

**USING DAISY (DIGITAL AUTOMATED IDENTIFICATION
SYSTEM) FOR AUTOMATED IDENTIFICATION OF
MOTHS OF THE SUPERFAMILY BOMBYCOIDEA OF
BORNEO**

LEONG YUEN MUNN

**DISSERTATION SUBMITTED IN FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE**

**INSTITUTE OF BIOLOGICAL SCIENCES
FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR**

2013

ABSTRACT

The health of our environment can be measured by monitoring moths. Moths are especially useful as biological indicators given that they are sensitive to environmental changes, widespread and can be found in various habitats. Therefore, by monitoring their ranges and numbers, we can obtain essential clues about our changing environment such as climate change, air pollution and the effects of new farming practices. In this study, I tested DAISY as a tool for automated identification of moths. Images of 210 species of the superfamily Bombycoidea from the book “The Moths of Borneo: Part 3: Lasiocampidae, Eupterotidae, Bombycidae, Brahmaeidae, Saturniidae, Sphingidae” were used as a training data-set for DAISY. The images were pre-processed to (600X400 pixels), mirroring of least torn wing was performed, and the file format was converted from JPEG to TIF. Training and testing of the system were performed using images of the right forewings of moths. Additionally, the hindwings of two species; *Actias maenas* and *A. Selene* were included as they have significant shape (very elongated hindwings compared to other species). Three test datasets were then used to evaluate the performance of DAISY: (i) distorted versions of the training images (ii) images from internet resources of same species to the ones in training dataset (Superfamily Bombycoidea), (iii) images from other volumes of The Moths of Borneo of species which were not included in the training dataset, from families; Notodontidae, Lymantriidae, Arctiidae, Drepaninae, Callidulidae, Geometridae, Notuidae and Noctuidae) and (iv) DAISY’s default sample images of species not in the training dataset (Belize sphingids). I classified the results of DAISY identifications into four categories for analysis; (i) test species in training set and correct species is given (true positive, TP), (ii) test species may or may not be in training set but incorrect species name is given (false positive, FP), (iii) test species not in training set and no species

name is given (true No-ID, TNI) and (iv) test species in training set but no species name is given (false No-ID, FNI). Based on these criteria I measured the precision as $TP / (TP+FP)$ and overall accuracy as $(TP+TNI) / (TP+TNI+FP+FNI)$. Overall, the precision of DAISY across all four test data-sets was 51% while overall accuracy was 50%. I discuss the potential reasons for the low observed accuracy. Finally, I make recommendations for features to be considered when designing an automated tool for moth identification.

ABSTRAK

Kesihatan persekitaran kita boleh diukur dengan pemantauan kupu-kupu. Kupu-kupu adalah amat berguna sebagai penunjuk biologi memandangkan mereka lebih sensitif terhadap perubahan alam sekitar, meluas dan boleh didapati dalam pelbagai habitat. Oleh itu, dengan memantau julat dan bilangan mereka, kita dapat petunjuk persekitaran penting yang berubah seperti perubahan iklim, pencemaran udara dan kesan-kesan amalan pertanian baru. Dalam kajian ini, saya menguji DAISY sebagai alat untuk mengenal pasti automatik kupu-kupu. 210 spesies imej daripada superfamili Bombycoidea dari buku "The Moths of Borneo: Part 3: Lasiocampidae, Eupterotidae, Bombycidae, Brahmaeidae, Saturniidae, Sphingidae" telah digunakan sebagai latihan data yang ditetapkan untuk DAISY imej telah pra-diproses untuk (600X400 piksel), mencerminkan sayap-kurangnya koyak telah dilaksanakan, dan format fail telah ditukar dari JPEG untuk Latihan TIF dan ujian sistem dijalankan menggunakan imej bergigit kanan kupu-kupu. Selain itu, yang kepak belakang dua spesies; *Actias maenas* dan *A. Selene* telah dimasukkan kerana mereka mempunyai bentuk yang ketara (kepak belakang yang panjang berbanding dengan spesies lain). Tiga ujian dataset kemudiannya digunakan untuk menilai prestasi DAISY: (i) imej latihan diputarbelitkan (ii) imej dari sumber internet spesies sama dalam dataset latihan (superfamili Bombycoidea), (iii) imej dari jilid The Moth of Borneo lain iaitu spesies yang tidak dimasukkan dalam dataset latihan, dari famili; Notodontidae, Lymantriidae, Arctiidae, Drepaninae, Callidulidae, Geometridae, Notuidae dan Noctuidae) dan (iv) imej dalam sampel lalai DAISY iaitu spesies tidak dalam dataset latihan (Belize sphingids). Saya mengklasifikasikan keputusan pengenalan DAISY kepada empat kategori untuk analisis spesies; (i) ujian dalam set latihan dan spesies betul diberikan (benar positif, TP), (ii) ujian spesies mungkin atau mungkin tidak dalam latihan yang ditetapkan tetapi nama

spesies tidak betul diberikan (palsu positif, FP), (iii) spesies ujian tidak dalam set latihan dan tiada nama spesies diberikan (benar No- ID, TNI) dan (iv) Ujian spesies dalam set latihan tetapi tiada nama spesies diberikan (palsu No- ID, FNI). Berdasarkan kriteria ini, saya mengukur ketepatan sebagai $TP / (TP + FP)$ dan accuracy keseluruhan $(TP + TNI) / (TP + TNI + FP + FNI)$. Secara keseluruhan, ketepatan DAISY di semua empat ujian dataset ialah 51% manakala ketepatan keseluruhan adalah 50%. Saya membincangkan sebab-sebab yang berpotensi untuk ketepatan observed yang rendah. Akhir sekali, saya memberi cadangan ciri-ciri yang perlu dipertimbangkan apabila mereka bentuk alat automatik untuk mengenal pasti kupu-kupu.

ACKNOWLEDGEMENTS

I would like to express my appreciation and gratitude to my supervisor, Dr. Sarinder Kaur Kashmir Singh, my co-supervisor Professor Dr. Susan Lim Lee Hong and for their invaluable advice, guidance, patience and constructive criticism throughout the course.

Special thanks to my consultant, Dato' Henry Barlow for contributing his expertise in moths and allowing us to use the moths of Borneo for this course.

I would also like to thank my family: Mrs. Woon Huey Sam, Ms. Nikki Leong, Ms. Yvonne Leong and Mr. Denver Tan for their unconditional love, support and encouragement. Without them, I would not have succeeded in completing this course.

I am also grateful to fellow friends: Ms. Vivien Tan Mui Koon, Ms. Kwong Soke Chee, and Mr. Gary Sing for their assistance and assurance when I am in doubt. Thank you for giving me the strength to complete this course.

This study was conducted using financial assistance from University of Malaya (University Malaya Research Grant (UMRG), project number FG 053/098US).

TABLE OF CONTENTS

ABSTRACT	iii
ABSTRAK	v
ACKNOWLEDGEMENTS	vii
LIST OF FIGURES	x
LIST OF TABLES	xii
LIST OF SYMBOLS AND ABBREVIATIONS	xiii
CHAPTER 1 : INTRODUCTION	1
1.1 Background	1
1.2 Research Objective	3
CHAPTER 2 : LITERATURE REVIEW	4
2.1 Computer-assisted taxonomy	4
2.2 Automated Bee Identification System (ABIS)	4
2.3 Digital Automated Identification System (DAISY)	5
2.4 Species Identification Automated (SPIDA)	10
2.5 Moths	11
CHAPTER 3 : MATERIALS AND METHOD	11
3.1 Image acquisition	12
3.2 Pre-processing of images	14
3.3 Species identification using DAISY	16
CHAPTER 4 : RESULTS	17
4.1 Distorted training images (Test 1)	17

4.2 Images of training dataset species from online databases (Test 2)	18
4.3 Image from species not present in the training dataset (Test 3)	20
4.4 Moth images from Genting Highlands moth collection (Test 4)	22
CHAPTER 5 : DISCUSSION	23
CHAPTER 6 : CONCLUSION	26
REFERENCES	27
APPENDIX 1	32
APPENDIX 2	37
APPENDIX 3	44
APPENDIX 4	52
APPENDIX 5	53
APPENDIX 6	58
APPENDIX 7	59
APPENDIX 8	60

LIST OF FIGURES

Figure		Page
Figure 3.1	An example of horizontal elongation of <i>Trabala ganesha</i> . The width of the moth relative to height was increased.	13
Figure 3.2	Duplication of the wing of <i>Trabala ganesha</i> . The original image was incomplete. The left wing portion was duplicated, flipped vertically and attached to the right side to form an even and whole image.	13
Figure 4.1	Different categories of identification provided by DAISY II for test images of moth species present in the training dataset taken from the Digital Moth of Japan website and the Barcode of Life Data Systems.	18
Figure 4.2	An image of <i>Brahmaeidae hearseyi</i> from Moths of Japan website which was identified correctly by DAISY.	19
Figure 4.3	An image of <i>Bombyx incomposita</i> from Moths of Japan website which DAISY returned as “not classifiable”.	19
Figure 4.4	Different categories of identification provided by DAISY II for test images of moth species not present in the training dataset taken from the other volumes of Moth of Borneo (Part 4, 5, 7, 8, 9, 12, 15 & 16) and Belize Sphingids in DAISY II.	21
Figure 4.5	An example of two incorrect identifications of images from Belize. <i>Protambulyx strigilis</i> (a) is followed by <i>Pachylioides resumens</i> (b). Both were incorrectly identified as <i>Ambulyx obliterata</i> (c). Notice that the wing shapes of both species are	21

very similar to *Ambulyx obliterata* from Sphingidae family.

Figure 4.6 Different categories of identification provided by DAISY II 22
for test images of moth species from Genting Highlands
collection.

LIST OF TABLES

Tables		Page
Table 3.1	Training and test datasets	13
Table 5.1	Summary of species identifications from all test datasets.	23

LIST OF SYMBOLS AND ABBREVIATIONS

ABIS	Automated Bee Identification System
ANN	Artificial Neural Network
BOLD	Barcode of Life Data Systems
DAISY	Digital Automated Identification System
DFE	DAISY Front End
DMJ	Digital Moths of Japan
DNA	DeoxyriboNucleic Acid
FNI	False No-ID
FP	False Positive
INBIO	Instituto Nacional de Biodiversidad
JPEG	Joint Photographic Experts Group
LDA	Linear Discriminant Analysis
No.	Number
No-ID	Not Identified/ Not Classifiable
NNC	Nearest-neighbour classification
NVD	Normalized Vector Difference
PCA	Principal Component Analysis
polyROI	Polygonal region of interest
PSOM	Plastic self-organizing map
SPIDA	Species Identification Automated
SVM	Support Vector Machine
TIFF	Tagged Image File Format
TNI	True No-ID
TP	True Positive

UK	United Kingdom
~	Approximately
%	Percentage

CHAPTER 1

INTRODUCTION

1.1 Background

Taxonomy is the basis for all scientific endeavors in biology and is particularly important in conservation planning. How do decision-makers decide which parts of the world or country should be labelled as protected areas if they do not know what can be found there to be protected? How does the agricultural industry sustainably combat invasive species if they cannot distinguish between harmful invasive species and harmless native species? How can we know that an organism belongs to a previously described species or is a novel species? How can we have knowledge about what species existed in the past, in the present and might be extinct in the future? What is the diversity of species in a certain area? All these questions can be answered by the knowledge we gain from taxonomy. In order to have useful taxonomies, we need taxonomists (Convention on Biological Diversity, 2012). However in recent years, the number of taxonomists has been declining due to retirement as well as lack of funding. Taxonomists have to go through many years of training and exposure to gain expertise and their research is usually undervalued and under appreciated.

Malaysia is one of the world's top twelve mega-diversity countries, rich in biodiversity resources, and has a National Biodiversity Index value of 0.809 (1.000 maximum) and 15,584km² of protected areas (CBD Secretariat, 2001), however, the species diversity is understudied and largely unknown. This could be due to lack of taxonomists and lack of funds for taxonomic research. Alternatives have been suggested to reduce the workload expert taxonomists and to increase taxonomic capacity to a

certain level. However it should be noted that ‘Unless the technology has advanced to a stage where the artificial intelligence has the ability to work like a human brain and take over decision making, we still need trained human taxonomists to point the way and to confirm results’ (Lim and Gibson, 2010).

Automated identification systems have been introduced to assist biologists in identifying organisms. Current automated identification systems are mostly based on either pattern or image recognition. Digital Automated Identification System (DAISY) (Hall and O’Neill, 2004), Species Identification Automated (SPIDA) (Russell et al., 2005) and Automated Bee Identification System (ABIS) (Arbuckle et al., 2001) are some of the existing image recognition systems. DAISY has been tested on various taxonomic groups such as bumble bees from Britain (Pajak, 2000), Hawkmoths from Costa Rica (Gaston and O’Neill, 2004) and British Moths (Watson et al., 2004), while SPIDA has been tested mainly on ground spiders from Australia (Russell et al., 2005), and ABIS has been tested on bees (Arbuckle et al., 2001). Both DAISY and SPIDA uses artificial neural network models while ABIS uses a support vector machine model. Watson et al. (2003) performed training and testing of DAISY using the right, dorsal forewings of live Macrolepidoptera, light trapped in Treborth Botanical Garden, Gwynedd, Wales. Given the reported success (83%) of DAISY with moth identification in the past studies of British moths (Watson et al., 2004) a license for DAISY was purchased from Tumbling Dice Pte (2009) to conduct this study. DAISY II, 2009 was the version available at time of purchase.

1.2 Research Objective

In this study, I aim to investigate the ability of DAISY to accurately assign unknown organisms to a species. Specifically, I plan to test the ability of DAISY to accurately assign Bombycoidea moths from Malaysian Borneo, to species.

CHAPTER 2

LITERATURE REVIEW

2.1 Computer-assisted taxonomy

A current trend in species identification technologies includes automated identification systems that are based on pattern recognition (Macleod, 2008). Research on image databases have been stimulated mostly by the increasing popularity of digital images. The leading image recognition systems that are currently available for species identification purposes include Automated Bee Identification System (ABIS), Digital Automated Identification System (DAISY) and Species Identification Automated (SPIDA).

2.2 Automated Bee Identification System (ABIS)

Automated Bee Identification System (ABIS) was funded by the German Research Council (1996 until 2001) and the German Ministry of Education and Research (2000 until 2003) (Arbuckle et al., 2001). The aim of this project was to develop an automated identification system for bee species using image analysis of wings (Arbuckle et al., 2001). This system was developed using geometrical image analysis, template matching, affine projection, discriminant analysis, kernel functions and Support Vector Machine models (SVMs) (Roth and Steinhage, 1999). ABIS is a knowledge-based image analysis of bee forewings and uses nonlinear discriminant analysis for identification (Arbuckle et al., 2001). Images of bee forewings were captured and input into ABIS. ABIS first detects the lines and intersections within the image. It then uses the information to search for three basal cells. A best first search will be conducted to detect the three cells. Once found, numerical features, employed to guide the future search, will be calculated.

These features are passed to a linear classification algorithm (Linear Discriminant Analysis, LDA) which allows the loading of a deformable template particular to a bee family. Using the hints from the template, the remaining cells are found. A new set of vectors will be calculated using the complete set of cells. Non-linear statistical methods such as SVM and Kernel Discriminant Analysis are applied to carry out the final recognition phase with the extended vectors.

2.3 Digital Automated Identification System (DAISY)

DAISY is a generic pattern matching system (Weeks et al., 1999). The development of DAISY was supported by United Kingdom Government's Darwin Initiative and the Biotechnology and Biological Sciences Research Council. DAISY was designed to enable automated identification of organisms to species level. Previously, this system has been tested on British bumblebees (Pajak, 2000), British moths (Watson et al., 2004), Costa Rican hawkmoths (O'Neill, 2010), Costa Rican parasitic wasps (O'Neill, 2010), palaeartic biting midges (Gauld et al., 2000), British butterflies (O'Neill, 2010), and lycosid spiders (O'Neill, 2010). DAISY was built in a Linux environment using artificial neural network (ANN) model and is capable of dynamic learning. ANNs are the most common approach used in pattern recognition. The information-processing structures of ANN are modelled after the immensely parallel structure of the brain (Gaston and O'Neill, 2004). ANN is a network which consists of interconnecting layers of nodes and they are not rule based, but trained on examples of the taxa to be identified, an iterative process that may be time-consuming as the internal network organization is reformed until it can successfully discriminate between these taxa (Gaston and O'Neill, 2004). The two main classes of ANNs that have been used in automated identification systems are supervised ANNs and unsupervised ANNs (PSOM) (Gaston and O'Neill, 2004). DAISY uses a hybrid identification scheme which consists of a Lucas continuous

n-tuple classifier and a plastic self-organizing map (PSOM). An n-tuple classifier works by comparing an unknown image with the trained images while PSOM is a variant of the continuous n-tuple classifier that can learn dynamically (Gaston and O'Neill, 2004).

In brief, there are four main components of DAISY with the DAISY front end (DFE) being the first component. DFE is the graphical user interface for the pattern recognition system and it provides a “point and click” style interface. DFE was based on the GTX+/Gnome X toolkit (O'Neill and Hilgerag, 2001). The next component is the ipm (image process management) which normalizes the input image and re-samples it to a standard size prior to identification. Region of interest in the images are extracted and passed to the next component which is the floret. This application is multi-threaded and is responsible for the identification of unknowns. The floret will recognise if the unknown can be identified and is sent to vhtml application for further process. Otherwise the unknown image is simply discarded. The last component is the vhtml which uses the identification tags to check through a database for information available on the species. Information about new species cannot be added to the vhtml database by the user, as it is currently restricted to the developer.

The first version of DAISY was developed based on Principle Component Analysis (PCA). In 1995, DAISY was built for the purpose of identifying insects using images of wings. The optical imagery of insects' wings help to reduce the problem to one of a tractable size. Using wings, which are two dimensional by nature, simplifies the preprocessing stage as perspective need not be considered. Furthermore, wings can be robustly displayed on slides for analysis. Wings can be easily detached from the body of insects, be it in the form of museum training materials or specimens caught in the field. The decision to use the wings for identification was also influenced by a computer-based identification model suggested by Weeks et al. (1997).

DAISY's initial implementation was mainly influenced by Turk and Pentland (1991). The (PCA)-based DAISY prototype based on Turk and Pentland's algorithm accurately identified five species of parasitic wasps after being trained with datasets of 8-10 wings (specimens) per species (Weeks et al., 1997). A similar result was achieved with 30 biting midge species (Weeks et al., 1999). These identifications were said to achieve more than 95% accuracy for biting midges using wings which were mounted in slides. The PCA-based approach also successfully precluded the time-consuming manual measuring of wing characters that were done in other systems originating from wing venation geometry (Lanne, 1981; Yu et al., 1992). In short, DAISY had the advantage of user-friendliness over other systems and as such, more funding was pursued and attained from UK government's Darwin initiative in collaboration with the University of Costa Rica and INBIO to build a fully functional DAISY prototype.

The first test was conducted to identify the challenging, yet biologically important, parasitic wasps in the genus *Enicospilus*, from Costa Rica. With some 50 species, this genus was chosen because it is relatively speciose the species are all extremely alike in appearance and have complicated wing venation, resulting in difficulties for non-taxonomists or non-biologists to distinguish between the species. This makes the genus a challenging experimental case for an automated species identification system. Thus, there are many wing-based characters that might be useful in diagnosing species. And so, the proficiency to identify members of *Enicospilus* with speed and accuracy will definitely benefit the ecologists.

PCA-based DAISY was able to distinguish 84% of the members of this challenging group. Linearity was one of the problems that were highlighted as a shortcoming of the system. The necessity to repopulate during the transformation from the high-dimensional pattern space to a low-dimensional component space whenever

new material is added or existing material is taken out from the system was another major setback. In circumstances where the system is non-closed and training data set are added or taken out regularly, PCA approach becomes a very expensive option. The rate at which the system computes principal components is $O(n^2)$, where n equals the number of taxa in the system, making the matter worse. This means that the period required to train the system escalate proportionately to the square of the total of species it encompasses. Definitely, this approach is not a fitting for van all-purpose system that is able to scale tens of thousands of taxa.

A second version of DAISY (DAISY II) was intended to mitigate the non-linearity and scaling issues of the PCA-based approach. The new version was built using a nearest-neighbour classification model (NNC) which is a simple yet effective classification scheme, first promoted by Aleksander and Stonham (1979). Even though it is easy to understand and execute the nearest-neighbour classification, algorithms that originated from nearest-neighbour classification delivered a $> 95\%$ accuracy compared to more complicated pattern-matching algorithms (Lucas, 1997).

“In its simplest form, NNC reduces to comparing an unknown (U) with a set of pattern vectors (P). If U is correlated with each member P_1 of (P), U is assumed to belong to the same class as pattern vector P_m for which it has the highest correlation affinity a_{nm} (where n is the number of taxa in the system and m is the number of elements in the pattern vectors compared). This is equivalent to a so-called “first-past-the-post” (FPTP) classification” (O’Neill, 2007).

NNC has significant advantages over PCA including the repopulation from high-dimensionality to low-dimensionality whenever material is added or removed from the training sets. This is because NNC effectively trains in real-time like that of Kohonen

self-organizing maps (Kohonen, 2001) and the closely related plastic self-organizing map (PSOM) (Lang and Warwick, 2002). NNC also has the advantage that it is ready to incorporate dynamic learning such as that used in the PSOM algorithm by Lang and Warwick (2002). Additionally, NNC can easily be deployed in hardware and has a simple, robust and linear nature. The linear nature makes NNC well suited to geometric parallelization. Ideal usages of computing resources are obtainable when these natures are manipulated within NNC/NVD DAISY to allow the scaling of tens of thousands of taxa seamlessly. Besides that, NNC is also able to cope well with non-linear morphospaces.

The typical cross-correlation was not used by DAISY II when computing the pattern-pattern correlation affinity. This is because DAISY II wanted to maximize the functionality of NNC algorithm. As such, a novel form of pattern-pattern correlation named the normalized vector difference (NVD) was developed. NVD was found to be a statistically more favourable method to compare patterns (O'Neill, 2007). NVD offers quicker computation and superior results to traditional cross-correlation for patterns with high similarity. DAISY II was tested on a broad range of pattern-recognition problems with good success, including identification of insects using wing shape and pattern (Gauld et al., 2000; Watson et al., 2003), identification of planktonic microfossils (MacLeod et al., 2003), vertebrate bones (Walsh et al., 2004) and categorisation of functional morphological (MacLeod et al., 2004).

DAISY was used in this study because it is a generic system, was used previously on similar taxa, and has the advantage over ABIS in that it does not require microscopic images.

2.4 Species Identification Automated (SPIDA)

Species Identification Automated is an internet-accessible automated identification system based on an artificial neural network model. This project was funded by the National Science Foundation Biocomplexity in the Environment Program: Instrumentation Development for Environmental Activities (Russell et al., 2008). SPIDA was developed with the aim to identify any arachnid species in a given family. Identifications of species uses digital images encoded for input to the artificial neural network using wavelet transformation. Spiders of order Araneae was used as the test group for the prototype of SPIDA. Images submitted to the system need to be cropped to a square, 51X51 pixels, converted to grayscale, and be in TIFF or JPEG format. An identification system for Australian ground spiders was developed using images of the external genitalia (ventral view of the female epigynum, ventral and retro lateral views of the male palp). Users must position the images according to the standard in SPIDA. SPIDA was tested using 15 genera consisting of 121 species from family Trochanteriidae. The accuracy of identification was related to the quality and quantity of the images used as training set (Russell et al., 2005). SPIDA had an accuracy of 81% (Russell et al., 2005).

2.5 Moths

Moths are insects in the order Lepidoptera of the class Insecta. Moths are commonly divided into two groups, macro-moths and micro-moths (Beadle and Leckie, 2012). Many moths are important agricultural pests but also play an important role in pollination of flowers (Balmford and Bond, 2005).

Moths are often used as biodiversity indicator species as they are sensitive to environmental changes and are widespread in many different habitats. Monitoring their numbers and ranges can give us vital clues to changes in our environment such as the effects of new farming practices, pesticides, air pollution and climate change (Balmford and Bond, 2005). Moths are hard to identify as there are potentially more than 1000 species present in any given area (Beadle and Leckie, 2012).

CHAPTER 3

MATERIALS AND METHOD

3.1 Image acquisition

DAISY identifies an unknown by matching the unknown image to images that the system “knows” prior to the identification. Therefore a training dataset must be created in order for DAISY to identify an image (Hall and O’Neill, 2004). Images from the “The Moths of Borneo: Part 3: Lasiocampidae, Eupterotidae, Bombycidae, Brahmaeidae, Saturniidae, Sphingidae” by J. D. Holloway (1998) containing Bombycoidea were used as the training dataset. This superfamily was chosen as the training dataset because it includes species with a wide variety of wing shapes.

The moth images used as the training dataset in this study were scanned and digitized from “The Moths of Borneo: Part 3: Lasiocampidae, Eupterotidae, Bombycidae, Brahmaeidae, Saturniidae, Sphingidae” by J.D. Holloway (1998). HP Scanjet 5590 was used to digitize the images from the book while Adobe Photoshop CS5 (Adobe, 2011) was used to pre-process them. DAISY was installed on a HP computer with Intel core duo 2 to conduct this study.

Images used as test datasets were sourced from (i) Digital Moths of Japan website (KeiHiroshi et al., 2012) and the Barcode of Life Datasystems (Ratnasingham and Hebert, 2007), (ii) the Moths of Borneo (Part 4-16) website (Holloway, 2010) and Belize Sphingids in the DAISY II default training set, and (iv) moth collected at Genting Highlands, Pahang in April, 2012 (Table 3.1).

Images from the training dataset were distorted and used as an additional test dataset. The images were elongated horizontally and vertically up to 25% to 30% using Adobe Photoshop CS5 (Adobe, 2011) (Fig.3.1).

Table 3.1: Training and test datasets

Sources	Family	Genus	Species	Images	Used As
Moths of Borneo (Part 3)	6	74	210	273	Training
Moths of Borneo (Part 3) Distorted	6	74	210	273	Test 1
Digital Moths of Japan website	5	16	28	33	Test 2
Moth collection in BOLD website	6	69	148	280	Test 2
Moths of Borneo (Part 4-16)	8	21	30	30	Test 3
Belize Sphingids in DAISY II	1	20	55	206	Test 3
Genting Moth Collections	3	8	8	10	Test 4

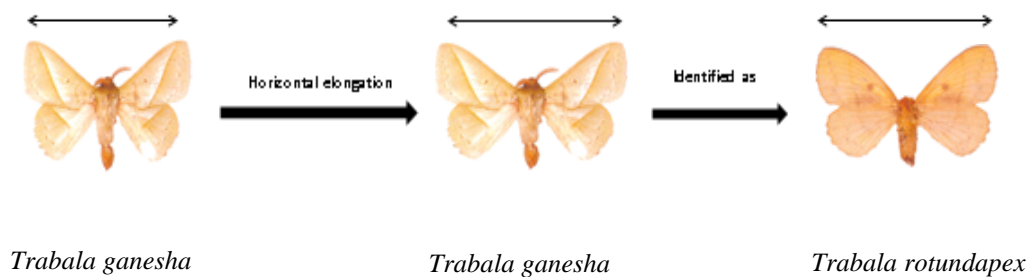


Figure 3.1: An example of horizontal elongation of *Trabala ganessa*. The width of the moth relative to the height was increased.

3.2 Pre-processing of images

Duplication of wings and standardization of the images were performed using Adobe Photoshop CS5 (Adobe, 2011). Duplication of wings ensures that the most representative wing is used as the training image (Fig. 3.2). When building a training set, it is important to decide which orientation of the wing is to be used, and it should be standardized throughout the training sets to simplify the usage later on. DAISY does not limit the training image to only one wing. However in this study, training was limited to one wing due to the time constraints and large number of species used. All images (training and test datasets) were converted into TIF files as required by DAISY. The image size was also standardized to 600X480 pixels, as this size fits the DAISY window.



Figure 3.2: Duplication of the wing of *Trabala ganeshha*. The original image was incomplete. The left wing portion was duplicated, flipped vertically and attached to the right side to form an even and whole image.

Images were imported into DAISY and the wing shape was pinpointed by creating a polygonal region of interest around the image using the polyROI function from DAISY's panel. When the wing shape was highlighted, it was important to ensure that the amount of space around the area of interest was as small as possible because the identification can be affected by the background. Pinpointing of wing shapes was done manually one image at a time. There are two ways to draw the PolyROI. The first is by using the stylus to trace out the wing area of the moths (continuous overlay mode) while the second is by using the point clicking of a mouse. For this study, I used the point

clicking of a mouse because it takes more time to master the art of stylus. Both the methods give the same result (O'Neill and Hilgerag, 2001). These images were used to train the artificial neural network of DAISY. Each species needs at least 10 images (different or duplicated) with wings pinpointed to build the training set. For this study, images were duplicated as the reference book only featured one image for each species.

3.3 Species identification using DAISY

Test images were imported and opened in DAISY II. The polyROI function was used to highlight the wing region of interest. The results of identification using DAISY II were summarized into four categories, (a) True Positive (TP) if the test species is in training set and is correctly identified, (b) True No-ID (TNI) if test species is not in training set and is (correctly) not identified by DAISY II, (c) False Positive (FP) if test species is identified to an incorrect species, (d) False No-ID (FNI) if test species is in training set but is not identified by DAISY II.

Precision, the fraction of queries identified to a species that belongs there and overall accuracy, the proportion of queries placed without error, were used to analyse the results. Precision was calculated using formula (1) whereas overall accuracy for each of the dataset was calculated using formula (2). ;

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \quad (1)$$

$$\text{Overall accuracy} = (\text{TP} + \text{TNI}) / (\text{TP} + \text{TNI} + \text{FP} + \text{FNI}) \quad (2)$$

CHAPTER 4

RESULTS

4.1 Distorted training images (Test 1)

Test done using the distorted training images returned a total of 257 correct identifications (True Positive) out of the 273 images tested. Thirteen images were identified wrongly as another species (False Positive) while the remaining three were not identified (False no-ID) (Appendix 2). Note that the original undistorted images were tested and all were identified correctly.

4.2 Images of training dataset species from online databases (Test 2)

Three hundred and thirteen moth images of species present in the training dataset were collected from the Digital Moths of Japan website (KeiHiroshi et al., 2012) and the Barcode of Life Data Systems (Ratnasingham and Herbert, 2007). These images were also used as tests for DAISY II. DAISY II was able to identify 127 images correctly (True Positive, e.g. Fig. 4.2), 37 images were not identifiable (False no-ID, e.g. Fig. 4.3) and 149 images were identified as an incorrect species (False Positive) (Fig. 4.1) (Appendix 3).

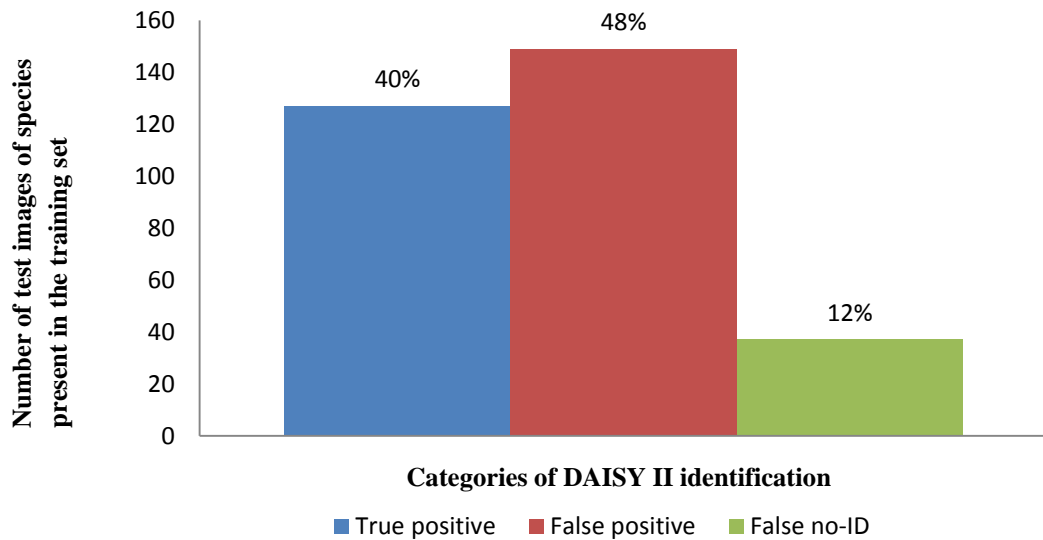


Figure 4.1: Different categories of identification provided by DAISY II for test images of moth species present in the training dataset taken from the Digital Moth of Japan website and the Barcode of Life Datasystems.



Figure 4.2: An image of *Brahmaeidae hearseyi* from Moths of Japan website which was identified correctly by DAISY.



Figure 4.3: An image of *Bombyx incomposita* from Moths of Japan website which DAISY returned as “not classifiable”.

4.3 Images from species not present in the training dataset (Test 3)

Images from other volumes of Moth of Borneo (Part 4, 5, 7, 8, 9, 12, 15 & 16) (Holloway, 2010) which cover different families (i.e. species not present in the training dataset) were also used as test datasets. All images from the volumes (were given numbers and 30 images were randomly selected by drawing numbers). Of the 30 images submitted to DAISY II for identification, seven were correctly returned as unclassifiable (True no-ID) and 23 were identified incorrectly (False Positive) (Appendix 4).

Belize sphingids in DAISY II default training set were also used as a test dataset. These species are from a family represented in the training dataset, but these are images of species not present in the training dataset. Out of 206 images from Belize, 185 were given a species identification (False Positive e.g. Fig. 4.4) and 21 were correctly given no-ID (True no-ID). In total for test 3 (i.e. species not present in the training dataset) 88% were given a species name by DAISY (False positive) (Fig. 4.5).

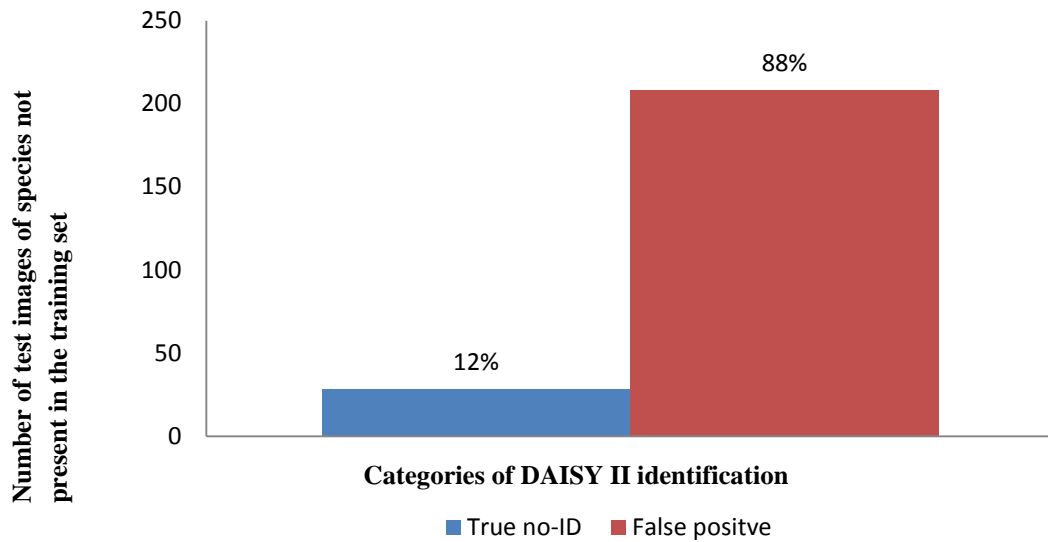


Figure 4.4: Different categories of identification provided by DAISY II for test images of moth species not present in the training dataset taken from the other volumes of Moth of Borneo (Part 4, 5, 7, 8, 9, 12, 15 & 16) and Belize Sphingids in DAISY II.



Figure 4.5: An example of two incorrect identifications of images from Belize. *Protambulyx strigilis* (a) is followed by *Pachylioides resumens* (b). Both were incorrectly identified as *Ambulyx obliterate* (c). Notice that the wing shapes of both species are very similar to *Ambulyx obliterate* from Sphingidae family.

4.4 Moths images from Genting Highlands moth collection (Test 4)

Moths were captured in Genting Highlands, Pahang (some of the species were present in training dataset and some were not present). These moths were identified independently by DNA barcoding (Reference) (Ratnasingham and Hebert, 2007). Out of the ten images tested, only two were correct (true no-ID and true positive) and eight were misidentified (false positive) (Fig. 4.6).

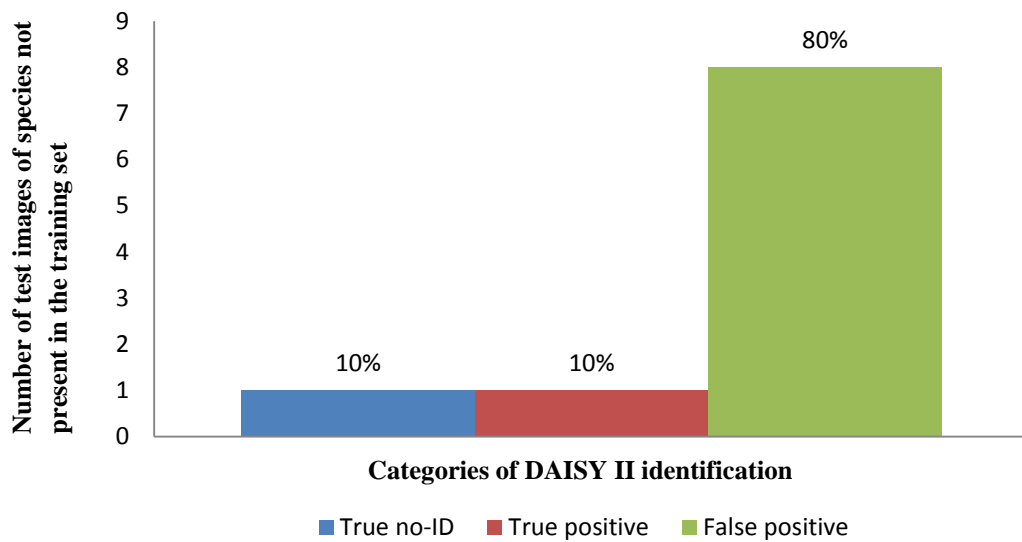


Figure 4.6: Different categories of identification provided by DAISY II for test images of moth species from Genting Highlands collection.

CHAPTER 5

DISCUSSION

All identification systems strive for making accurate identification processes automatic. However, currently there are no fully automated systems as images are too complex and there is no general method to solve all problems (Zhang, 1996). DAISY II, one of available alternatives, was tested in this study to identify Bombycoidea moths to species. Testing was done using (i) images from species present in training dataset and (ii) images from species not present in training dataset. Accuracy of DAISY was tested in (i) and the ability of DAISY II to detect novel species was tested in (ii). In this study the level of precision and overall accuracy for the identifications by DAISY II were 51% and 50% respectively (Table 5.1). This was much lower than my expectation, based on the reported successes of DAISY II in previous studies (e.g. Watson et al. 2003).

Table 5.1: Summary of species identifications from all test datasets.

Dataset	No. of images	TP	FP	TNI	FN I	Precision	Overall accuracy
Distorted Moths of Borneo (Part 3)	273	257	13	0	3	0.95	0.94
Digital Moths of Japan	33	11	17	0	5	0.39	0.33
The Barcode of Life	280	116	132	0	32	0.47	0.41
Moths of Borneo (Part 4-16)	30	0	23	7	0	0.00	0.23
Belize Sphingids (DAISY)	206	0	185	21	0	0.00	0.10
Moths from Genting Highlands	10	1	8	1	0	0.11	0.20
Combined	832	384	370	28	40	0.51	0.50

In my opinion, the low accuracy was caused by the number of training images used. In Watson et al. (2003), their training set consisted of 10-15 different training images for each species, while I only used one duplicated training image for each species. A variety of training images allows the system to factor intraspecific variation, potentially improving the accuracy of the system.

Another factor that might have contributed to the low precision and accuracy was the low variation of wing positioning in the training dataset. All the training images consist of specimens with their wings spread nicely and which had been carefully prepared by drying on a board. Some of the images in the test datasets were captured from live moths in natural resting posture.

It is also worth noting that in this study the training dataset consisted of 210 species, compared to the study by Watson et al. (2003) based on 35 species. The number of species used increased the possibility of misidentification by DAISY as moths' species in a particular family tend to have very similar wing shapes, although they belong to different species (Holloway, 1998).

Based on the results from this study, I believe DAISY II is not very suitable for species identification. I suggest DAISY II be further tested on genus or family level identifications as it may be able to give higher accuracy at a more inclusive taxonomic level.

Based on my findings, I suggest future work be focussed on the effects of adding images of moths in their natural habitat to the training dataset. Otherwise I recommend that or only slide mounted moth images are suitable subjects for identification by DAISY II. This may have some use in a museum setting.

It is obvious that by using wing shapes as the only feature in identification might not be sufficient for species identification. As DAISY II only identified ~49.5% test images correctly, future work can also be conducted on the wing pattern and genitalia of moths as they are two of the important morphological characteristics for species identification. Although moth genitalia part is an important characteristic, it requires dissection of the moths, which is a specialized skill (Holloway, 2008).

The accuracy of DAISY could also potentially be improved by examining two regions of interest in the image, instead of restricting the system to one.

When comparing DAISY to SPIDA and ABIS, I think ABIS could have done better as it reported the highest percentage of accuracy (>90%) by using wing venation of bees. However, if my aim was to distinguish between training dataset species and novel species, SPIDA could have been a better choice as SPIDA reported 95% accuracy when identifying novel species from training species.

Although automated image identification is one of the alternatives, there are also other automated and semi-automated tools available to assist humans in the identification of species. Lucid keys allows expert knowledge to be replicated and shared via CD or on the internet (Trigg et al., 2013) and DNA barcoding, uses a standardized segment of DNA to identify species (Ratnasingham and Herbert, 2007).

CHAPTER 6

CONCLUSION

In conclusion, DAISY is generic system that returned nearest possible species and this is a hindrance especially to distinguish between an existing species and a novel species. It is also not suitable to be used in moth species identification, perhaps better on genus or family level.

From the findings of this study, I conclude that a good species identification system should be specific to a particular organism, must be able to distinguish between novel and existing species, user-friendly and allow users to select more than one region of interest for identification at one time. Additionally, I suggest that much more research activities should be geared up towards evaluating elements of artificial intelligence in finding the most suitable algorithm that returns an accuracy of almost 100%.

As we are running out of biodiversity specialists, it is timely that we train more taxonomists or perhaps make use of the digital technology in such a way that we can ease the routine burden of taxonomists.

REFERENCES

- Adobe. (2011). Adobe Photoshop CS5. <http://www.photoshopsupport.com/photoshop-cs5/free-trial/photoshop-cs5-free-trial.html>
- Aleksander, I., & Stonham, T. J. (1979). Guide to pattern recognition using the random-access memories. *Computers and Digital Techniques*, 2, 29-40.
- Arbuckle, T., Schröder, S., Steinhage, V. & Wittmann, D. (2001). *Biodiversity Informatics in Action: Identification and Monitoring of Bee Species using ABIS*. Proceedings of 15th International Symposium Informatics for Environmental Protection, Zürich.
- Balmford, A. & Bond, W. (2005). Trends in the state of nature and their implications for human well being. *Ecology Letters*, 8, 1218–1234.
- Beadle, D. & Leckie, S. (2012). Peterson Field Guide to Moths of Northeastern North America. *Houghton Mifflin Harcourt*.
- Butterfly Conservation. (2012). Importance of Moths. <http://www.mothscount.org> CBD Secretariat. (2001). Global Biodiversity Outlook 1, Montreal. <http://www.cbd.int/gbo1/>
- Convention on biological diversity. (2012) Why is Taxonomy Important. <http://www.cbd.int/gti/importance.shtml>
- Gaston, K. J., & O'Neill, M. A. (2004). Automated species identification: why not?. *Philos Trans R Soc Lond B Biol Sci.*, 359(1444), 655–667.
- Gauld, I.D., O'Neill, M.A., & Gaston, K.J. (2000). Driving Miss DAISY: the performance of an automated insect identification system. In: A.D. Austin & M. Dowton (Eds.), *Hymenoptera: Evolution, Biodiversity and Biological Control*. Canberra: CSIRO.

- Hall, Z. E., & O'Neill, M. A. (2004). *Daisy User Manual*. London.
- Holloway, J. D. (2010). Moths of Borneo. <http://www.mothsofborneo.com/>.
- Holloway, J. D. (1998). The Moths Of Borneo Part 3 Superfamily Bombycoidea: Families Lasiocampidae, Eupterotidae, Bombycidae, Brahmaeidae, Saturniidae, Sphingidae. *Southdene Sdn. Bhd.* (Vol. 3).
- KeiHiroshi, Y., Teruhiko, F., Jinbo, S. & Nakao, K . (2012). Digital Moths of Japan. <http://www.jpmoth.org/~dmoth/list.htm>
- Kohonen, T. (2001). Self- Organising Maps. *Information Sciences*, 30 (3), 501.
- Lang, R., & Warwick, K. (2002). The plastic self organising map. <http://www.cyber.rdg.ac.uk/research/publications/01183.pdf>
- Lanne, R.P. (1981). Allometry in size in three species of biting midges (Diptera, Ceratopogonidae, Culicoides). *Journal of Natural History*, 15, 775-788.
- Lim, L. H. S., & Gibson, D. I. (2010). Taxonomy, Taxonomists & Biodiversity. In: Manurung, R., Zaliha, C., Abdullah, Z.C., Ahmad, F.B. & Kuek, C. Biodiversity-biotechnology: Gateway to discoveries, sustainable utilization and wealth creation. Proceedings of the Internaional Symposium held in Kuching, Sarawak, Malaysia, 10th.-21st. November 2008. The Sarawak Biodiversity Centre, Kuching. 458 pp.
- Lucas, S.M. (1997). Face recognition with the continuous n-tuple classifier. Paper presented at the Proceedings of the British Machine Vision Conference.
- MacLeod, N., O'Neill, M.A., & Walsh, S.A. (2004). A comparison between morphometric and unsupervised, articial, neural-net approaches to automated species identification in foraminifera. Paper presented at the Proceedings of the Micropalaeontological Society Calcareous Plankton Spring Meeting.

- MacLeod, N. (2008). *Automated Taxon Identification in Systematics: Theory, Approaches and Applications*. CRC Press.
- MacLeod, N., O'Neill, M.A., & Walsh, S.A. (2003). PaleoDAISY: An integrated and adaptive system for the automated recognition of fossil species. *The Geological Society of America, Abstracts and Programs*, 35, 316.
- O'Neill, M. A. (2007). DAISY: A Practical Computer-Based Tool for Semi-Automated Species Identification. *Automated Taxon Identification in Systematics: Theory, Approaches and Applications*, 101 - 114.
- O'Neill, M. A. (2010). DAISY: A Practical Tool for Automated Species Identification. <http://www.tumblingdice.co.uk/daisy/doverview.pdf>.
- O'Neill, M.A., & Hilgerag, C.C. (2001). The portable UNIX programming system [PUPS]: a computational environment for the dynamical representation and analysis of complex neurobiological data. *Philosophical Transactions of the Royal Society of London, Series B*, 356, 1259-1276.
- Pajak, M. (2000). *Identification of British Bombus and Megabombus using DAISY*, B.A. 3rd Year Honours Project, University of Oxford, United Kingdom.
- Ratnasingham, S., & Herbert, P. D. N. "BOLD: The Barcode of Life Data System (<http://www.barcodinglife.org>)." *Molecular ecology notes* 7.3 (2007): 355-364.
- Roth, V., & Steinhage, V. (1999). *Nonlinear discriminant analysis using kernel functions*. Paper presented at the Proceedings of 13th Annual Conference on Neural Information Processing Systems, NIPS.
- Russell, K. N., Do, M. T., & Plantick, N.I. (2005). *Introducing SPIDA-web: An automated identification system for biological species*. Paper presented at the Proceedings of TDWG, St Petersburg, Russia.

- Russell, K. N., Do M. T., & Plantick, N.I. (2008). Introducing Spida-Web: Wavelets, Neural Networks and Internet Accessibility in an Image-Based Automated Identification System. *Automated Taxon Identification in Systematics: Theory, Approaches and Applications*, 131-152.
- Trigg, H., Jacobucci, S., Henkel, M., & Steinberg, J. (2013). Human Impacts Pollen Database, an Illustrated Key. Andrew Fiske Center for Archaeological Research, University of Massachusetts Boston.
- Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3, 71-86.
- Walsh, S.A., MacLeod, N., & O'Neill, M.A. (2004). *Analysis os spheniseid tarsometatarsus and humerus morphological variability using DAISY automated digital image recognition*. Paper presented at the Proceedings of the 6th International Meeting of the Society of Avian Paleontology and Evolution, Quillan.
- Watson, A. T., O'Neill, M. A., & Kitching, I. J. (2003). A qualitative study investigating automated identification of living macrolepidoptera using the Digital Automated Identification System (DAISY). *Systematics & Biodiversity*, 1, 287-300.
- Watson, A. T., O'Neill, M. A., & Kitching, I. J. (2004). Automated identification of live moths (Macrolepidoptera) using digital automated identification System (DAISY). *Systematics and Biodiversity* 1(3), 287-300.
- Weeks, P.J.D, O'Neill, M.A, Gaston, K.J, & Gauld, I.D. (1999). Automating insect identification: exploring the limitations of a prototype system. *Journal of Applied Entomology*, 123, 1-8.

- Weeks, P.J.D., Gauld, I.D., Gaston, K.J., & O'Neill, M.A. (1997). Automating the identification of insects: a new solution to an old problem. *Bulletin of Entomological Research*, 87, 203-211.
- Yu, Kokko, D.S., Baron, E.G., Schaalje, G.B. & Gowen, B.E. (1992). Identification of Ichneumoid wasps using image analysis of wings. *Systematic Entomology*, 17, 389-395.
- Zhang, J. W. (1996). A classification and comparison of evaluation techniques for image segmentation. *Journal of Image and Graphics*, 1(2), 151-158.

APPENDIX 1

Table shows the detailed breakdown of training dataset.

Family	Genus	Species	Specimen
Bombycidae	Bombyx	Bombyx incomposita	1
	Gunda	Gunda ochracea	2
		Gunda javanica	2
		Gunda subnotata	2
	Ocinara	Ocinara albicollis	1
		Ocinara bifurcula	2
		Ocinara albiceps	1
	Trilocha	Trilocha friedeli	1
	Penicillifera	Penicillifera apicalis	1
		Penicillifera purpurascens	1
	Ernolatia	Ernolatia lida	1
Ernolatia moorei		1	
Mustilia	Mustilia dierli	1	
Andraca	Andraca apodecta	2	
Prismosticta	Prismosticta tiretta	1	
Brahmaeidae	Brahmaea	Brahmaea hearseyi	1
Eupterotidae	Eupterote	Eupterote multiarcuata	1
		Eupterote naessigi	1
		Eupterote obsoleta	1
		Eupterote niassana	1
		Eupterote asclepiades	1
		Eupterote muluana	2
		Eupterote harmani	2
		Ganisa	Ganisa plana
	Ganisa similis		1
	Pseudojana	Pseudojana perspicuifascia	1
		Pseudojana obscura	1
Melanothrix	Melanothrix nymphaliaria albidior	2	
	Melanothrix latevittata	2	
	Melanothrix fumosa	2	
	Melanothrix alternans	2	
Lasiocampidae	Bhima	Bhima borneana	1
	Metanastria	Metanastria gemella	2
	Lebeda	Lebeda cognata	2
		Lebeda brauni	1
		Lebeda pruetti	2
		Lebeda intermedia	1
	Paralebeda	Paralebeda lucifuga	2
Paralebeda uniformis		2	

Continue Appendix 1

Lasiocampidae	Streblote	Streblote helpsi	2
	Suana	Suana concolor	2
		Suana sundana	1
	Lajonquierea	Lajonquierea mediofasciata	2
		Lajonquierea derunoides	1
		Lajonquierea variabile	3
		Lajonquierea piccoloptera	1
		Lajonquierea jermyi	1
	Kunugia	Kunugia basimacula	1
		Kunugia rectifascia	1
		Kunugia leucopicta	2
		Kunugia suanoides	2
		Kunugia austroplacida	2
		Kunugia ferox	2
		Kunugia gynandra	2
		Kunugia quadrilineata	1
		Kunugia basinigra	1
		Kunugia drakei	2
	Takanea	Takanea diehli	2
	Arguda	Arguda rectilinea	2
		Arguda insulindiana	1
		Arguda rosemariae	1
		Arguda albiscutulata	1
	Radhica	Radhica holoxantha	1
		Radhica elisabethae	2
		Radhica himerta	1
	Syrastrena	Syrastrena sumatrana	1
		Syrastrena lanaoensis	1
Syrastrena tamsi		1	
Hallicarnia	Hallicarnia albipectus	2	
Odonestis	Odonestis erectilinea	1	
	Odonestis lipara	1	
	Odonestis vita	1	
	Odonestis angulata	1	
Odontocraspis	Odontocraspis hasora	1	
Chonopla	Chonopla tecta	1	
Alompra	Alompra roepkei	1	
	Alompra ferruginea bidiensis	1	
Trabala	Trabala irrorata	2	
	Trabala krishna	2	
	Trabala viridana	2	
	Trabala ganeshia	2	
	Trabala shiva	1	

Continue Appendix 1

Lasiocampidae	Trabala	Trabala pallida	3
		Trabala hantu	2
		Trabala garuda	2
		Trabala gautama	1
		Trabala rotundapex	2
		Trabala bouraq	1
Lasiocampidae	Gastropacha	Gastropacha leopoldi	2
	Paradoxopla	Paradoxopla cardinalis	1
	Micropacha	Micropacha roepkei	1
	Euthrix	Euthrix laeta	1
		Euthrix sp.	1
Saturniidae	Actias	Actias selene	1
		Actias maenas	2
	Antheraea	Antheraea helferi	2
		Antheraea diehli	2
		Antheraea larissa	1
		Antheraea assamensis	2
		Antheraea jana	2
		Antheraea korintjiana	1
		Antheraea celebensis	3
		Antheraea moultoni	4
		Antheraea alleni	3
		Antheraea brunei	2
		Antheraea rosieri	3
	Loepa	Loepa sikkima	1
		Loepa megacore	1
	Lemaireia	Lemaireia loepoides	1
	Cricula	Cricula trifenestrata	2
		Cricula bornea	1
		Cricula elaezia	1
	Attacus	Attacus atlas	1
Archaeoattacus	Archaeoattacus staudingeri	1	
Samia	Samia tetrica	1	
Sphingidae	Agrius	Agrius convolvuli	1
	Megacorma	Megacorma obliqua	1
	Acherontia	Acherontia lachesis	1
		Acherontia styx	1
	Meganoton	Meganoton analis	1
		Meganoton nyctiphanes	1
		Meganoton rufescens	1
	Psilogamma	Psilogamma menephron	2
	Dolbina	Dolbina krikkeni	1
Amplypterus	Amplypterus panopus	1	
Ambulyx	Ambulyx joiceyi	1	

Continue Appendix 1

Sphingidae	Ambulyx	Ambulyx obliterata	1
		Ambulyx substrigilis	1
		Ambulyx tattina	1
		Ambulyx pryeri	1
		Ambulyx clavata	1
		Ambulyx canescens	1
		Ambulyx subocellata	1
	Clanis	Clanis bilineata	1
		Clanis stenosema	1
	Marumba	Marumba cristata	1
		Marumba sperchius	1
		Marumba juvencus	1
		Marumba tigrina	1
		Marumba dyras	1
		Marumba spectabilis	1
	Daphnusa	Daphnusa ocellaris	1
	Cypa	Cypa decolor	2
	Smerinthulus	Smerinthulus terranea	1
		Smerinthulus diehli	1
		Smerinthulus quadripunctatus	1
	Degmaptera	Degmaptera olivacea	1
	Callambulyx	Callambulyx rubricosa	1
	Sataspes	Sataspes infernalis	1
	Cephonodes	Cephonodes picus	1
	Gnathothlibus	Gnathothlibus erotus	1
	Daphnis	Daphnis hypothous	1
		Daphnis placida	1
	Elibia	Elibia dolichus	1
	Acosmeryx	Acosmeryx anceus	1
		Acosmeryx shervillii	2
	Gehlenia	Gehlenia falcata	1
	Panacra	Panacra busiris	1
		Panacra variolosa	1
Panacra psaltria		1	
Panacra dohertyi		1	
Panacra automedon		1	
Panacra malayana		1	
Panacra mydon		1	
Angonyx	Angonyx testacea	1	
Enpinanga	Enpinanga vicens	2	
	Enpinanga borneensis	2	
Eurypteryx	Eurypteryx bhaga	1	
	Eurypteryx shelfordi	1	
	Eurypteryx falcata	1	
Giganteopalpus	Giganteopalpus mirabilis	1	

Continue Appendix 1

Macroglossum	Macroglossum lepidum	1
	Macroglossum fritzei	1
	Macroglossum passalus	1
	Macroglossum faro	1
	Macroglossum mitchelli	1
	Macroglossum multifascia	1
	Macroglossum prometheus	1
	Macroglossum mediiovitta	1
	Macroglossum corythus	1
	Macroglossum sylvia	1
	Macroglossum semifasciata	1
	Macroglossum hemichroma	1
	Macroglossum aquila	1
	Macroglossum variegatum	1
	Macroglossum heliophila	1
	Macroglossum insipida	1
	Macroglossum pseudungues	1
	Macroglossum gyrans	1
	Hippotion	Hippotion velox
Hippotion celerio		1
Hippotion echeclus		1
Hippotion boerhaviae		1
Hippotion rosetta		1
Theretra	Theretra nessus	1
	Theretra boisduvali	1
	Theretra rhesus	1
	Theretra clotho	1
	Theretra latreillei	1
	Theretra alecto	1
	Theretra suffusa	1
	Theretra oldenlandiae	1
	Theretra silhetensis	1
Rhyncholaba	Rhyncholaba acteus	1
Rhagastis	Rhagastis castor	1
	Rhagastis albomarginatus	1
	Rhagastis rubetra	1
Cechenena	Cechenena lineosa	1
	Cechenena helops	1
	Cechenena aegrota	1
TOTAL		273

APPENDIX 2

Table shows the test dataset: Distorted training images tested against training dataset: Superfamily Bombycoidea.

Test Image Family	No.	Test Images	Results
Family Lasiocampidae	1	<i>Alompra ferruginea</i>	True positive
	2	<i>Alompra roepkei</i>	True positive
	3	<i>Arguda albiscutulata</i>	True positive
	4	<i>Arguda insulindiana</i>	True positive
	5	<i>Arguda rectilinea</i> 1	True positive
	6	<i>Arguda rectilinea</i> 2	True positive
	7	<i>Arguda rosemariae</i>	True positive
	8	<i>Bhima borneana</i>	True positive
	9	<i>Chonopla tecta</i>	True positive
	10	<i>Euthrix</i> sp.	True positive
	11	<i>Euthrix laeta</i>	True positive
	12	<i>Gastropacha leopoldi</i> 1	True positive
	13	<i>Gastropacha leopoldi</i> 2	True positive
	14	<i>Hallicarnia albipectus</i> 1	True positive
	15	<i>Hallicarnia albipectus</i> 2	True positive
	16	<i>Kunugia austroplacida</i> 1	True positive
	17	<i>Kunugia austroplacida</i> 2	True positive
	18	<i>Kunugia basimacula</i>	True positive
	19	<i>Kunugia basinigra</i>	True positive
	20	<i>Kunugia drakei</i> 1	True positive
	21	<i>Kunugia drakei</i> 2	True positive
	22	<i>Kunugia ferox</i> 1	True positive
	23	<i>Kunugia ferox</i> 2	True positive
	24	<i>Kunugia gynandra</i> 1	True positive
	25	<i>Kunugia gynandra</i> 2	True positive
	26	<i>Kunugia leucopicta</i> 1	True positive
	27	<i>Kunugia leucopicta</i> 2	True positive
	28	<i>Kunugia quadrilineata</i>	True positive
	29	<i>Kunugia rectifascia</i>	True positive
	30	<i>Kunugia suanoides</i> 1	True positive
	31	<i>Kunugia suanoides</i> 2	True positive
	32	<i>Lajonquierea derunoides</i>	True positive
	33	<i>Lajonquierea jermyi</i>	True positive
	34	<i>Lajonquierea mediofasciata</i>	True positive
	35	<i>Lajonquierea mediofasciata</i>	False positive
	36	<i>Lajonquierea piccolopecta</i>	True positive
	37	<i>Lajonquierea variabile</i> 1	True positive
	38	<i>Lajonquierea variabile</i> 2	True positive

Continue Appendix 2

39	Lajonquierea variabile 3	True positive
40	Lebeda brauni	True positive
41	Lebeda cognata 1	True positive
42	Lebeda cognata 2	True positive
43	Lebeda intermedia	True positive
44	Lebeda pruetti 1	True positive
45	Lebeda pruetti 2	True positive
46	Metanastria gemella 1	True positive
47	Metanastria gemella 2	Metanastria gemella
48	Micropacha roepkei	True positive
49	Odonestis angulata	True positive
50	Odonestis erectilinea	True positive
51	Odonestis lipara	True positive
52	Odonestis vita	True positive
53	Odontocraspis hasora	True positive
54	Paradoxopla cardinalis	True positive
55	Paralebeda lucifuga 1	True positive
56	Paralebeda lucifuga 2	True positive
57	Paralebeda uniformis 1	True positive
58	Paralebeda uniformis 2	True positive
59	Radhica elisabethae 1	True positive
60	Radhica elisabethae 2	True positive
61	Radhica himerta	True positive
62	Radhica holoxantha	True positive
63	Streblote helpsi 1	True positive
64	Streblote helpsi 2	True positive
65	Suana concolor 1	True positive
66	Suana concolor 2	True positive
67	Suana sundana	True positive
68	Syastrena lanaoensis	True positive
69	Syastrena sumatrana	True positive
70	Syastrena tamsi	True positive
71	Takanea diehli 1	True positive
72	Takanea diehli 2	True positive
73	Trabala bouraq	False positive
74	Trabala ganesha 1	True positive
75	Trabala ganesha 2	False positive
76	Trabala garuda 1	True positive
77	Trabala garuda 2	True positive
78	Trabala gautama	True positive
79	Trabala hantu 1	True positive
80	Trabala hantu 2	False positive
81	Trabala irrorata 1	True positive
82	Trabala irrorata 2	True positive

Continue Appendix 2

	83	Trabala krishna 1	True positive
	84	Trabala krishna 2	True positive
	85	Trabala pallida 1	True positive
	86	Trabala pallida 2	True positive
	87	Trabala pallida 3	True positive
	88	Trabala rotundapex 1	True positive
	89	Trabala rotundapex 2	True positive
	90	Trabala shiva	False positive
	91	Trabala viridana 1	True positive
	92	Trabala viridana 2	True positive
Family Eupterotidae	93	Eupterote asclepiades	False positive
	94	Eupterote harmani 1	True positive
	95	Eupterote harmani 2	True positive
	96	Eupterote multiarcuata	True positive
	97	Eupterote muluana 1	True positive
	98	Eupterote muluana 2	False no-ID
	99	Eupterote naessigi	True positive
	100	Eupterote niassana	True positive
	101	Eupterote obsoleta	True positive
	102	Ganisa plana 1	True positive
	103	Ganisa plana 2	True positive
	104	Ganisa similis	True positive
	105	Melanothrix alternans 1	True positive
	106	Melanothrix alternans 2	True positive
	107	Melanothrix fumosa 1	True positive
	108	Melanothrix fumosa 2	True positive
	109	Melanothrix latevittata 1	True positive
	110	Melanothrix latevittata 2	True positive
	111	Melanothrix nymphaliaria 1	True positive
	112	Melanothrix nymphaliaria 2	True positive
	113	Pseudojana obscura	True positive
	114	Pseudojana prespicuifascia	True positive
Family Bombycidae	115	Andraca apodecta 1	True positive
	116	Andraca apodecta 2	False no-ID
	117	Bombyx incomposita	Bombyx incomposita
	118	Ernolatida lida	True positive
	119	Ernolatida moorei	True positive
	120	Gunda javanica 1	True positive
	121	Gunda javanica 2	True positive
	122	Gunda ochracea 1	True positive
	123	Gunda ochracea 2	True positive
	124	Gunda subnotata 1	True positive

Continue Appendix 2

	125	<i>Gunda subnotata</i> 2	True positive
	126	<i>Mustilia dierli</i>	True positive
	127	<i>Ocinara albiceps</i>	True positive
	128	<i>Ocinara albicollis</i>	True positive
	129	<i>Ocinara bifurcula</i> 1	True positive
	130	<i>Ocinara bifurcula</i> 2	True positive
	131	<i>Penicillifera apicalis</i>	True positive
	132	<i>Penicillifera purpurascens</i>	False positive
	133	<i>Prismosticta tiretta</i>	True positive
	134	<i>Trilocha friedeli</i>	True positive
Family Brahmaeidae	135	<i>Brahmaea hearseyi</i>	True positive
Family Saturniidae	136	<i>Actias maenas</i> 1	True positive
	137	<i>Actias maenas</i> 2	True positive
	138	<i>Actias selene</i>	True positive
	139	<i>Antheraea alleni</i> 1	True positive
	140	<i>Antheraea alleni</i> 2	True positive
	141	<i>Antheraea alleni</i> 3	True positive
	142	<i>Antheraea assamensis</i> 1	True positive
	143	<i>Antheraea assamensis</i> 2	True positive
	144	<i>Antheraea brunei</i> 1	True positive
	145	<i>Antheraea brunei</i> 2	True positive
	146	<i>Antheraea celebensis</i> 1	True positive
	147	<i>Antheraea celebensis</i> 2	True positive
	148	<i>Antheraea celebensis</i> 3	True positive
	149	<i>Antheraea diehli</i> 1	True positive
	150	<i>Antheraea diehli</i> 2	True positive
	151	<i>Antheraea helferi</i> 1	True positive
	152	<i>Antheraea helferi</i> 2	True positive
	153	<i>Antheraea jana</i> 1	True positive
	154	<i>Antheraea jana</i> 2	True positive
	155	<i>Antheraea korintjiana</i>	True positive
	156	<i>Antheraea larissa</i>	True positive
	157	<i>Antheraea moultoni</i> 1	True positive
	158	<i>Antheraea moultoni</i> 2	True positive
	159	<i>Antheraea moultoni</i> 3	True positive
	160	<i>Antheraea moultoni</i> 4	True positive
	161	<i>Antheraea rosieri</i> 1	True positive
	162	<i>Antheraea rosieri</i> 2	True positive
	163	<i>Antheraea rosieri</i> 3	True positive
	164	<i>Archaeoattacus staudingeri</i>	True positive
	165	<i>Attacus atlas</i>	True positive
	166	<i>Cricula bornea</i>	True positive
	167	<i>Cricula elaezia</i>	True positive

Continue Appendix 2

168	<i>Cricula trifenestrata</i> 1	True positive
169	<i>Cricula trifenestrata</i> 2	True positive
170	<i>Lemaireia loepoides</i>	True positive
171	<i>Loepa megacore</i>	True positive
172	<i>Loepa sikkima</i>	True positive
173	<i>Samia tetrica</i>	True positive
174	<i>Acherontia lachesis</i>	True positive
175	<i>Acherontia styx</i>	True positive
176	<i>Acosmeryx anceus</i>	True positive
177	<i>Acosmeryx shervillii</i> 1	True positive
178	<i>Acosmeryx shervillii</i> 2	True positive
179	<i>Agrius convolvuli</i>	True positive
180	<i>Ambulyx canescens</i>	True positive
181	<i>Ambulyx clavata</i>	True positive
182	<i>Ambulyx joiceyi</i>	True positive
183	<i>Ambulyx oblitterata</i>	True positive
184	<i>Ambulyx pryeri</i>	True positive
185	<i>Ambulyx subocellata</i>	True positive
186	<i>Ambulyx substrigilis</i>	True positive
187	<i>Ambulyx tattina</i>	True positive
188	<i>Amplipterus canopus</i>	True positive
189	<i>Angonyx testacea</i>	True positive
190	<i>Callambulyx rubricosa</i>	True positive
191	<i>Cechenena aegrota</i>	True positive
192	<i>Cechenena helops</i>	True positive
193	<i>Cechenena lineosa</i>	True positive
194	<i>Cephonodes picus</i>	True positive
195	<i>Clanis bilineata</i>	True positive
196	<i>Clanis stenosema</i>	True positive
197	<i>Cypa decolor</i> 1	True positive
198	<i>Cypa decolor</i> 2	True positive
199	<i>Daphnis hypothous</i>	True positive
200	<i>Daphnis placida</i>	True positive
201	<i>Daphnusa ocellaris</i>	False positive
202	<i>Degmaptera olivacea</i>	True positive
203	<i>Dolbina krikkeni</i>	True positive
204	<i>Elibia dolichus</i>	True positive
205	<i>Enpinanga borneensis</i> 1	True positive
206	<i>Enpinanga borneensis</i> 2	True positive
207	<i>Enpinanga vigens</i> 1	True positive
208	<i>Enpinanga vigens</i> 2	True positive
209	<i>Eurypteryx bhaga</i>	True positive
210	<i>Eurypteryx falcata</i>	True positive
211	<i>Eurypteryx shelfordi</i>	True positive

Continue Appendix 2

212	<i>Gehlenia falcata</i>	True positive
213	<i>Giganteopalpus mirabilis</i>	False positive
214	<i>Gnathothlibus erotus</i>	True positive
215	<i>Hippotion velox</i>	True positive
216	<i>Hippotion boerhaviae</i>	True positive
217	<i>Hippotion celerio</i>	True positive
218	<i>Hippotion echeclus</i>	True positive
219	<i>Hippotion rosetta</i>	True positive
220	<i>Macroglossum aquila</i>	True positive
221	<i>Macroglossum corythus</i>	True positive
222	<i>Macroglossum faro</i>	False positive
223	<i>Macroglossum fritzei</i>	True positive
224	<i>Macroglossum gyrans</i>	True positive
225	<i>Macroglossum heliophila</i>	True positive
226	<i>Macroglossum hemichroma</i>	True positive
227	<i>Macroglossum insipida</i>	True positive
228	<i>Macroglossum lepidum</i>	True positive
229	<i>Macroglossum mediovitta</i>	True positive
230	<i>Macroglossum mitchelli</i>	True positive
231	<i>Macroglossum multifascia</i>	True positive
232	<i>Macroglossum passalus</i>	True positive
233	<i>Macroglossum prometheus</i>	True positive
234	<i>Macroglossum pseudungues</i>	False positive
235	<i>Macroglossum semifasciata</i>	True positive
236	<i>Macroglossum sylvia</i>	False no-ID
237	<i>Macroglossum variegatum</i>	True positive
238	<i>Marumba cristata</i>	True positive
239	<i>Marumba dyras</i>	True positive
240	<i>Marumba juvenicus</i>	False positive
241	<i>Marumba spectabilis</i>	True positive
242	<i>Marumba sperchius</i>	True positive
243	<i>Marumba tigrina</i>	True positive
244	<i>Megacorma obliqua</i>	True positive
245	<i>Meganoton analis</i>	True positive
246	<i>Meganoton nyctiphanes</i>	True positive
247	<i>Meganoton rufescens</i>	True positive
248	<i>Panacra automedon</i>	True positive
249	<i>Panacra busiris</i>	True positive
250	<i>Panacra dohertyi</i>	True positive
251	<i>Panacra malayana</i>	True positive
252	<i>Panacra mydon</i>	True positive
253	<i>Panacra psaltria</i>	True positive
254	<i>Panacra variolosa</i>	True positive
255	<i>Psilogramma menephron 1</i>	True positive

Continue Appendix 2

	256	<i>Psilogramma menephron 2</i>	True positive
	257	<i>Rhagastis albomarginatus</i>	True positive
	258	<i>Rhagastis castor</i>	True positive
	259	<i>Rhagastis rubetra</i>	True positive
	260	<i>Rhyncholaba acteus</i>	True positive
	261	<i>Sataspes infernalis</i>	True positive
	262	<i>Smerinthulus diehli</i>	True positive
	263	<i>Smerinthulus quadripunctatus</i>	True positive
	264	<i>Smerinthulus terranea</i>	True positive
	265	<i>Theretra alecto</i>	True positive
	266	<i>Theretra boisduvali</i>	True positive
	267	<i>Theretra clotho</i>	True positive
	268	<i>Theretra latreillei</i>	True positive
	269	<i>Theretra nessus</i>	True positive
	270	<i>Theretra oldenlandiae</i>	True positive
	271	<i>Theretra rhesus</i>	False positive
	272	<i>Theretra silhetensis</i>	True positive
	273	<i>Theretra suffusa</i>	True positive

APPENDIX 3

Table shows detailed results of test dataset: Web images of Superfamily Bombycoidea tested against training dataset: Superfamily Bombycoidea.

Source	No.	Test Images	Results
Barcode of Life Database	1	Bold001	False no-ID
	2	Bold002	False positive
	3	Bold003	False positive
	4	Bold004	True positive
	5	Bold005	True positive
	6	Bold006	True positive
	7	Bold007	True positive
	8	Bold008	True positive
	9	Bold009	False no-ID
	10	Bold010	False positive
	11	Bold011	False no-ID
	12	Bold012	False positive
	13	Bold013	False positive
	14	Bold014	False positive
	15	Bold015	False positive
	16	Bold016	False no-ID
	17	Bold017	True positive
	18	Bold018	True positive
	19	Bold019	False no-ID
	20	Bold020	False no-ID
	21	Bold021	False positive
	22	Bold022	False no-ID
	23	Bold023	True positive
	24	Bold024	False positive
	25	Bold025	True positive
	26	Bold026	True positive
	27	Bold027	True positive
	28	Bold028	False positive
	29	Bold029	False positive
	30	Bold030	False positive
	31	Bold031	False no-ID
	32	Bold032	True positive
	33	Bold033	False no-ID
	34	Bold034	False no-ID
	35	Bold035	True positive
	36	Bold036	True positive
	37	Bold037	False positive
	38	Bold038	False positive
	39	Bold039	False positive

Continue Appendix 3

40	Bold040	True positive
41	Bold041	True positive
42	Bold042	True positive
43	Bold043	True positive
44	Bold044	True positive
45	Bold045	False positive
46	Bold046	False positive
47	Bold047	False positive
48	Bold048	False positive
49	Bold049	False positive
50	Bold050	False positive
51	Bold051	False positive
52	Bold052	False positive
53	Bold053	False positive
54	Bold054	False no-ID
55	Bold055	False positive
56	Bold056	False no-ID
57	Bold057	False positive
58	Bold058	True positive
59	Bold059	False no-ID
60	Bold060	True positive
61	Bold061	True positive
62	Bold062	True positive
63	Bold063	False positive
64	Bold064	False positive
65	Bold065	True positive
66	Bold066	True positive
67	Bold067	False positive
68	Bold068	False positive
69	Bold069	False positive
70	Bold070	True positive
71	Bold071	False positive
72	Bold072	False positive
73	Bold073	False positive
74	Bold074	False positive
75	Bold075	False positive
76	Bold076	False positive
77	Bold077	False positive
78	Bold078	True positive
79	Bold079	False positive
80	Bold080	True positive
81	Bold081	False positive
82	Bold082	False positive
83	Bold083	False positive

Continue Appendix 3

84	Bold084	False positive
85	Bold085	False no-ID
86	Bold086	False no-ID
87	Bold087	False no-ID
88	Bold088	True positive
89	Bold089	False positive
90	Bold090	True positive
91	Bold091	False no-ID
92	Bold092	False positive
93	Bold093	True positive
94	Bold094	True positive
95	Bold095	False positive
96	Bold096	False no-ID
97	Bold097	False no-ID
98	Bold098	False no-ID
99	Bold099	True positive
100	Bold100	False positive
101	Bold101	False no-ID
102	Bold102	True positive
103	Bold103	False positive
104	Bold104	True positive
105	Bold105	False no-ID
106	Bold106	False positive
107	Bold107	False no-ID
108	Bold108	False positive
109	Bold109	False no-ID
110	Bold110	True positive
111	Bold111	False no-ID
112	Bold112	False positive
113	Bold113	False no-ID
114	Bold114	False no-ID
115	Bold115	True positive
116	Bold116	True positive
117	Bold117	True positive
118	Bold118	True positive
119	Bold119	True positive
120	Bold120	False positive
121	Bold121	False positive
122	Bold122	True positive
123	Bold123	False positive
124	Bold124	False positive
125	Bold125	False positive
126	Bold126	True positive
127	Bold127	False positive

Continue Appendix 3

128	Bold128	True positive
129	Bold129	False no-ID
130	Bold130	False positive
131	Bold131	False no-ID
132	Bold132	False positive
133	Bold133	True positive
134	Bold134	False positive
135	Bold135	False positive
136	Bold136	True positive
137	Bold137	True positive
138	Bold138	True positive
139	Bold139	True positive
140	Bold140	True positive
141	Bold141	True positive
142	Bold142	True positive
143	Bold143	False positive
144	Bold144	False positive
145	Bold145	True positive
146	Bold146	True positive
147	Bold147	True positive
148	Bold148	True positive
149	Bold149	True positive
150	Bold150	False positive
151	Bold151	False positive
152	Bold152	False positive
153	Bold153	True positive
154	Bold154	True positive
155	Bold155	True positive
156	Bold156	True positive
157	Bold157	False positive
158	Bold158	True positive
159	Bold159	True positive
160	Bold160	False positive
161	Bold161	True positive
162	Bold162	True positive
163	Bold163	True positive
164	Bold164	False positive
165	Bold165	False positive
166	Bold166	False positive
167	Bold167	False positive
168	Bold168	False positive
169	Bold169	True positive
170	Bold170	False positive
171	Bold171	False positive

Continue Appendix 3

172	Bold172	True positive
173	Bold173	True positive
174	Bold174	True positive
175	Bold175	False positive
176	Bold176	True positive
177	Bold177	False positive
178	Bold178	True positive
179	Bold179	True positive
180	Bold180	True positive
181	Bold181	True positive
182	Bold182	False positive
183	Bold183	True positive
184	Bold184	True positive
185	Bold185	True positive
186	Bold186	True positive
187	Bold187	False positive
188	Bold188	False positive
189	Bold189	True positive
190	Bold190	True positive
191	Bold191	True positive
192	Bold192	False positive
193	Bold193	True positive
194	Bold194	True positive
195	Bold195	True positive
196	Bold196	True positive
197	Bold197	True positive
198	Bold198	True positive
199	Bold199	True positive
200	Bold200	False positive
201	Bold201	False positive
202	Bold202	True positive
203	Bold203	True positive
204	Bold204	True positive
205	Bold205	False positive
206	Bold206	False positive
207	Bold207	False positive
208	Bold208	False positive
209	Bold209	False positive
210	Bold210	True positive
211	Bold211	False no-ID
212	Bold212	True positive
213	Bold213	True positive
214	Bold214	True positive
215	Bold215	False positive

Continue Appendix 3

216	Bold216	False positive
217	Bold217	False positive
218	Bold218	False positive
219	Bold219	False positive
220	Bold220	False positive
221	Bold221	False positive
222	Bold222	False positive
223	Bold223	True positive
224	Bold224	False no-ID
225	Bold225	False positive
226	Bold226	False positive
227	Bold227	False positive
228	Bold228	False positive
229	Bold229	False positive
230	Bold230	False positive
231	Bold231	False positive
232	Bold232	False positive
233	Bold233	False positive
234	Bold234	False positive
235	Bold235	False positive
236	Bold236	False positive
237	Bold237	False no-ID
238	Bold238	False positive
239	Bold239	False positive
240	Bold240	False positive
241	Bold241	False positive
242	Bold242	False positive
243	Bold243	False positive
244	Bold244	True positive
245	Bold245	False positive
246	Bold246	False positive
247	Bold247	False positive
248	Bold248	False positive
249	Bold249	False positive
250	Bold250	True positive
251	Bold251	False positive
252	Bold252	False positive
253	Bold253	True positive
254	Bold254	True positive
255	Bold255	False positive
256	Bold256	True positive
257	Bold257	True positive
258	Bold258	True positive
259	Bold259	True positive

Continue Appendix 3

	260	Bold260	True positive
	261	Bold261	True positive
	262	Bold262	False positive
	263	Bold263	False positive
	264	Bold264	True positive
	265	Bold265	True positive
	266	Bold266	False positive
	267	Bold267	True positive
	268	Bold268	True positive
	269	Bold269	False positive
	270	Bold270	True positive
	271	Bold271	False positive
	272	Bold272	False positive
	273	Bold273	False positive
	274	Bold274	True positive
	275	Bold275	True positive
	276	Bold276	False positive
	277	Bold277	False positive
	278	Bold278	True positive
	279	Bold279	True positive
	280	Bold280	True positive
Digital Moths of Japan	281	DMJ001	False positive
	282	DMJ002	True positive
	283	DMJ003	False positive
	284	DMJ004	True positive
	285	DMJ005	True positive
	286	DMJ006	False no-ID
	287	DMJ007	False no-ID
	288	DMJ008	True positive
	289	DMJ009	False positive
	290	DMJ010	False positive
	291	DMJ011	True positive
	292	DMJ012	False positive
	293	DMJ013	False positive
	294	DMJ014	False positive
	295	DMJ015	False positive
	296	DMJ016	False positive
	297	DMJ017	False positive
	298	DMJ018	False positive
	299	DMJ019	False positive
	300	DMJ020	False no-ID
	301	DMJ021	False positive
	302	DMJ022	False positive
	303	DMJ023	False positive

Continue Appendix 3

	304	DMJ024	True positive
	305	DMJ025	True positive
	306	DMJ026	True positive
	307	DMJ027	False positive
	308	DMJ028	False positive
	309	DMJ029	True positive
	310	DMJ030	False no-ID
	311	DMJ031	False no-ID
	312	DMJ032	True positive
	313	DMJ033	True positive

APPENDIX 4

Table shows results of test dataset: Other parts of Borneo against training dataset: Superfamily Bombycoidea.

Source	No.	Test Image	Result
Part 4 (Family Notodontidae)	1	Dudusa vethi	True no-ID
	2	Phycidopsis albovittata	False positive
	3	Porsica punctifascia	False positive
	4	Tarsolepis sommeri	False positive
Part 5 (Family Lymantriidae)	5	Lymantria brunneiplaga	False positive
	6	Lymantria ganara	False positive
	7	Lymantria lepcha	False positive
	8	Lymantria marginalis	False positive
	9	Lymantria singapura	False positive
Part 7 (Family Arctiidae)	10	Cyana maiae	False positive
	11	Cyana conclusa	False positive
	12	Cyana effracta	False positive
	13	Oeonistis altica	False positive
	14	Vamuna remelana	False positive
Part 8 (Family Drepaninae & Callidulidae)	15	Horityhyatira decorate	True no-ID
	16	Tethea oberthueri	True no-ID
	17	Tetragonus catamitus	False positive
Part 9 (Family Geometridae)	18	Sarcinoides perakaria	False positive
	19	Sarcinoides vultuaria	False positive
Part 12 (Family Notuidae)	20	Agrotis epsilon	False positive
	21	Agrotis kinabaluensis	False positive
	22	Diarsia nigrosira	True no-ID
	23	Dictyestra dissectus	True no-ID
	24	Tiracola plagiata	False positive
	25	Helicoverpa assulta	True no-ID
Part 15 & 16 (Family Noctuidae)	26	Achaea janata	False positive
	27	Macaldenia palumba	False positive
	28	Ophisma pallescens	False positive
	29	Ophiura trapezium	False positive
	30	Thyas honesta	True no-ID

APPENDIX 5

Table shows detailed results of test dataset: Belize sphingids tested against training dataset: Superfamily Bombycoidea.

No.	Test Images	Returned As
1	Adhemarius gannascus 1	False positive
2	Adhemarius gannascus 2	False positive
3	Adhemarius gannascus 3	False positive
4	Adhemarius gannascus 4	False positive
5	Adhemarius gannascus 5	False positive
6	Adhemarius ypsilon 1	False positive
7	Adhemarius ypsilon 2	True no-ID
8	Adhemarius ypsilon 3	False positive
9	Adhemarius ypsilon 4	False positive
10	Adhemarius ypsilon 5	True no-ID
11	Callionima falcifera 1	False positive
12	Callionima falcifera 2	True no-ID
13	Callionima falcifera 3	True no-ID
14	Callionima falcifera 4	False positive
15	Callionima falcifera 5	True no-ID
16	Callionima_inuus 1	False positive
17	Callionima_inuus 2	False positive
18	Callionima_inuus 3	True no-ID
19	Callionima_inuus 4	False positive
20	Callionima_inuus 5	False positive
21	Callionima_nomius 1	False positive
22	Callionima_nomius 2	True no-ID
23	Callionima_nomius 3	False positive
24	Callionima_nomius 4	False positive
25	Callionima_nomius 5	True no-ID
26	Callionima_parce 1	True no-ID
27	Callionima_parce 2	False positive
28	Callionima_parce 3	False positive
29	Callionima_parce 4	False positive
30	Callionima_parce 5	False positive
31	Cautethia_spuria 1	False positive
32	Cautethia_spuria 2	False positive
33	Cautethia_spuria 3	False positive
34	Cautethia_spuria 4	False positive
35	Cautethia_spuria 5	False positive
36	Cautethia_yucatanana 1	False positive
37	Cautethia_yucatanana 2	False positive
38	Cautethia_yucatanana 3	False positive
39	Cautethia_yucatanana 4	False positive

Continue Appendix 5

40	Cautethia_yucatana 5	False positive
41	Cocytius_duponchel 1	True no-ID
42	Cocytius_duponchel 2	False positive
43	Cocytius_duponchel 3	False positive
44	Cocytius_duponchel 4	False positive
45	Cocytius_duponchel 5	False positive
46	Cocytius_lucifer 1	False positive
47	Cocytius_lucifer 2	False positive
48	Enyo_gorgon	False positive
49	Enyo_lugubris 1	False positive
50	Enyo_lugubris 2	True no-ID
51	Enyo_lugubris 3	False positive
52	Enyo_lugubris 4	False positive
53	Enyo_lugubris 5	False positive
54	Enyo_ocypete 1	False positive
55	Enyo_ocypete 2	False positive
56	Enyo_ocypete 3	False positive
57	Enyo_ocypete 4	True no-ID
58	Enyo_ocypete 5	False positive
59	Erinnyis_crameri 1	False positive
60	Erinnyis_crameri 2	False positive
61	Erinnyis_crameri 3	False positive
62	Erinnyis_crameri 4	False positive
63	Erinnyis_crameri 5	False positive
64	Erinnyis_ello 1	False positive
65	Erinnyis_ello 2	False positive
66	Erinnyis_ello 3	False positive
67	Erinnyis_ello 4	False positive
68	Erinnyis_ello 5	False positive
69	Erinnyis_oenotrus 1	False positive
70	Erinnyis_oenotrus 2	True no-ID
71	Erinnyis_oenotrus 3	False positive
72	Erinnyis_oenotrus 4	True no-ID
73	Erinnyis_oenotrus 5	False positive
74	Erinnyis_yucatana 1	False positive
75	Erinnyis_yucatana 2	False positive
76	Erinnyis_yucatana 3	False positive
77	Eumorpha_anchemolus	False positive
78	Eumorpha_obliquus	False positive
79	Eumorpha_satellititia 1	False positive
80	Eumorpha_satellititia 2	False positive
81	Eumorpha_satellititia 3	False positive
82	Eumorpha_satellititia 4	False positive
83	Eumorpha_satellititia 5	False positive

Continue Appendix 5

84	<i>Eumorpha triangulum</i> 1	True no-ID
85	<i>Eumorpha triangulum</i> 2	False positive
86	<i>Eumorpha triangulum</i> 3	False positive
87	<i>Eumorpha triangulum</i> 4	False positive
88	<i>Eumorpha triangulum</i> 5	False positive
89	<i>Hemeroplanes triptolemus</i>	False positive
90	<i>Hyles lineata</i>	False positive
91	<i>Isognathus rimosa</i> 1	False positive
92	<i>Isognathus rimosa</i> 2	False positive
93	<i>Isognathus rimosa</i> 3	False positive
94	<i>Isognathus rimosa</i> 4	False positive
95	<i>Isognathus rimosa</i> 5	False positive
96	<i>Madoryx bubastus</i>	False positive
97	<i>Madoryx oiclus</i> 1	False positive
98	<i>Madoryx oiclus</i> 2	False positive
99	<i>Madoryx oiclus</i> 3	False positive
100	<i>Madoryx oiclus</i> 4	False positive
101	<i>Madoryx oiclus</i> 5	False positive
102	<i>Madoryx plutonius</i> 1	False positive
103	<i>Madoryx plutonius</i> 2	False positive
104	<i>Madoryx plutonius</i> 3	True no-ID
105	<i>Manduca albiplaga</i> 1	False positive
106	<i>Manduca albiplaga</i> 2	False positive
107	<i>Manduca albiplaga</i> 3	False positive
108	<i>Manduca corallina</i> 1	False positive
109	<i>Manduca corallina</i> 2	False positive
110	<i>Manduca corallina</i> 3	False positive
111	<i>Manduca corallina</i> 4	False positive
112	<i>Manduca corallina</i> 5	False positive
113	<i>Manduca florestan</i> 1	False positive
114	<i>Manduca florestan</i> 2	False positive
115	<i>Manduca florestan</i> 3	False positive
116	<i>Manduca florestan</i> 4	False positive
117	<i>Manduca florestan</i> 5	False positive
118	<i>Manduca hannibal</i>	False positive
119	<i>Manduca lanuginosa</i>	False positive
120	<i>Manduca rustica</i>	False positive
121	<i>Manduca sexta</i> 1	False positive
122	<i>Manduca sexta</i> 2	False positive
123	<i>Neococytius cluentius</i> 1	False positive
124	<i>Neococytius cluentius</i> 2	False positive
125	<i>Neococytius cluentius</i> 3	True no-ID
126	<i>Neococytius cluentius</i> 4	True no-ID
127	<i>Neococytius cluentius</i> 5	False positive

Continue Appendix 5

128	Nyceryx_riscus	False positive
129	Pachylia_ficus 1	False positive
130	Pachylia_ficus 2	False positive
131	Pachylia_ficus 3	False positive
132	Pachylia_ficus 4	True no-ID
133	Pachylia_ficus 5	False positive
134	Pachylia_syces	False positive
135	Pachylioides_resumens 1	False positive
136	Pachylioides_resumens 2	False positive
137	Pachylioides_resumens 3	False positive
138	Pachylioides_resumens 4	False positive
139	Pachylioides_resumens 5	False positive
140	Perigonia_ilus 1	False positive
141	Perigonia_ilus 2	False positive
142	Perigonia_ilus 3	True no-ID
143	Perigonia_ilus 4	False positive
144	Perigonia_ilus 5	False positive
145	Protambulyx_eurycles	False positive
146	Protambulyx_strigilis 1	False positive
147	Protambulyx_strigilis 2	False positive
148	Protambulyx_strigilis 3	False positive
149	Protambulyx_strigilis 4	False positive
150	Protambulyx_strigilis 5	False positive
151	Pseudosphinx_tetrio 1	False positive
152	Pseudosphinx_tetrio 2	False positive
153	Pseudosphinx_tetrio 3	False positive
154	Pseudosphinx_tetrio 4	False positive
155	Pseudosphinx_tetrio 5	False positive
156	Xylophanes_anubus 1	False positive
157	Xylophanes_anubus 2	False positive
158	Xylophanes_anubus 3	True no-ID
159	Xylophanes_anubus 4	False positive
160	Xylophanes_anubus 5	False positive
161	Xylophanes_belti 1	False positive
162	Xylophanes_belti 2	False positive
163	Xylophanes_belti 3	False positive
164	Xylophanes_belti 4	False positive
165	Xylophanes_belti 5	False positive
166	Xylophanes_ceratomioides 1	False positive
167	Xylophanes_ceratomioides 2	False positive
168	Xylophanes_ceratomioides 3	False positive
169	Xylophanes_ceratomioides 4	False positive
170	Xylophanes_ceratomioides 5	False positive
171	Xylophanes_chiron 1	False positive

Continue Appendix 5

172	Xylophanes_chiron 2	False positive
173	Xylophanes_chiron 3	False positive
174	Xylophanes_chiron 4	False positive
175	Xylophanes_chiron 5	False positive
176	Xylophanes_cyrene 1	False positive
177	Xylophanes_cyrene 2	False positive
178	Xylophanes_cyrene 3	False positive
179	Xylophanes_cyrene 4	False positive
180	Xylophanes_cyrene 5	False positive
181	Xylophanes_libya 1	False positive
182	Xylophanes_libya 2	False positive
183	Xylophanes_libya 3	False positive
184	Xylophanes_libya 4	False positive
185	Xylophanes_libya 5	False positive
186	Xylophanes_neoptolemus 1	False positive
187	Xylophanes_neoptolemus 2	False positive
188	Xylophanes_neoptolemus 3	False positive
189	Xylophanes_neoptolemus 4	False positive
190	Xylophanes_neoptolemus 5	False positive
191	Xylophanes_pluto 1	False positive
192	Xylophanes_pluto 2	False positive
193	Xylophanes_pluto 3	False positive
194	Xylophanes_pluto 4	False positive
195	Xylophanes_pluto 5	False positive
196	Xylophanes_porcus	False positive
197	Xylophanes_thyelia 1	False positive
198	Xylophanes_thyelia 2	False positive
199	Xylophanes_thyelia 3	False positive
200	Xylophanes_thyelia 4	False positive
201	Xylophanes_thyelia 5	False positive
202	Xylophanes_titana 1	False positive
203	Xylophanes_titana 2	False positive
204	Xylophanes_tyndarus 1	False positive
205	Xylophanes_tyndarus 2	False positive
206	Xylophanes_tyndarus 3	False positive

APPENDIX 6

Table shows results of test dataset: Genting moth collections against training dataset: Superfamily Bombycoidea.

Specimen	Species from BOLD	Identification by DAISY
Live023	<i>Arcte modesta</i>	True no-ID
Live026	<i>Lyssa zampa</i>	False positive
Live034	<i>Ambulyx canescens</i>	True Positive
Live030	<i>Erebus caprimulgus</i>	False positive
Live001	<i>Asota caricae</i>	False positive
Live031	<i>Arcte modesta</i>	False positive
Live036	<i>Erebus caprimulgus</i>	False positive
Live028	<i>Daphnis hypothous</i>	True Positive
Live032	<i>Thyas honesta</i>	False positive
Live003	<i>Achaea serva</i>	False positive

APPENDIX 7

Random selection of training images:



Alompra ferruginea



Arguda rectilinea



Gastropacha leopoldi



Lajonquiereea derunoides



Archerontia lachesis



Daphnis hypothous



Gehlenia falcata



Satapes infernalis

APPENDIX 8

Random selection of test images:



Gunda ochracea



Mustilia dierli



Ocinara albicollis



Ernolatia moorei



Antheraea brunei



Trabala irrorata



Ganisa similis



Brahmaea hearseyi