Classification of *Suncus murinus* species complex (Soricidae: Crocidurinae) in Peninsular Malaysia using Image Analysis and Machine Learning Approaches

Arpah Abu\(^1\text{*}\), Lee Kien Leow\(^1\text{*}\), Rosli Ramli\(^1\text{*}\), Hasmahzaiti Omar\(^1\text{*}\)

\(^1\text{Institute of Biological Sciences, Faculty of Science, University of Malaya, 50603 Kuala Lumpur, Malaysia}\)

\(^*\text{These authors contributed equally to this work}\)

\(^\text{§}\text{Corresponding author}\)

Email addresses:

AA: arpah@um.edu.my

LLK: leowlkien@gmail.com

RR: rosliramli@um.edu.my

HO: zaiti_1978@um.edu.my
Abstract

Background
Taxonomists frequently identify specimen from various populations based on the morphological characteristics and molecular data. This study looks into another invasive process in identification of house shrew (*Suncus murinus*) using image analysis and machine learning approaches. Thus, an automated identification system is developed to assist and simplify this task. In this study, seven descriptors namely area, convex area, major axis length, minor axis length, perimeter, equivalent diameter and extent which are based on the shape are used as features to represent digital image of skull that consists of dorsal, lateral and jaw views for each specimen. An Artificial Neural Network (ANN) is used as classifier to classify the skulls of *S. murinus* based on region (northern and southern populations of Peninsular Malaysia) and sex (adult male and female). Thus, specimen classification using Training data set and identification using Testing data set were performed through two stages of ANNs.

Results
At present, the classifier used has achieved an accuracy of 100% based on skulls’ views. Classification and identification to regions and sexes have also attained 72.5%, 87.5% and 80.0% of accuracy for dorsal, lateral, and jaw views, respectively. This results show that the shape characteristic features used are substantial because they can differentiate the specimens based on regions and sexes up to the accuracy of 80% and above. Finally, an application was developed and can be used for the scientific community.

Conclusions
This automated system demonstrates the practicability of using computer-assisted systems in providing interesting alternative approach for quick and easy identification of unknown species.