Tropical Biomedicine 40(1): 80-87 (2023)
https://doi.org/10.47665/tb.40.1.019

RESEARCH ARTICLE

Identification of medically and forensically relevant flies using a decision tree-learning method

Tanajitaree, C.[1], Sanit, S.[2], Sukontason, K.L.[2], Sukontason, K.[2], Somboon, P.[2], Anakkamatee, W.[3], Amendt, J.[4], Limsopatham, K.[2*]

[1]Graduate Master's Degree Program in Parasitology, Faculty of Medicine, Chiang Mai University, Chiang Mai 50200, Thailand
[2]Center of Insect Vector Study, Department of Parasitology, Faculty of Medicine, Chiang Mai University, Chiang Mai 50200, Thailand
[3]Department of Mathematics, Faculty of Science, Naresuan University, Phitsanulok 65000, Thailand
[4]Institute of Legal Medicine, Goethe University, Frankfurt 60596, Germany
*Corresponding author: kwankamol.l@cmu.ac.th

ARTICLE HISTORY

Received: 9 December 2022
Revised: 3 February 2023
Accepted: 6 February 2023
Published: 31 March 2023

ABSTRACT

Blow flies, flesh flies, and house flies can provide excellent evidence for forensic entomologists and are also essential to the fields of public health, medicine, and animal health. In all questions, the correct identification of fly species is an important initial step. The usual methods based on morphology or even molecular approaches can reach their limits here, especially when dealing with larger numbers of specimens. Since machine learning already plays a major role in many areas of daily life, such as education, business, industry, science, and medicine, applications for the classification of insects have been reported. Here, we applied the decision tree method with wing morphometric data to construct a model for discriminating flies of three families [Calliphoridae, Sarcophagidae, Muscidae] and seven species [*Chrysomya megacephala* (Fabricius), *Chrysomya rufifacies* (Macquart), *Chrysomya* (*Ceylonomyia*) *nigripes* Aubertin, *Lucilia cuprina* (Wiedemann), *Hemipyrellia ligurriens* (Wiedemann), *Musca domestica* Linneaus, and *Parasarcophaga* (*Liosarcophaga*) *dux* Thomson]. One hundred percent overall accuracy was obtained at a family level, followed by 83.33% at a species level. The results of this study suggest that non-experts might utilize this identification tool. However, more species and also samples per specimens should be studied to create a model that can be applied to the different fly species in Thailand.

Keywords: Decision tree; wing morphometric; identification model; forensic entomology.

INTRODUCTION

Filth flies (Order Diptera) are considered medically and forensically important insects. In Thailand, these flies include three main families: Calliphoridae (blow flies), Sarcophagidae (flesh flies), and Muscidae (house fly and relatives). On the negative side, adult flies are not only a nuisance to humans and animals, but also a mechanical vector for various pathogens (Greenberg, 2019). Furthermore, most larval species are myiasis-producing agents in humans and livestock (Wall & Lovatt, 2015). Conversely, adult blow flies are important pollinators in agriculture, their larvae can improve wound treatment in medicine and act as entomological evidence in forensic investigation (Amendt *et al*., 2004; Sherman *et al*., 2013; Byrd & Tomberlin, 2019). The latter is well-known to be used in estimating the minimal post-mortem interval (PMI_{min}) or time since death of a human corpse because blow flies are usually the first insects to arrive on a dead body. For this application, the reliable identification of insect specimens found associated with the body is typically an early important step (Amendt *et al*., 2007). There are two main techniques for identification, based on morphology (Tumrasvin & Shinonaga, 1982; Kurahashi & Bunchu, 2011; Kurahashi & Samerjai, 2018) and DNA (Wells & Stevens, 2008; Sontigun *et al*., 2018; Samerjai *et al*., 2019), but some limitations may occur. For example, external or internal specific features of the larval or adult insect are difficult for non-experts who are not familiar with. Although molecular technique (e.g., sequencing of the so-called barcoding region of the mitochondrial genome) yields high specificity and sensitivity and can be applied to all stages (egg, larva, pupa, and adult) of flies, it requires sophisticated equipment, high budget, and know-how for analyzing and matching the results. In addition, unclear results in cases of low amounts of DNA or degraded DNA have been reported (Saigusa *et al*., 2009; Sonet *et al*., 2013).

Over the last decades, data mining (known as knowledge discovery in databases) has been widely utilized in various applications, e.g., economics, finance, marketing, industry, education, engineering, science, or medicine. It extracts useful patterns of information from huge datasets and transforms it into important knowledge (Larose & Larose, 2014). While geometric morphometrics is widely used as a tool for insect identification due to its low-cost, ease of use, and high accuracy ($\geq$ 80% identification success), data mining might encourage the benefit when applied simultaneously. In general, insect exoskeletons do not change much after developing into the adult stage. Shape quantification based

on geometric morphometrics has been an effective approach to describe morphological variation (Tatsuta *et al*., 2018). Within medical and forensic important Diptera, morphometric analysis of wings has been investigated widely for species identification, e.g., blow flies, flesh flies, house flies and relatives, and mosquitoes (Grzywacz *et al*., 2017; Sontigun *et al*., 2017; Sontigun *et al*., 2019; Szpila *et al*., 2019; Champakaew *et al*., 2021; Limsopatham *et al*., 2021). However, these studies mainly focused on the difference between units (genera/species) of specimens based on traditional statistical analyses without addressing the question of how to predict and assign unidentified (blind) specimens to such a unit. Therefore, an identification model trained on the reference or research database (identified samples served as a training group) should be established to determine the blind specimens (unidentified samples served as a testing group).

The decision tree, used for classification purposes based on machine learning, is one of the appropriate methods for creating an identification model in the context of data mining (Quinlan, 1986; Bell, 1999; De'ath & Fabricius, 2000). It is a top-down flowchart-like structure consisting of a root node (nodes that have no incoming edge) and leaf nodes (terminal nodes). Each root and internal node contain a splitting rule for partitioning input into two or more subsets according to the outcome, represented by outgoing edges (branches) of the node. The data item follows the tree from root to leaf based on splitting rules and assigned to its prediction class. The paths from root to leaf represent classification rules (criteria) (Witten *et al*., 2016).

The present study aims to build a decision tree model based on the dataset of wing geometric morphometrics of common fly species in Thailand [the blow flies *Chrysomya megacephala* (Fabricius), *Chrysomya rufifacies* (Macquart), *Chrysomya* (*Ceylonomyia*) *nigripes* Aubertin, *Lucilia cuprina* (Wiedemann), and *Hemipyrellia ligurriens* (Wiedemann), the house fly *Musca domestica* Linneaus, and the flesh fly; *Parasarcophaga* (*Liosarcophaga*) *dux* Thomson], and to validate that model with blind samples. The model could be useful not just for non-taxonomists to identify forensically important fly species in Thailand and can also be considered as a template for future studies, i.e., analyzing other species in many different regions of the world and potentially distinguishing populations of one and the same species.

MATERIALS AND METHODS

Fly samples

Training group

In order to obtain various and reliable data, laboratory and field strains were included. Colonies of *C. megacephala*, *C. rufifacies*, *C. nigripes*, *L. cuprina*, *H. ligurriens*, *M. domestica*, and *P. dux* have been maintained for 12 years under ambient temperature and humidity at the Department of Parasitology, Faculty of Medicine, Chiang Mai University. The rearing methods followed Sukontason *et al*. (2008). For field strains, the specimens were captured from six provinces in Thailand (Chiang Mai, Lampang, Nan, Phitsanulok, Ubon Ratchathani, and Songkla) using a hand-held fly net, sweeping it over a bait of 1-day-old beef offal and then kept alive before transporting back to the laboratory for sacrificing by freezing at -20 °C for 2 hours. After that, all flies were pinned and identified under a stereomicroscope (Olympus, Japan) based on the diagnostic morphological characters described by Tumrasvin and Shinonaga (1982), Kurahashi and Bunchu (2011), and Kurahashi and Samerjai (2018). The identified specimens were kept at -20 °C in an individual container until used for wing preparation. The number of specimens used for training is shown in Table 1.

Testing group

Thirty flies of each species (*C. megacephala*, *C. rufifacies*, *C. nigripes*, *L. cuprina*, *H. ligurriens*, *M. domestica*, and *P. dux*) were captured in the field (Table 2). Sampling and identification were carried out as described above.

Specimen preparation and image processing

A total of 2,054 adult *C. megacephala*, *C. rufifacies*, *C. nigripes*, *L. cuprina*, *H. ligurriens*, *M. domestica*, and *P. dux* were prepared by removing their right wings with fine forceps. Each wing was submerged in xylene to avoid bubbles before placing on a drop of Permount™ mounting medium (Fisher Scientific, USA) on a glass slide. Then, a thin layer of Permount™ was added to the wing and a cover slip was placed on top. After drying at room temperature for a week, each wing was photographed using a Nikon D5100 digital camera attached to a stereomicroscope (Olympus SZ51, Japan) at 1.5x magnification. JPG files of each image were converted into tps files using TpsUtil V.1.74 software (Rohlf, 2015b) to minimize a possible bias when digitizing the landmarks. Eighteen landmarks (Figure 1) were digitized using TpsDig2 V.2.30 software (Hall *et al*., 2014; Rohlf, 2015a). To reduce the measurement error, digitization was undertaken twice (Arnqvist & Martensson, 1998).

Model construction, validation, and data analysis

The established tps files, containing digitized coordinates of 18 landmarks (1x1y to 18x18y) from all wings, were subjected to Microsoft Excel for distance calculation. Twenty-three distances (µm) were selected (Figure 2). To remove the wing size effect, each distance was normalized before determining the difference of each distance data between the fly family and species. For each sample, the sum of length of the distances a, b, c, and d was rescaled to 5000 µm to obtain the normalized distances. Therefore, the word "distances" refers to the normalized distances here. Only significant distances were used as features in the model construction process for family and species identification within the RapidMiner software (https://rapidminer.com/). Independent *t*-test was used to analyze the difference of each distance data between fly family and species ($p < 0.05$; SPSS program version 22.0).

For model validation, all distances of each blind specimen (testing group) were used to predict its family/species by applying into the identification model (created based on the training group). Overall and individual accuracy were calculated by comparing the predicted class and the actual class of each specimen within the RapidMiner software (https://rapidminer.com/).

Ethics approval

The protocol of this study was approved by the Research Ethics Committee (Institutional Animal Care and Use Committee) (Protocol Number 30/2563) of Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand, for consideration before performing the experiments.

RESULTS

Families identification

Wing morphometric distances analysis

For each distance, statistical analysis revealed significant differences between three families (Independent *t*-test, $p < 0.05$; Supplementary Table S1), except distance a, d, o, r, and w for Calliphoridae-Sarcophagidae and distance j, q, t, and v for Muscidae-Sarcophagidae. Hence, all distances were considered (used with some or most families) and not removed for further model construction.

Table 1. Fly specimens used for training

Family	Species	Code of specimens	Location	GPS reference		No.	
				Latitude	Longitude	Male	Female
Calliphoridae	*Chrysomya megacephala*	CM	Fly-rearing room, CMU	–	–	110	130
			Hangdong, Chiang Mai	18.7734790	98.8602090	16	26
			Ko Kha, Lampang	18.1275210	99.4123955	16	23
			Na Muen, Nan	18.2116498	100.6666276	–	4
			Doi Suthep, Chiang Mai	18.4820252	98.5434238	15	30
			E-James swamp, Ubon Ratchathani	15.1250239	104.9133332	15	27
					Total	**412**	
	Chrysomya rufifacies	CR	Fly-rearing room, CMU	–	–	110	114
			Ko Kha, Lampang	18.1275210	99.4123955	10	30
			Faculty of Veterinary Science Prince of Songkla University, Songkhla	7.0064652	100.5014770	10	30
			Suan Pa Kaokrayang, Phitsanulok	16.8454373	100.7479117	16	30
					Total	**350**	
	Chrysomya (*Ceylonomyia*) *nigripes*	CN	Fly-rearing room, CMU	–	–	110	110
			Faculty of Agriculture, Mae Hia, CMU	18.7666944	98.9340278	–	4
					Total	**224**	
	Lucilia cuprina	LC	Fly-rearing room, CMU	–	–	133	176
			Male medical dormitory 1, CMU	18.7907510	98.9717779	24	8
					Total	**341**	
	Hemipyrellia ligurriens	HL	Fly-rearing room, CMU	–	–	110	110
					Total	**220**	
Muscidae	*Musca domestica*	MD	Fly-rearing room, CMU	–	–	110	112
			Longan orchard, Mae Hia, Chiang Mai	18.455666	98.554013	–	24
			Palm garden, Mae Hia, Chiang Mai	18.4527841	98.5548515	–	6
					Total	**252**	
Sarcophagidae	*Parasarcophaga* (*Liosarcophaga*) *dux*	PD	Fly-rearing room, CMU	–	–	109	116
			Mae Khanin, Hangdong, Chiang Mai	18.7928889	98.7908611	15	15
					Total	**255**	

Family identification model

The tree flow in a top-down manner from the root node through the internal nodes and finally to the leaf nodes (the predicted family) as shown in Figure 3. When distance [Dis.] b was less than or equal to 737.62 µm, Muscidae (a leaf node) was an identified result. On the other hand, if the outcome of distance b was more than 737.62 µm, another internal node (e.g., Dis. g > 620.70 µm) was considered. If distance g was more than 620.70 µm, the node of Dis. e > 1342.00 µm was then considered. In case of "yes", the predicted family was Sarcophagidae. If distance e was less than or equal to 1342.00 µm, the predicted family was Calliphoridae. While the outcome of distance g was less than or equal to 620.70 µm, the node of Dis. f > 1111.08 µm was considered. Finally, seven distances (a, b, e, f, g, p, and v) were assigned to the model's splitting rules.

Validation of the model

The performance of the family model showed a 100% overall accuracy for all blind/tested specimens and a 100% individual family accuracy (Calliphoridae, Sarcophagidae, and Muscidae) (Table 3).

Species identification

Wing morphometric distances analysis

For each distance, statistical analysis revealed significant difference between seven species (Independent *t*-test, $p < 0.05$; Supplementary Table S2), except distance a for *C. rufifacies-C. megacephala* and *L. cuprina-M. domestica*; distance e for *C. nigripes-H. ligurriens*, *C. nigripes-L. cuprina*, and *H. ligurriens-L. cuprina*; distance g for *H. ligurriens-L. cuprina*; distance h for *C. rufifacies-C. megacephala* and *H. ligurriens-P. dux*; distance i for *C. rufifacies-L. cuprina* and *C. rufifacies-C. megacephala*, *H. ligurriens-P. dux*, and *L. cuprina-C. megacephala*; distance j for *C. nigripes-M. domestica*, *C. nigripes-P. dux*, *C. rufifacies-C. megacephala*, *L. cuprina-C. megacephala*, and *M. domestica-P. dux*; distance k for *C. nigripes-C. megacephala*; distance l for *C. rufifacies-L. cuprina*; distance m for *H. ligurriens-L. cuprina* and *C. megacephala-P. dux*; distance o for *C. nigripes-P. dux*; distance p for *C. rufifacies-C. megacephala*; distance q for *L. cuprina-C. megacephala* and *M. domestica-P. dux*; distance r for *C. nigripes-C. megacephala*, *C. nigripes-P. dux*, and *L. cuprina-P. dux*; distance t for *C. rufifacies-C. megacephala* and *M. domestica-P.*

Table 2. Fly specimens used for testing

Family	Species	Code of specimens	Location	GPS reference		No.	
				Latitude	Longitude	Male	Female
Calliphoridae	*Chrysomya megacephala*	CM	Na Muen, Nan	18.2116498	100.6666276	8	22
					Total	**30**	
	Chrysomya rufifacies	CR	Hangdong, Chiang Mai	18.773479	98.860209	10	20
					Total	**30**	
	Chrysomya (*Ceylonomyia*) *nigripes*	CN	Faculty of Agriculture, Mae Hia, CMU	18.7666944	98.9340278	16	14
					Total	**30**	
	Lucilia cuprina	LC	Na Muen, Nan	18.2116498	1006666276	12	8
			Male medicine dormitory 1, CMU	18.7907510	98.9717779	–	10
					Total	**30**	
	Hemipyrellia ligurriens	HL	Na Muen, Nan	18.2116498	100.6666276	4	14
			E-James swamp, Ubon Ratchathani	15.1250239	104.9133332	4	8
					Total	**30**	
Muscidae	*Musca domestica*	MD	Forest area, Mae Hia, Chiang Mai	18.460108	98.56083	1	13
			Longan orchard, Mae Hia, Chiang Mai	18.455666	98.554013	9	–
			Palm garden, Mae Hia, Chiang Mai	18.4527841	98.5548515	–	**7**
					Total	**30**	
Sarcophagidae	*Parasarcophaga* (*Liosarcophaga*) *dux*	PD	Mae Khanin Hangdong, Chiang Mai	18.7928889	98.7908611	15	15
					Total	**30**	

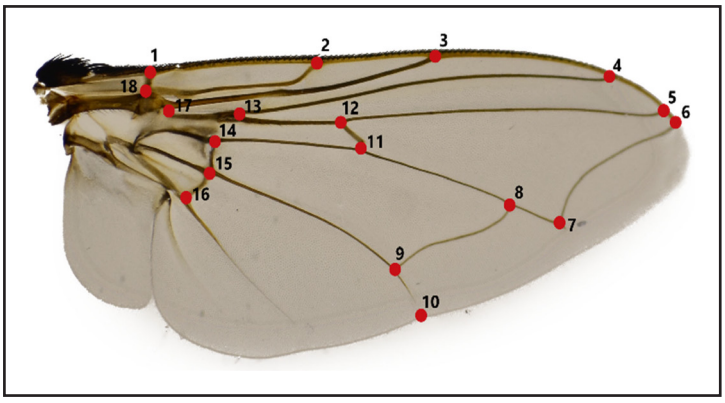

Figure 1 Right wing of *C. rufifacies* showing the 18 landmarks modified from Hall *et al*. (2014).

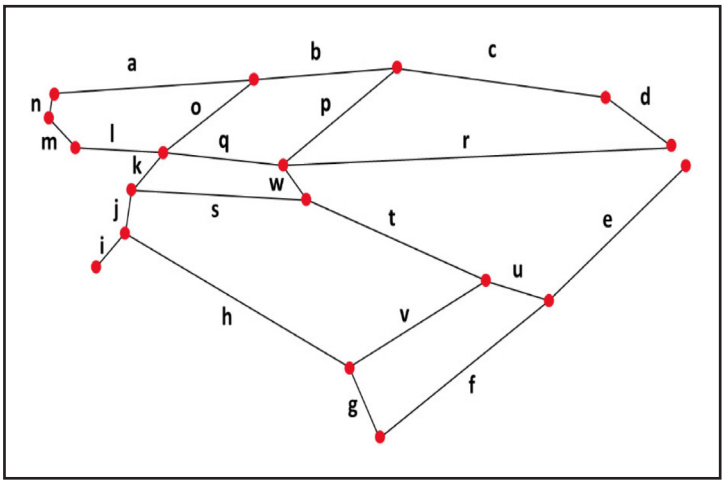

Figure 2 Illustration of 23 distances on wing for model construction.

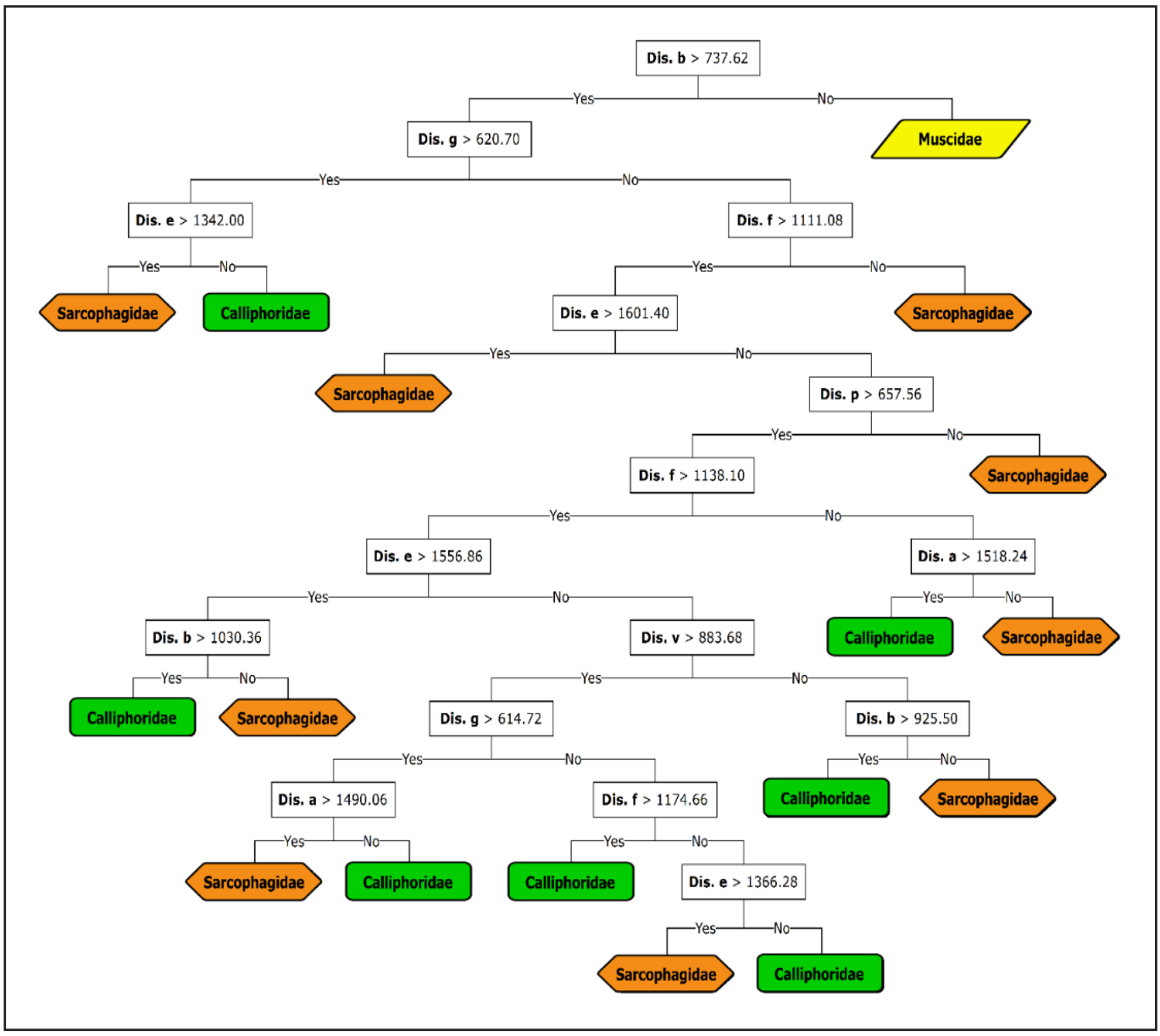

Figure 3. Family identification model; wing morphometric distances (µm).

Table 3. Performance of family identification model

	True families		
	Calliphoridae	Muscidae	Sarcophagidae
Predicted families			
Calliphoridae	**30**	0	0
Muscidae	0	**30**	0
Sarcophagidae	0	0	**30**
Individual family accuracy (%)	**100.00**	**100.00**	**100.00**
Overall accuracy (%)		**100.00**	

dux; distance v for *M. domestica-P. dux*. Similar to the family level, all distances were still used and no data were removed for further model construction.

Species identification model

Figure 4 showed Dis. b > 737.62 µm as the root node of the tree. When the outcome of distance b was less than or equal to 737.62 µm, *M. domestica* (leaf node) was an identified result. On the other hand, if the outcome of distance b was more than 737.62 µm, another internal node (e.g., Dis. b > 1026.04 µm) was considered. In case of "yes", the tree flow of other internal nodes starting with the node of Dis. e > 1403.12 µm was followed. Otherwise, the tree went along with the internal node of Dis. g > 393.34 µm. Regarding these internal nodes, twelve distances (a, b, c, d, e, f, g, h, l, r, s, and u) were assigned in the splitting rules of the model. Finally, the end of each leaf node indicated the predicted fly species (CN, *C. nigripes*; CR, *C. rufifacies*; HL, *H. ligurriens*; LC, *L. cuprina*; CM, *C. megacephala*; MD, *M. domestica*; PD, *P. dux*).

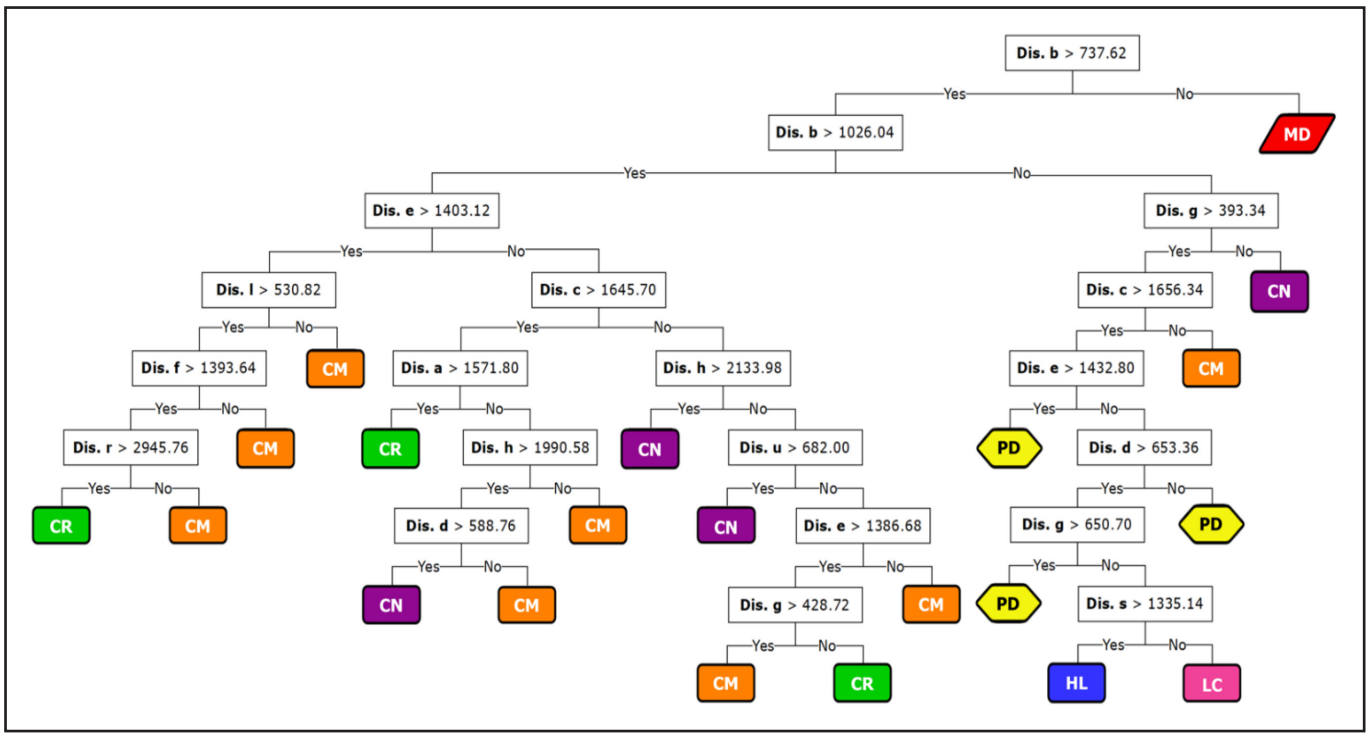

Figure 4. Species identification model; wing morphometric distances (µm).

Table 4. Performance of species identification model

	True species						
	CM	HL	LC	CR	CN	MD	PD
Predicted species							
CM	**20**	3	0	1	0	0	0
HL	1	**24**	18	0	0	0	0
LC	0	3	**12**	0	0	0	0
CR	8	0	0	**29**	0	0	0
CN	1	0	0	0	**30**	0	0
MD	0	0	0	0	0	**30**	0
PD	0	0	0	0	0	0	**30**
Individual species accuracy (%)	**66.67**	**80.00**	**40.00**	**96.67**	**100.00**	**100.00**	**100.00**
Overall accuracy (%)	**83.33**						

Abbreviation: CN, *C. nigripes*; CR, *C. rufifacies*; HL, *H. ligurriens*; LC, *L. cuprina*; CM, *C. megacephala*; MD, *M. domestica*; PD, *P. dux.*

Validation of the model

Overall accuracy was 83.33% in all seven blind/tested species. As for the individual species accuracy, *C. nigripes*, *M. domestica*, and *P. dux* showed 100% correctly identified blind specimens, followed by *C. rufifacies* (96.67%), *H. ligurriens* (80.00%), *C. megacephala* (66.67%), and *L. cuprina* (40.00%), respectively (Table 4).

The problematic of misidentification was represented in *C. megacephala*, *H. ligurriens*, and *L. cuprina*. For *H. ligurriens*, this species was misidentified with *C. megacephala* (predicted sp. *n* = 3/true sp. *n* = 30) and *L. cuprina* (predicted sp. *n* = 3/true sp. *n* = 30). While the lowest percentage of individual species accuracy was observed in *C. megacephala* and *L. cuprina*, the former species was misidentified with *H. ligurriens* (predicted sp. *n* = 1/true sp. *n* = 30), *C. rufifacies* (predicted sp. *n* = 8/true sp. *n* = 30), and *C. nigripes* (predicted sp. *n* = 1/true sp. *n* = 30), respectively and the latter species was misidentified with *H. ligurriens* (predicted sp. *n* = 18/true sp. *n* = 30) (Table 4).

DISCUSSION

As species identification is an early and important step in forensic entomology, very accurate and reliable methods are required. The results in this study are based on machine learning (ML), i.e., the application of computer algorithms that automatically improve themselves through experience and the use of data (here: wing morphometric data). ML algorithms build a model based on training data, in order to make predictions or decisions without being explicitly programmed to do so. There are many sophisticated ML methods, such as decision tree, support vector machine, cluster and principal component analysis, artificial neural network, linear

regression, and deep learning (Xu & Jackson, 2019). Here we have focused on the decision tree method because it is simple to understand and interpret.

When identifying families, Muscidae was individually separated from the root node (Dis. b ≤ 737.62 µm), whereas Calliphoridae and Sarcophagidae still need other criteria for identification. The reason for the easy identification of the Muscidae could be their wing venation, as this feature, together with the thorax and the chaetotaxy on the legs, is one of the most important features for identifying adult Muscidae (Tumrasvin & Shinonaga, 1982). The species identification model agreed with the family model in which *M. domestica* is separated near the root node. Furthermore, *P. dux* was placed as a terminal leaf node if distance b was less than or equal to 1026.04 µm while the tree flow in a top-down manner from the node of Dis. e > 1403.12 µm represented most calliphorid species in a terminal leaf node. From this, it can be assumed that distance b could be important for the classification.

The model performance showed high predictive power with a 100% overall accuracy in the family model. At the species level, the overall accuracy was 83.33%, which represents the major issue in *H. ligurriens*, *C. megacephala*, and *L. cuprina*. Most blind specimens of *C. megacephala* was identified as *C. rufifacies*, whereas *L. cuprina* were mainly identified as *H. ligurriens*. The reason might be related with their genetic relationship, i.e., same genus (*C. megacephala* vs *C. rufifacies*) or same subfamily (*L. cuprina* vs *H. ligurriens*). Especially Luciliinae, based on cytochrome oxidase I gene (*COI*) analysis revealed *H. ligurriens* embedded within the *L. cuprina* clade (Wells *et al*., 2007; Williams *et al*., 2016). In fact, the external morphology between *C. megacephala* vs *C. rufifacies* and *L. cuprina* vs *H. ligurriens* is easily recognized by gena color and body color. Orange gena indicated *C. megacephala*, whereas cupreous body is indicated *L. cuprina*. Therefore, comprising some characteristic with wing morphometric data might be improved the accuracy of the identification model.

Decision tree method has been widely applied for identification/prediction purpose in many fields (e.g., medicine, biological science, economics, etc.). For instance, mosquitoes differentiation between species (*Aedes albopictus*, *Aedes vexans*, and *Culex* spp.) and sexes from wing beat frequency and optical cross section data showed 71.6% of the performance (Genoud *et al*., 2020). Csosz *et al*. (2015) reported high predictivity (>95%) of myrmicine ants identification from 22 continuous morphometric traits. Furthermore, fish-school identification by acoustic echo trace data led to a 73% overall accuracy, followed by individual accuracy of myctophid (20%), mackerel (76%), herring (94%), and layer (100%), respectively (Fernandes, 2009). Identifying eight bat species based on a decision tree approach using ultrasonic calls showed a 70% overall accuracy (Herr *et al*., 1997). Besides animal taxonomy, plant identification using trait databases has been successfully applied too, with an accuracy of more than 89.1% (Almeida *et al*., 2020). In ecology, decision tree method was applied for identifying and monitoring the mangrove forest change from time series in the Pearl River Estuary using multi-temporal Landsat TM data and ancillary GIS data. Here, high accuracy of mangrove identification between 81.0-87.2% were reached (Liu *et al*., 2008). In the present study, the identification model created by a decision tree method showed high performance (≥ 80%) which can be applied to families (Calliphoridae, Muscidae, and Sarcophagidae) and species identification (*C. megacephala*, *C. rufifacies*, *C. nigripes*, *L. cuprina*, *H. ligurriens*, *P. dux*, and *M. domestica*). However, increasing both number and species of training and testing groups of medical and forensic flies in these three families should be provided to obtain more data for further model improvement.

Our results provide non-taxonomists a new alternative tool for identifying adult flies from, e.g., a field survey or a crime scene. No knowledge of fly morphology (e.g., notopleuron setae, acrostichal bristles, chaetotaxy on leg, genitalia, etc.) is needed. It allows non-taxonomists to solely measure the wing distance and follow the identification model. While taxonomists have to individual pinned and individual identified the morphology point by point of sample under stereo microscope using their personal skill. However, this tool might be a new facilitating method to help taxonomist when incomplete body of flies are found. Nevertheless, for correct identification one should pay attention to the condition of the wings, the preparation of specimen (wings on flies or on slides), the perspective when photographing, and last but not least the limitation of the families and species studied here.

In conclusion, using decision tree method is the first report of fly's families and species identification model in Thailand. Due to high identification success, this information might be an initial template not only for flies' identification, but also other insects. Furthermore, the output from this work might be useful to develop an application, which is easy for the user in the future.

Conflicts of interest

The authors declare no conflicts of interest.

Acknowledgement

This research was partially funded by the Chiang Mai University, Thailand.

REFERENCES

Amendt, J., Campobasso, C.P., Gaudry, E., Reiter, C., LeBlanc, H.N. & Hall, M.J. (2007). Best practice in forensic entomology—standards and guidelines. *International Journal of Legal Medicine* **121**: 90-104. https://doi.org/10.1007/s00414-006-0086-x

Amendt, J., Krettek, R. & Zehner, R. (2004). Forensic entomology. *Naturwissenschaften* **91**: 51-65. https://doi.org/10.1007/s00114-003-0493-5

Almeida, B.K., Garg, M., Kubat, M. & Afkhami, M.E. (2020). Not that kind of tree: assessing the potential for decision tree–based plant identification using trait databases. *Applications in Plant Sciences* **8**: e11379. https://doi.org/10.1002/aps3.11379

Arnqvist, G. & Martensson, T. (1998). Measurement error in geometric morphometrics: empirical strategies to assess and reduce its impact on measures of shape. *Acta Zoologica Academiae Scientiarum Hungaricae* **44**: 73-96.

Bell, J.F. (1999). Tree-based methods. In: Machine Learning Methods for Ecological Applications, Fielding, A.H. (editor) 1st edition. Boston: Springer, pp. 89-105. https://doi.org/10.1007/978-1-4615-5289-5_3

Byrd, J.H. & Tomberlin, J.K. (2019). Forensic entomology: the utility of arthropods in legal investigations, 3rd edition. Boca Raton: CRC press, pp. 1-620. https://doi.org/10.4324/9781351163767

Champakaew, D., Junkum, A., Sontigun, N., Sanit, S., Limsopatham, K., Saeung, A., Somboon, P. & Pitasawat, B. (2021). Geometric morphometric wing analysis as a tool to discriminate female mosquitoes from different suburban areas of Chiang Mai province, Thailand. *PLoS One* **16**: e0260333. https://doi.org/10.1371/journal.pone.0260333

Csosz, S., Heinze, J. & Miko, I. (2015). Taxonomic synopsis of the Ponto-Mediterranean ants of *Temnothorax nylanderi* species-group. *PLoS One* **10**: e0140000. https://doi.org/10.1371/journal.pone.0140000

De'ath, G. & Fabricius, K.E. (2000). Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology* **81**: 3178-3192. https://doi.org/10.1890/0012-9658(2000)081[3178:CARTAP]2.0.CO;2

Fernandes, P.G. (2009). Classification trees for species identification of fish-school echotraces. *ICES Journal of Marine Science* **66**: 1073-1080. https://doi.org/10.1093/icesjms/fsp060

Genoud, A.P., Gao, Y., Williams, G.M. & Thomas, B.P. (2020). A comparison of supervised machine learning algorithms for mosquito identification from backscattered optical signals. *Ecological Informatics* **58**: 101090. https://doi.org/10.1016/j.ecoinf.2020.101090

Greenberg, B. (2019). Flies and disease: II. biology and disease transmission. Princeton: Princeton University Press, pp. 1-464.

Grzywacz, A., Ogiela, J. & Tofilski, A. (2017). Identification of Muscidae (Diptera) of medico-legal importance by means of wing measurements. *Journal of Parasitology Research* **116**: 1495-1504. https://doi.org/10.1007/s00436-017-5426-x

Hall, M.R., MacLeod, N. & Wardhana, A. (2014). Use of wing morphometrics to identify populations of the Old World screwworm fly, *Chrysomya bezziana* (Diptera: Calliphoridae): a preliminary study of the utility of museum specimens. *Acta Tropica* **138**: S49-S55. https://doi.org/10.1016/j.actatropica.2014.03.023

Herr, A., Klomp, N.I. & Atkinson, J.S. (1997). Identification of bat echolocation calls using a decision tree classification system. *Complexity International* **4**: 1-9.

Kurahashi, H. & Bunchu, N. (2011). The blow flies recorded from Thailand, with the description of a new species of *Isomyia* Walker (Diptera: Calliphoridae). *Japanese Journal of Systematic Entomology* **17**: 237-278.

Kurahashi, H. & Samerjai, C. (2018). Revised keys to the flesh flies of Thailand, with the establishment of a new genus (Diptera: Sarcophagidae). *Medical Entomology and Zoology* **69**: 67-93. https://doi.org/10.7601/mez.69.67

Larose, D.T. & Larose, C.D. (2014). Discovering knowledge in data: an introduction to data mining, 2nd edition. Hoboken: John Wiley & Sons, pp. 1-336. https://doi.org/10.1002/9781118874059

Limsopatham, K., Klong-Klaew, T., Fufuang, N., Sanit, S., Sukontason, K.L., Sukontason, K., Somboon, P. & Sontigun, N. (2021). Wing morphometrics of medically and forensically important muscid flies (Diptera: Muscidae). *Acta Tropica* **222**: 106062. https://doi.org/10.1016/j.actatropica.2021.106062

Liu, K., Li, X., Shi, X. & Wang, S. (2008). Monitoring mangrove forest changes using remote sensing and GIS data with decision-tree learning. *Wetlands* **28**: 336-346. https://doi.org/10.1672/06-91.1

Quinlan, J.R. (1986). Induction of decision trees. *Machine learning* **1**: 81-106. https://doi.org/10.1007/BF00116251

Rohlf, F. (2015a). TpsDig2, digitize landmarks and outlines [software version 2.20]. State University of New York.

Rohlf, F. (2015b). TpsUtil, file utility program: department of ecology and evolution. State University of New York.

Saigusa, K., Matsumasa, M., Yashima, Y., Takamiya, M. & Aoki, Y. (2009). Practical applications of molecular biological species identification of forensically important flies. *Legal Medicine* **11**: S344-S347. https://doi.org/10.1016/j.legalmed.2009.01.026

Samerjai, C., Sukontason, K.L., Sontigun, N., Sukontason, K., Klong-Klaew, T., Chareonviriyaphap, T., Kurahashi, H., Klimpel, S., Kochmann, J., Saeung, A. *et al*. (2019). Mitochondrial DNA-based identification of forensically important flesh flies (Diptera: Sarcophagidae) in Thailand. *Insects* **11**: 2. https://doi.org/10.3390/insects11010002

Sherman, R.A., Mumcuoglu, K.Y., Grassberger, M. & Tantawi, T.I. (2013). Maggot therapy. In: Biotherapy – History, Principles and Practice, Grassberger, M., Sherman, R., Gileva, O., Kim, C. & Mumcuoglu, K. (editors) 1st edition. Dordrecht: Springer, pp. 5-29. https://doi.org/10.1007/978-94-007-6585-6

Sonet, G., Jordaens, K., Braet, Y., Bourguignon, L., Dupont, E., Backeljau, T., De Meyer, M. & Desmyter, S. (2013). Utility of GenBank and the Barcode of Life Data Systems (BOLD) for the identification of forensically important Diptera from Belgium and France. *Zookeys* **365**: 307-328. https://doi.org/10.3897/zookeys.365.6027

Sontigun, N., Samerjai, C., Sukontason, K., Wannasan, A., Amendt, J., Tomberlin, J.K. & Sukontason, K.L. (2019). Wing morphometric analysis of forensically important flesh flies (Diptera: Sarcophagidae) in Thailand. *Acta Tropica* **190**: 312-319. https://doi.org/10.1016/j.actatropica.2018.12.011

Sontigun, N., Sukontason, K.L., Amendt, J., Zajac, B.K., Zehner, R., Sukontason, K., Chareonviriyaphap, T. & Wannasan, A. (2018). Molecular analysis of forensically important blow flies in Thailand. *Insects* **9**: 159. https://doi.org/10.3390/insects9040159

Sontigun, N., Sukontason, K.L., Zajac, B.K., Zehner, R., Sukontason, K., Wannasan, A. & Amendt, J. (2017). Wing morphometrics as a tool in species identification of forensically important blow flies of Thailand. *Parasites & Vectors* **10**: 229. https://doi.org/10.1186/s13071-017-2163-z

Sukontason, K., Piangjai, S., Siriwattanarungsee, S. & Sukontason, K.L. (2008). Morphology and developmental rate of blowflies *Chrysomya megacephala* and *Chrysomya rufifacies* in Thailand: application in forensic entomology. *Parasitology Research* **102**: 1207-1216. https://doi.org/10.1007/s00436-008-0895-6

Szpila, K., Żmuda, A., Akbarzadeh, K. & Tofilski, A. (2019). Wing measurement can be used to identify European blow flies (Diptera: Calliphoridae) of forensic importance. *Forensic Science International* **296**: 1-8. https://doi.org/10.1016/j.forsciint.2019.01.001

Tatsuta, H., Takahashi, K.H. & Sakamaki, Y. (2018). Geometric morphometrics in entomology: basics and applications. *Entomological Science* **21**: 164-184. https://doi.org/10.1111/ens.12293

Tumrasvin, W. & Shinonaga, S. (1982). Studies on medically important flies in Thailand: VIII. Report on 73 species of muscid flies (excluding Muscinae and Stomoxyinae) with the taxonomic keys (Diptera: Muscidae). *Japanese Journal of Sanitary Zoology* **33**: 181-199.

Wall, R. & Lovatt, F. (2015). Blowfly strike: biology, epidemiology and control. *In Practice* **37**: 181-188. https://doi.org/10.1136/inp.h1434

Wells, J.D. & Stevens, J.R. (2008). Application of DNA-based methods in forensic entomology. *Annual Review of Entomology* **53**: 103-120. https://doi.org/10.1146/annurev.ento.52.110405.091423

Wells, J.D., Wall, R. & Stevens, J.R. (2007). Phylogenetic analysis of forensically important *Lucilia* flies based on cytochrome oxidase I sequence: a cautionary tale for forensic species determination. *International Journal of Legal Medicine* **121**: 229-233. https://doi.org/10.1007/s00414-006-0147-1

Williams, K.A., Lamb, J. & Villet, M.H. (2016). Phylogenetic radiation of the greenbottle flies (Diptera, Calliphoridae, Luciliinae). *Zookeys* **568**: 59-86. https://doi.org/10.3897/zookeys.568.6696

Witten, I.H., Frank, E., Hall, M.A. & Pal, C.J. (2016). Data mining: practical machine learning tools and techniques, 4th edition. Cambridge: Morgan Kaufmann, pp. 1-654.

Xu, C. & Jackson, S.A. (2019). Machine learning and complex biological data. *Genome Biology* **20**: 1-4. https://doi.org/10.1186/s13059-019-1689-0

SUPPLEMENTARY DATA

https://msptm.org/files/Vol40No1/tb-40-1-019-Tanajitaree-C-supplementary-data.pdf