. CONCLUSION

This paper presents the `pAssessor`, a novel and comprehensive place type quality assessment technique, with respect to buildings textually and variously defined for places. In particular, `pAssessor` relies on (i) the building-place ontology – to enable the semantic detection; (ii) the building semantic selection algorithm – to enable the selection of the most applicable building semantic among different potential semantics; (iii) the boosted feature extraction – to promote the visibility of places as buildings; (iv) the supervised-learning algorithm – to learn the building-place relations and hence perform place classification; and finally (v) the assessment algorithm – to evaluate the quality of place types and classify such quality into 3 categories: *high*, *ambiguous* and *low* quality. The experimental results showed that the efficiency of `TanRabad` in assessing the quality of place types is greater than 87.5%.

Future works focus on (i) improving the ambiguous place type quality assessment; (ii) embedding `pAssessor` as part of `TanRabad QUALITY` to notify the poorness of place types; and (iii) providing effective trainings and guidelines for all public health officials (users) to achieve the place type quality improvement.

ACKNOWLEDGMENT

The authors would like to thank all public health officials throughout Thailand for the continuous collection of visual larval survey data via `TanRabad SURVEY` and the Department of Disease Control (DDC) for having `TanRabad` as a surveillance tool for *Aedes*-borne diseases. This paper and the research behind it would not have been possible without the need from public health officials for the data quality improvement.

REFERENCES

[1] *Degue Infection Summary*. Accessed: 2019. [Online]. Available: https://ddc.moph.go.th/th/site/office/view/thaivbd
[2] NSTDA and DDC. (2016). *Tanrabad: Software Suite for Dengue Surveillance and Control*. [Online]. Available: https://www.tanrabad.org
[3] A. D. Focks. *A Review of Entomological Sampling Methods and Indicators For Dengue Ventor*. Accessed: 2003. [Online]. Available: https://apps.who.int/iris/bitstream/handle/10665/68575/TDR_IDE_DEN_03.1.pdf
[4] *Vector Surveillance*. Accessed: 2020. [Online]. Available: https://www.who.int/denguecontrol/monitoring/vector_surveillance/en/
[5] N. Sahavechaphan, M. Rattananen, P. Panichphol, W. Wongwilai, S. Iamsiri, and P. Sadakorn, "TanRabad: Software suite for dengue epidemic surveillance and control," *Int. J. Infectious Diseases*, vol. 53, p. 118, Dec. 2016.
[6] *Global Strategy for Dengue Prevention Control*, World Health Org., Geneva, Switzerland, 2012.
[7] P. Viswanathan, T. Southey, J. J. Little, and A. Mackworth, "Automated place classification using object detection," in *Proc. Can. Conf. Comput. Robot Vis.*, Jun. 2010, pp. 324–330.
[8] A. Rottmann, O. M. Mozo, C. Stachniss, and W. Burgard, "Semantic place classification of indoor environments with mobile robots using boosting," in *Proc. 20th Nat. Conf. Artif. Intell. 70th Innov. Appl. Artif. Intell. Conf.*, Jul. 2005, pp. 1306–1311.
[9] O. M. Mozos, C. Stachniss, and W. Burgard, "Supervised learning of places from range data using Adaboost," in *Proc. Int. Conf. Robot. Autom.*, Apr. 2005, pp. 1730–1735.
[10] M. F. Kilkenny and K. M. Robinson, "Data quality: 'Garbage in–garbage out,"' *Health Inf. Manage. J.*, vol. 47, no. 3, pp. 103–105, Sep. 2018.
[11] J. Goyvaerts. *Regular Expressions the Complete Tutorial*. Accessed: Jul. 2007. [Online]. Available: https://www.princeton.edu/~mlovett/reference/Regular-Expressions.pdf

[12] N. F. Noy and D. L. McGuinness, "Ontology development 101: A guide to creating your first ontology," Stanford Knowl. Syst. Lab., Stanford Univ., Stanford, CA, USA, Tech. Rep. KSL-01-05 and SMI-2001-0880, 2001.
[13] W. Tungkwampian, A. Theerarungchaisri, and M. Buranarach, "Development of thai herbal medicine knowledge base using ontology technique," *Thai J. Pharmaceutical Sci.*, vol. 39, no. 3, pp. 102–109, July-Sep. 2015.
[14] J. Panawong, N. Kaewboonma, and W. Chansanam, "Building an ontology of flora of Thailand for developing semantic electronic dictionary," in *Proc. 3rd Int. Conf. Appl. Sci. Technol.*, Sep. 2018, pp. 020118-1–020118-7.
[15] A. Thunkijjanukij, A. Kawtrakul, S. Panichsakpatana, and U. Veesommai, "Ontology development: A case study for thai rice," *Kasetsart J.*, vol. 42, pp. 594–604, Dec. 2009.
[16] *Longest Common Subsequence*. Accessed: Sep. 30, 2019. [Online]. Available: https://www.geeksforgeeks.org/longest-common-subsequence-dp-4/
[17] C.-Y. J. Peng, K. L. Lee, and G. M. Ingersoll, "An introduction to logistic regression analysis and reporting," *J. Educ. Res.*, vol. 96, no. 1, pp. 3–14, 2002.
[18] C. Zhang and Y. Ma, "Random forests," in *Ensemble Machine Learning: Methods and Applications*. New York, NY, USA: Springer, Jan. 2011, pp. 157–176, ch. 5.
[19] T. Evgeniou and M. Pontil, "Support vector machines: Theory and applications," in *Machine Learning and Its Applications: Advanced Lectures*. New York, NY, USA: Springer, Jan. 2001.
[20] S. Taheri and M. Mammadov, "Learning the naive Bayes classifier with optimization models," *Int. J. Appl. Math. Comput. Sci.*, vol. 23, no. 4, pp. 787–795, Dec. 2013.

NAIYANA SAHAVECHAPHAN received the B.S. degree in computer science from Chiang Mai University, Chiang Mai, Thailand, in 1991, and the M.S. and S.Dc. degrees in computer science from the University of Massachusetts at Lowell, Lowell, MA, USA, in 1993 and 2006, respectively.

Since 1994, she has been working with the National Electronics and Computer Technology Center (NECTEC), Ministry of High Education, Science, Research and Innovation, Thailand, as a System Analyst, Researcher, and currently a Senior Researcher. Her research interests focus on three main objectives: (i) to develop a real-time data platform to simplify the development of any real-time and event-driven software systems; (ii) to develop an identity platform to support the authentication and authorization of both end-users and systems; and (iii) to build real-world software systems based on the previous platforms mainly but not limited to epidemiological and environmental domains. Essentially, she has expertise in software engineering, database design, data integration, complex event processing, event-driven architecture, and data stream management.

Dr. Sahavechaphan's awards and honors include an Invention Award for "TanRabad" from the National Research Council of Thailand in 2019, the Governmental Service Award for "TanRabad" from the Office of the Civil Service Commission in 2019, and two Star Awards from the National Electronics and Computer Technology Center in 2019.

JUKKRAPONG PONHARN received the B.S. degrees in computer science from Mahasarakham University, Mahasarakham, Thailand, in 2008. He is currently pursuing the M.S. degree in computer science with Thammasat University, Pathum Thani, Thailand.

Since 2008, he has been working with the National Electronics and Computer Technology Center, Ministry of High Education, Science, Research and Innovation, Thailand, as a Research Assistant. His interesting research includes the development of predictive maintenance for automatic weather station (AWS) using machine learning techniques.

Mr. Ponharn's awards and honors include an Invention Award for "TanRabad" from the National Research Council of Thailand in 2019, the Governmental Service Award for "TanRabad" from the Office of the Civil Service Commission in 2019, and the two Star Awards from the National Electronics and Computer Technology Center in 2019.

ASAMAPORN CHATRATTIKORN received the B.S. degrees in computer science from Naresuan University, Phitsanulok, Thailand, in 2007, and the M.S. degree in software engineering from the National Institute of Development Administration, Bangkok, Thailand, in 2013.

Since 2007, she has been working at the National Electronics and Computer Technology Center, Ministry of High Education, Science, Research and Innovation, Thailand, as a Senior Engineer. Her interesting research includes the development of data collection, real-time processing techniques, and analytics using a larval survey and Aedes-borne Disease.

Ms. Chatrattikorn's awards and honors include an Invention Award for "TanRabad" from the National Research Council of Thailand in 2019, the Governmental Service Award for "TanRabad" from the Office of the Civil Service Commission in 2019, and the two Star Awards from the National Electronics and Computer Technology Center in 2019.

PONGSAKORN SADAKORN received the B.S. degree in agriculture from Maejo University, Chiang Mai, Thailand, in 2009. He also received a Certificate in English Program (English) for Academic Purpose course from Massey University, New Zealand, in 2017, and Bridging English Program from The University of Queensland, Australia, in 2018.

From March 2011 to April 2014, he was an Entomologist at the Office of Disease Prevention and Control 1 Bangkok, Department of Diseases Control, Ministry of Public Health, Thailand. His responsibility was to prevent and control vector-borne diseases in three provinces (Ayutthaya, Nonthaburi, and Pathum Thani). Since May 2014, he has been the Public Health Technical Officer at the Division of Vector-Borne Diseases, Department of Diseases Control, Ministry of Public Health, Thailand. He is currently the Program Manager of the Aedes Control Program in the national level with over nine years of experience in public health and expertise in vector-borne diseases prevention and control, epidemiological, and entomological surveillance.

Mr. Sadakorn's awards and honors include the Best Employee Award of the Department of Disease Control (2014), the Invention Award for "TanRabad" from the National Research Council of Thailand in 2019, the Governmental Service Award for "TanRabad" from the Office of the Civil Service Commission in 2019, and the Pitching Rising Star Award in 2019.

SOPON IAMSIRITHAWORN received the B.Med. degree from the Ramathibodi Medical School, Mahidol University, Thailand, in 1994, and the M.P.H. degree in public health and the Ph.D. degree in epidemiology from the University of California at Los Angeles, Los Angeles, CA, USA, in 2001 and 2006, respectively.

He is currently the Director of the Division of Communicable Diseases, Department of Disease Control (DDC), Ministry of Public Health, Thailand. He has served as the Medical Epidemiologist and also the Manager of infectious diseases program for years. Moreover, he has direct experiences in establishing surveillance systems and research studies of influenza illness and other infectious diseases. During the COVID-19 crisis, he was appointed as an Incident Commander of the DDC's Emergency Operations Center for COVID-19. He is responsible for national planning and management of the COVID-19 situation, including outbreaks control. His expertise in emerging infectious disease epidemiology and public health has been useful for successful COVID-19 response. He is a co-investigator of the serological survey that aims to determine COVID-19 immunity among high-risk groups and the general population in Thailand.