

RESEARCH

Open Access

Discriminative histogram taxonomy features for snake species identification

Alex Pappachen James^{1,2*}, Bincy Mathews², Sherin Sugathan² and Dileep Kumar Raveendran³

* Correspondence: apj@ieee.org
¹School of Engineering, Nazarbayev University, Astana, Kazakhstan
²Enview R&D Labs, Thiruvananthapuram, India
Full list of author information is available at the end of the article

Abstract

Background: Incorrect snake identification from the observable visual traits is a major reason for death resulting from snake bites in tropics. So far no automatic classification method has been proposed to distinguish snakes by deciphering the taxonomy features of snake for the two major species of snakes i.e. Elapidae and Viperidae. We identify 38 different taxonomically relevant features to develop the Snake database from 490 sample images of *Naja Naja* (Spectacled cobra), 193 sample images of *Ophiophagus Hannah* (King cobra), 88 images of *Bungarus caeruleus* (Common krait), 304 sample images of *Daboia russelii* (Russell's viper), 116 images of *Echis carinatus* (Saw scaled viper) and 108 images of *Hypnale hypnale* (Hump Nosed Pit Viper).

Results: Snake identification performances with 13 different types of classifiers and 12 attribute elevator demonstrate that 15 out of 38 taxonomically relevant features are enough for snake identification. Interestingly, these features were almost equally distributed from the logical grouping of top, side and body views of snake images, and the features from the bottom view of snakes had the least role in the snake identification.

Conclusion: We find that only few of the taxonomically relevant snake features are useful in the process of snake identification. These discriminant features are essential to improve the accuracy of snake identification and classification. The presented study indicate that automated snake identification is useful for practical applications such as in medical diagnosis, conservation studies and surveys by interdisciplinary practitioners with little expertise in snake taxonomy.

Keywords: Snake classification; Snake database; Taxonomy; Classifiers; Feature analysis

Background

Snake is a cold blooded reptile that is in majority perceived to be deadly to humans [1-5]. Since the ancient times, Snakes have been worshipped, feared and disliked by people across the world. Snake remain a painful reality in the daily life of millions of affected people and is largely one of the most misunderstood species [6,7]. At the same time, they are more perilous than the wild animals due to their close existence near human habitation [2]. World Health organization reports around five million snake bites every year resulting in millions of envenomation, hundreds of thousands of amputations and deaths. In cities like Thiruvananthapuram in Kerala, that has high humidity environment, where we started our study, on daily approximately 25–30 Snakes

sightings are reported. Majority of these sighted snakes were identified to equip with enough venom to kill a human in the course of few hours.

In tropical regions of the world, most of the snake bite cases are caused by four venomous snakes often referred to as “Big Four” snakes [8]. They include Spectacled Cobra (*Naja naja*), Common Krait (*Bungarus caeruleus*), Russell’s Viper (*Daboia russelii*) and Saw Scaled Viper (*Echis carinatus*) [7]. Another snakes which causes major snake bite cases and is very commonly found are King cobra (*Ophiophagus Hannah*) and Hump nosed Pit Viper (*Hypnale hypnale*). Due to this reason we restrict our study in this paper to these six deadly snakes [9,10].

Although anti-venom is produced in sufficient quantities by several public and private manufacturers, most snake bite victims don’t have access to good quality care, and in populated countries like India, both morbidity and mortality due to snake bite is high. Because of serious misreporting, the true burden of snake bite is not known. Doctors mostly inject polyvalent anti-venom to the snake bite victim. This is injected without considering which snake has bitten the person, even under the situation when the patient has knowledge about some observational features of the snake under consideration. The taxonomy of the snake is not well understood by majority of the medical practitioners making the correct identification of the snake from the remarks of the victims or eye witness. The polyvalent anti-venom injected by the medical practitioner contains antibodies raised against two or more species of snake, which may neutralize the venom injected by a single snake bite. Since there is only one type of venom injected by a snake bite, the remaining non-neutralized part of the polyvalent anti-venom used for treating the patient creates further risk to the human health. So proper identification of the snake is very important for the proper medical treatment to save the life of the snake bite victims [9-11].

To our knowledge, there has been no research reported yet on computer based approach to automatically distinguish snake classes. This may be largely due the lack of database for this purpose and less awareness of snake taxonomy research. The lack of database of venomous snakes in India makes this research very challenging, as the collection of images often involve well trained snake catchers, photographers and expert biologists. Through this paper we provide an early set of snake images that are collected in a view to identify relevant features based on snake taxonomy. In addition, the images contain a wide range of features from different snakes that can help with gaining newer understanding on snake taxonomy. The Indian snake taxonomy is a topic that is not investigated with rigor and there is lack of expert taxonomists. This makes the first line snake identification difficult in life threatening situations that are essential for recommending accurate treatment to the snake bite victims.

Materials and methods

Snake database

The snake images for the experiment were collected from forest across different parts of Kerala, India with the help of snake catchers from Pujappura Panchakarma Serpentarium, Trivandrum, India, through the close and 1 year long interaction with the subjects under study. The total number of images used for this experiment is 1299 that are obtained from 10–15 wild snakes of each species taken at different occasions and time.

Table 1 shows the taxonomically relevant features and their logical grouping based on the top, bottom, side or body view of the snake in the captured image, and Figure 1

shows the visual description of taxonomy features for each of snake class. The descriptions of the snakes are included as a supplementary file (Additional file 1). In total, 38 taxonomy based features are identified for creation of the feature database from 1299 snake images collected. There are a total of 490 images of spectacled cobra, 304 images of Russell's viper, 193 images of king cobra, 88 images of common krait, 116 images of saw scaled viper and 108 images of hump nosed pit viper. For creating the feature database, the 1299 snake images are manually converted by taxonomist to form feature vectors representing 38 taxonomically relevant features. This database file is included as a supplementary material to this article (Additional file 2).

Feature ranking and selection

Out of 38 taxonomically relevant features, top features that have highest impact on classification are determined. In order to find the top features from the complete database following 12 Attribute Elevators are used: ChiSquared AttributeEval [12], CfsSubsetEval [13], ConsistencySubsetEval [14], FilteredAttributeEval [15], FilteredSubsetEval [16], GainRatioAttributeEval [17], InfoGainAttributeEval [18], OneRAttributeEval [19], PrincipalComponents [20], ReliefAttributeEval [21], SVMAttributeEval [19], SymmetricalUncertAttributeEva [19], along with combination of certain search methods [21,22] like Genetic Search, Greedy Stepwise, Linear Forward Selection, Rank Search, Scatter Search, Subset Size Forward Selection and Ranker. The histogram of the feature counts from these attribute elevators is then plotted to get the ranking of the taxonomically relevant features that are most useful for the classification as shown in Figure 2. The concept of ranking and histograms used in this method is useful for identifying the relevance of the features [23-25]. The rank table is made with the help of this histogram based on the total number of repetitions of each features in the experiment. The repetitions of the feature results from the repeated ranking of features using different feature ranking method. The features that share same number of repetitions are then ranked on the basis of their average classification score taken independently for that feature i.e. features with highest average classification score among the features with same repetition is ranked first. Table 2 shows the ranking of all the 38 features using the attribute elevators with search method and classification score. The rank list of features is used to prepare 38 feature subsets with different numbers of features from 1 to 38 starting from the top feature to the last feature of Table 2. The numbers of features in the feature subsets are referred to as feature size.

Classifier selection and training

In order to perform automated snake classification following 13 classifiers are used: Bayes Net [26], Naïve Bayes [27], Multilayer perception [26], Ada BoostM1 [28], Multi BoostAB [29], RBF network [30], IB1 [30], IBk [31], LWL [32], NB Tree [33], J48 [34], Random Sub Space [35], and Bagging [36]. In the setting up the classification experiment, the database is split into training and test set. The training set is the one that will train the classifier parameter, while the test set is used to assess the performance of the classifier in terms of classification accuracy, F-score value, the area under the receiver operator characteristic curve, precision and recall rates. The selection of less number of samples per snake class in the training set makes the problem challenging and performance

Table 1 The table shows the grouping of the taxonomy features and its idealistic feature values for the creation of the database for automatic classification purpose

Feature group	Features	Feature name	Spectacled cobra	King cobra	Common krait	Russel's viper	Saw scaled viper	Hump nosed pit viper
Top	F_1	Rostral	1	1	1	1	1	1
	F_2	Internasal	2	2	2	1	1	2
	F_3	Prefrontal	2	2	2	1	1	2
	F_4	Supraocular	2	2	2	1	1	2
	F_5	Frontal	2	2	2	1	1	2
	F_6	Parietals	2	2	2	1	1	2
	F_7	V mark on head	0	0	0	1	0	0
	F_8	Triangular head	0	0	0	1	0	1
	F_9	Two dark patch on head	0	0	0	1	0	0
	F_{10}	Number of scales between Supraoculars	1	1	1	6-9	6-9	1
Side	F_{11}	Big occipital	0	1	0	0	0	0
	F_{12}	Plus sign in the head	0	0	0	0	1	0
	F_{13}	Small nostril	1	1	1	0	1	1
	F_{14}	Round pupil	1	1	1	0	0	0
	F_{15}	Big nostril	0	0	0	1	0	0
	F_{16}	Elliptical pupil	0	0	0	1	1	1
	F_{17}	Loreal	0	0	0	1	1	1
	F_{18}	Nasorostral	0	0	0	1	0	0
	F_{19}	Supranasal	0	0	0	1	0	0
	F_{20}	Triangular brown streaks below/behind eyes	0	0	0	1	0	0
	F_{21}	Subocular	0	0	0	1	1	1
	F_{22}	Nasal	2	2	2	1	3	1
	F_{23}	Preoculars	1	1	1	4	4	4
	F_{24}	Postoculars	3	3	2	4	4	4
	F_{25}	Supralabial scale	6-7	6-7	6-7	9-11	9-11	9-11
F_{38}	Pit between eyes and nose	0	0	0	0	0	1	
Bottom	F_{26}	Mental	1	1	1	1	1	1
	F_{27}	Asterior sublingual	1	1	1	1	1	1
	F_{28}	Posterior sublingual	1	1	1	1	1	1
Body	F_{29}	Round/smooth scale	1	1	1	0	0	0
	F_{30}	Hood	1	0	0	0	0	0
	F_{31}	Spectacled mark on hood	1	0	0	0	0	0
	F_{32}	Keeled scale	0	0	0	1	1	1
	F_{33}	Spots on dorsal scale	0	0	0	1	1	1
	F_{34}	White/yellow stripes on dorsal scale	0	1	1	0	0	0
	F_{35}	Black stripes on ventral scale	0	1	0	0	0	0

Table 1 The table shows the grouping of the taxonomy features and its idealistic feature values for the creation of the database for automatic classification purpose (Continued)

F_{36}	Enlarged and Hexagonal vertebral scale	0	0	1	0	0	0
F_{37}	Ventral scale	1	1	1	1	1	1

If certain features are visible in the image, corresponding values are assigned else for every invisible or missing feature '0' is assigned.

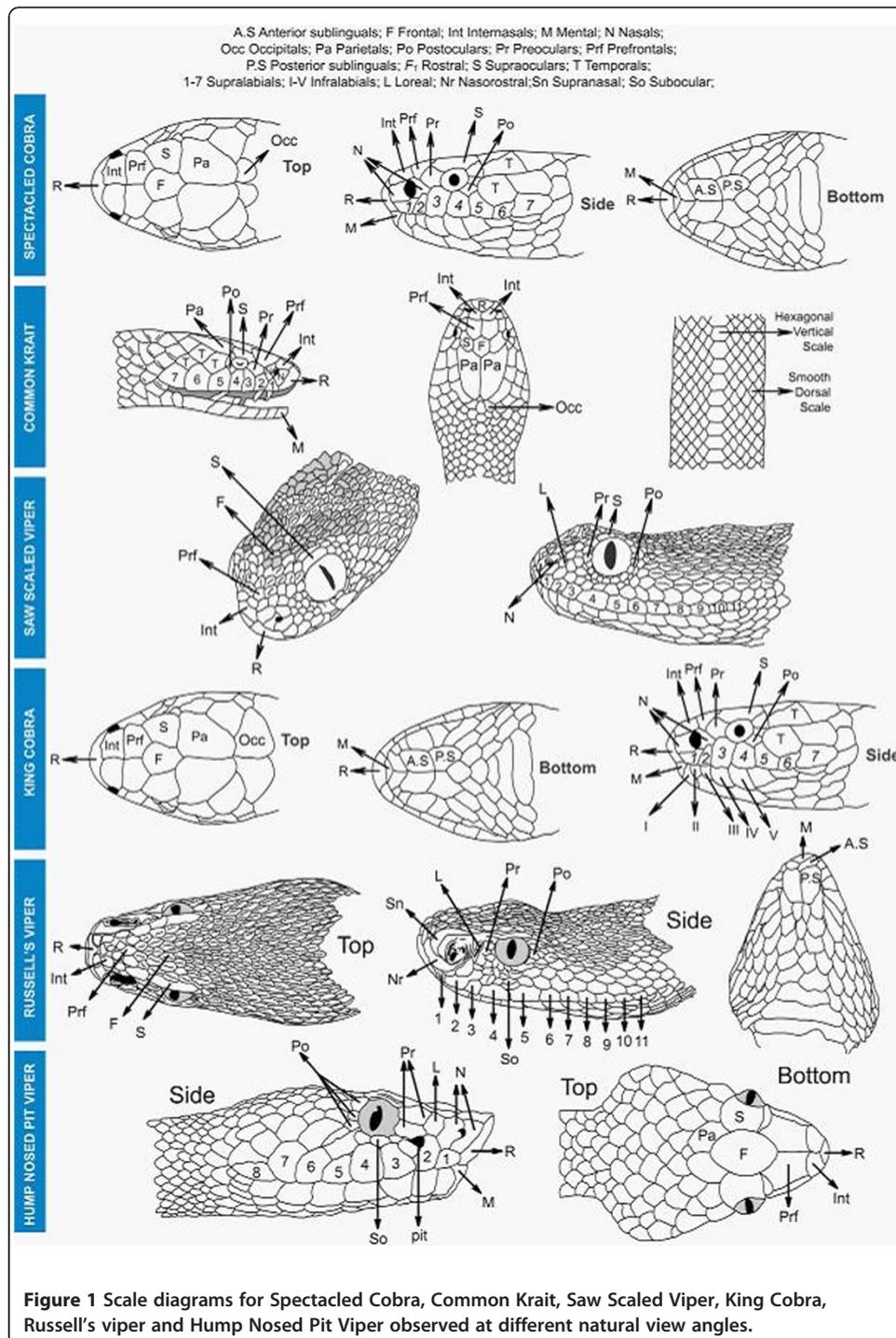
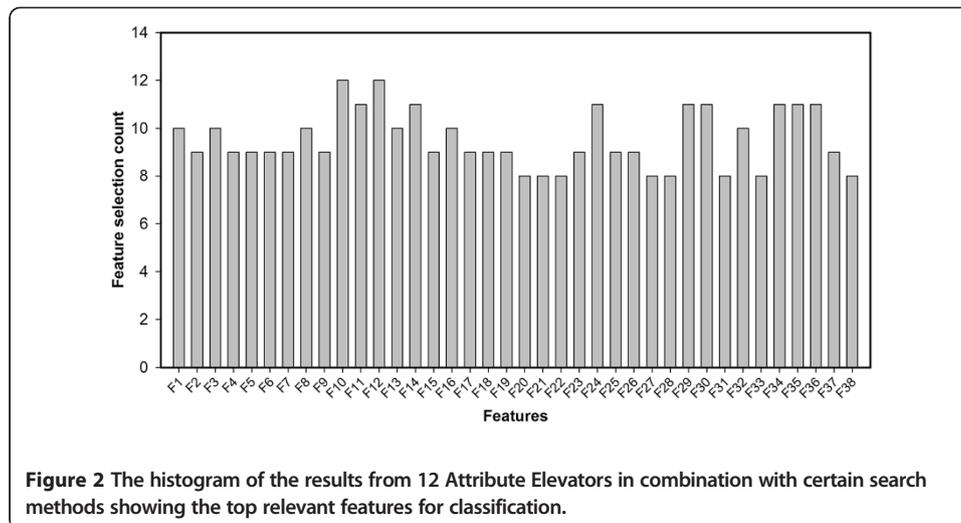


Figure 1 Scale diagrams for Spectacled Cobra, Common Krait, Saw Scaled Viper, King Cobra, Russell's viper and Hump Nosed Pit Viper observed at different natural view angles.



measures in such situations indicates classifiers applicability in practice. In our study, we use 5% of the samples from each snake class for the training set, while remaining 95% is selected as test set. The classifier that performs the best in terms of performance measures can be selected as a possible candidate for implementation.

The research and work submitted do conform to the guidelines for care and use of animals in scientific research. We've followed the guidelines published by Indian National Science Academy. The Ethics committee of Enview R&D Labs gave approval for the research work.

Results and discussion

The feature database of the snakes is as explained in Table 1 and Figure 1 is used for analysing the classification performance of this six class classification problem. The feature database contains 38 features of each sample. Now using Table 2, we perform our further experiments for databases with different feature size. The samples in the databases are randomly split into 5% samples in training set and 95% in test set and performance evaluated on individual classifiers. The selection of features is performed on the training set. To ensure statistical correctness, the selection and testing is repeated 100 times, and the resulted reported in Table 3. The testing is done such that test and training set are non-overlapping in samples. Table 3 shows the comparisons of average performance measures of 38 feature size databases. The performances indicated are percentage accuracy of correct classification, F-score value, the area under the receiver operator characteristic curve, precision and recall rates. Table 3 shows the variation of performance measures with the increase in feature size i.e. the number of features in the feature-subset. As shown in Table 3, the correct classification accuracy increases considerably till feature size 15 which contain top 15 features of rank list in the database and tend to drop from feature size 31. This proves that these top 15 features are alone enough for the automated snake identification instead of 38 taxonomically relevant features.

Tables 4 and 5 shows the performance of the automatic snake classification using Bayes Net [37], Naive Bayes [27], Multilayer perception [26], Ada BoostM1 [28], Multi BoostAB [29], RBF network [30], IB1 [31], IBk [31], LWL [32], NB Tree [33], J48 [34],

Table 2 Ranking of all the 38 features based on the results from 12 attribute elevators with certain search method and the average classification score taken individually for all the features

Feature number	Number of times repeated	Classification - score
F_{12}	30	43.66 ± 1.51
F_{29}	28	57.02 ± 1.27
F_{35}	28	47.95 ± 1.11
F_{34}	28	47.59 ± 1.23
F_{10}	27	51.16 ± 2.59
F_{24}	27	46.46 ± 3.70
F_{30}	25	47.51 ± 1.47
F_{36}	25	39.92 ± 1.89
F_{32}	23	54.79 ± 1.28
F_{11}	22	36.80 ± 1.13
F_{16}	20	50.60 ± 1.50
F_{13}	20	44.73 ± 2.23
F_{14}	17	50.23 ± 1.46
F_1	16	36.82 ± 1.84
F_8	15	45.80 ± 2.05
F_{25}	15	45.77 ± 3.11
F_{37}	15	36.24 ± 1.96
F_{23}	14	45.56 ± 2.88
F_{26}	14	43.14 ± 1.92
F_3	14	36.47 ± 1.42
F_{15}	12	50.23 ± 1.61
F_{33}	12	41.16 ± 2.01
F_{38}	12	37.50 ± 1.46
F_{17}	11	48.50 ± 1.78
F_9	11	41.22 ± 1.61
F_{27}	11	39.67 ± 2.48
F_{22}	11	37.81 ± 3.47
F_6	11	36.42 ± 1.40
F_{19}	10	49.05 ± 1.67
F_{18}	10	48.46 ± 1.91
F_{20}	10	48.02 ± 1.86
F_7	10	44.97 ± 2.06
F_{31}	10	36.83 ± 0.98
F_4	10	36.30 ± 1.48
F_2	10	36.26 ± 2.14
F_5	10	36.21 ± 1.58
F_{21}	9	48.60 ± 1.63
F_{28}	9	41.26 ± 2.48

Random Sub Space [35], and Bagging [36] classification methods for top 15 selected snake feature database and 38 snake feature database respectively. The performances indicated are percentage accuracy of correct classification, F-score value, the Area under the receiver operator characteristic curve, precision and recall rates. The RBF

Table 3 Comparison of average classification result from 13 classifier in different feature size snake database with 5% train and 95% test of the total samples

Feature size	% Correct	F-score	AUC	Precision (%)	Recall (%)
1	43.69 ± 1.51	0.55 ± 0.00	0.55 ± 0.02	39 ± 0.02	95 ± 0.03
2	63.16 ± 1.83	0.81 ± 0.01	0.75 ± 0.02	64 ± 0.02	91 ± 0.03
3	72.41 ± 2.38	0.89 ± 0.01	0.84 ± 0.03	79 ± 0.03	92 ± 0.02
4	73.73 ± 3.01	0.93 ± 0.01	0.88 ± 0.03	86 ± 0.05	93 ± 0.01
5	75.23 ± 2.65	0.93 ± 0.01	0.88 ± 0.03	85 ± 0.05	93 ± 0.02
6	76.81 ± 2.83	0.94 ± 0.01	0.89 ± 0.03	85 ± 0.04	95 ± 0.02
7	77.21 ± 2.82	0.94 ± 0.01	0.89 ± 0.02	85 ± 0.03	94 ± 0.01
8	77.86 ± 2.84	0.94 ± 0.01	0.90 ± 0.03	86 ± 0.04	94 ± 0.01
9	77.89 ± 2.95	0.94 ± 0.02	0.89 ± 0.03	84 ± 0.05	96 ± 0.03
10	77.82 ± 2.96	0.94 ± 0.02	0.89 ± 0.03	84 ± 0.05	95 ± 0.03
11	77.65 ± 2.92	0.94 ± 0.02	0.89 ± 0.03	84 ± 0.05	91 ± 0.03
12	77.76 ± 3.03	0.94 ± 0.02	0.88 ± 0.03	84 ± 0.05	92 ± 0.03
13	78.05 ± 2.97	0.94 ± 0.02	0.88 ± 0.03	83 ± 0.05	93 ± 0.03
14	77.66 ± 2.95	0.94 ± 0.02	0.88 ± 0.03	83 ± 0.05	93 ± 0.03
15	78.80 ± 3.05	0.94 ± 0.02	0.89 ± 0.03	83 ± 0.05	95 ± 0.03
16	78.19 ± 3.14	0.94 ± 0.02	0.88 ± 0.03	83 ± 0.05	94 ± 0.04
17	78.31 ± 3.03	0.94 ± 0.02	0.88 ± 0.03	83 ± 0.05	94 ± 0.04
18	78.29 ± 3.06	0.94 ± 0.02	0.88 ± 0.03	83 ± 0.05	96 ± 0.04
19	78.05 ± 2.93	0.94 ± 0.02	0.88 ± 0.03	83 ± 0.05	96 ± 0.04
20	78.00 ± 2.94	0.94 ± 0.02	0.88 ± 0.03	83 ± 0.05	96 ± 0.04
21	78.44 ± 2.55	0.94 ± 0.02	0.88 ± 0.03	83 ± 0.05	96 ± 0.04
22	78.52 ± 2.74	0.94 ± 0.02	0.88 ± 0.03	83 ± 0.05	96 ± 0.04
23	78.61 ± 2.80	0.94 ± 0.02	0.88 ± 0.03	83 ± 0.05	96 ± 0.04
24	78.47 ± 2.79	0.94 ± 0.02	0.88 ± 0.03	82 ± 0.05	96 ± 0.04
25	78.41 ± 2.84	0.94 ± 0.02	0.88 ± 0.03	82 ± 0.05	95 ± 0.04
26	78.24 ± 2.85	0.94 ± 0.02	0.87 ± 0.03	82 ± 0.05	95 ± 0.04
27	78.10 ± 2.87	0.94 ± 0.02	0.87 ± 0.03	82 ± 0.05	95 ± 0.04
28	78.09 ± 2.89	0.94 ± 0.02	0.87 ± 0.03	82 ± 0.05	95 ± 0.04
29	78.03 ± 2.90	0.94 ± 0.02	0.87 ± 0.03	82 ± 0.05	95 ± 0.04
30	78.05 ± 2.89	0.94 ± 0.02	0.87 ± 0.03	82 ± 0.05	95 ± 0.04
31	77.99 ± 2.92	0.94 ± 0.02	0.87 ± 0.03	82 ± 0.05	95 ± 0.04
32	77.97 ± 2.92	0.94 ± 0.02	0.87 ± 0.03	82 ± 0.05	95 ± 0.04
33	77.96 ± 2.94	0.94 ± 0.02	0.87 ± 0.03	82 ± 0.05	95 ± 0.04
34	77.90 ± 2.91	0.94 ± 0.02	0.87 ± 0.03	82 ± 0.05	95 ± 0.04
35	77.83 ± 2.95	0.94 ± 0.02	0.87 ± 0.03	82 ± 0.05	95 ± 0.04
36	77.74 ± 2.99	0.93 ± 0.02	0.87 ± 0.03	82 ± 0.05	95 ± 0.04
37	77.67 ± 2.98	0.93 ± 0.02	0.87 ± 0.03	82 ± 0.05	94 ± 0.04
38	77.55 ± 2.96	0.93 ± 0.02	0.87 ± 0.03	82 ± 0.05	94 ± 0.04

network, IBk and IB1 classifiers showed higher classification performance as opposed other classifiers. The classification accuracy of above 85% indicates robustness of the taxonomically relevant features in the automatic classification process. Multilayer perception [26], RBF Network [30], IB1 [31], IBk [31], and J48 [34] shows good recognition

Table 4 Comparison of different classifiers when 5% of the class samples are used as gallery and remaining 95% of sample are used as test on top 15 selected snake feature database

Method	% Correct	F-score	AUC	Precision (%)	Recall (%)
Bayes net [37]	81.26 ± 4.00	0.98 ± 0.01	0.92 ± 0.03	88 ± 0.05	96 ± 0.04
Naïve Bayes [27]	81.64 ± 3.05	0.98 ± 0.01	0.93 ± 0.03	91 ± 0.04	96 ± 0.04
Multilayer perceptron [26]	86.64 ± 2.71	0.97 ± 0.01	0.92 ± 0.02	90 ± 0.04	95 ± 0.03
Ada BoostM1 [28]	57.52 ± 1.27	0.80 ± 0.03	0.75 ± 0.02	63 ± 0.04	95 ± 0.04
Multi BoostAB [29]	57.52 ± 1.27	0.80 ± 0.03	0.75 ± 0.02	63 ± 0.04	95 ± 0.04
RBF network [30]	88.75 ± 2.69	0.97 ± 0.02	0.94 ± 0.02	93 ± 0.04	96 ± 0.03
IB1 [31]	86.05 ± 3.35	0.93 ± 0.03	0.91 ± 0.04	89 ± 0.04	94 ± 0.07
IBk [31]	87.50 ± 2.35	0.95 ± 0.01	0.92 ± 0.02	88 ± 0.04	96 ± 0.03
LWL [32]	69.57 ± 4.17	0.97 ± 0.01	0.86 ± 0.04	77 ± 0.06	96 ± 0.03
J48 [33]	84.71 ± 2.90	0.95 ± 0.02	0.91 ± 0.03	87 ± 0.06	96 ± 0.03
Random sub space [34]	79.77 ± 4.01	0.98 ± 0.01	0.89 ± 0.03	82 ± 0.06	98 ± 0.02
Bagging [35]	81.34 ± 3.93	0.97 ± 0.02	0.90 ± 0.03	85 ± 0.06	96 ± 0.03
NB Tree [36]	82.10 ± 4.02	0.96 ± 0.02	0.91 ± 0.03	87 ± 0.05	96 ± 0.03

performance among the tested classifiers at 5% training data. While increasing the training dataset size to 30% the multilayer perception [26] classifier results in 94.31 ± 1.00% classification accuracy. The results indicate the difficulty of automatic classification of snakes, nonetheless, is indicative of the practical use in as a first line prediction of the snake classification. These early results opens up two major directions of research: (1) as to identify the taxonomy features of unknown snakes using feature automatic feature analysis and (2) to develop accurate feature classification and recognition methods for automatic snake. To use of real-time applications such as in diagnosis an ambitious 100% accuracy is preferred, which is by far a challenging problem posed through these results. In addition, the results on 5% training data, is likely to be more useful in real-time systems as in real applications the size of the test data keeps on

Table 5 Comparison of different classifiers when 5% of the class samples are used as gallery and remaining 95% of sample are used as test on 38 snake feature database

Method	% Correct	F-score	AUC	Precision (%)	Recall (%)
Bayes net [37]	78.81 ± 2.27	0.98 ± 0.01	0.89 ± 0.03	83 ± 0.07	97 ± 0.03
Naïve Bayes [27]	77.69 ± 2.11	0.97 ± 0.01	0.89 ± 0.02	90 ± 0.05	89 ± 0.03
Multilayer perceptron [26]	86.85 ± 2.59	0.98 ± 0.01	0.92 ± 0.02	90 ± 0.04	94 ± 0.04
Ada BoostM1 [28]	57.39 ± 1.44	0.80 ± 0.03	0.75 ± 0.02	62 ± 0.04	95 ± 0.05
Multi BoostAB [29]	57.39 ± 1.44	0.80 ± 0.03	0.75 ± 0.02	62 ± 0.04	95 ± 0.05
RBF network [30]	85.00 ± 3.05	0.95 ± 0.02	0.91 ± 0.03	92 ± 0.05	89 ± 0.04
IB1 [31]	85.82 ± 2.45	0.93 ± 0.02	0.91 ± 0.02	88 ± 0.04	93 ± 0.04
IBk [31]	86.38 ± 2.47	0.94 ± 0.01	0.91 ± 0.02	88 ± 0.04	94 ± 0.03
LWL [32]	68.37 ± 6.26	0.97 ± 0.01	0.82 ± 0.05	72 ± 0.07	96 ± 0.03
NB Tree [36]	83.79 ± 2.87	0.95 ± 0.02	0.91 ± 0.03	88 ± 0.05	95 ± 0.04
J48 [33]	78.92 ± 4.37	0.98 ± 0.01	0.87 ± 0.04	80 ± 0.06	97 ± 0.03
Random sub space [34]	80.50 ± 3.29	0.97 ± 0.02	0.90 ± 0.03	85 ± 0.06	96 ± 0.04
Bagging [35]	80.91 ± 4.48	0.94 ± 0.03	0.88 ± 0.04	83 ± 0.06	93 ± 0.04

growing at a rate higher than the training data, mainly because of the labor intensive processes involved in the preparation and validation of the training data.

Conclusion

In this paper, we presented an automatic snake identification problem by developing a taxonomy based feature targeted for use by the computer scientist and herpetologist. The feature-subset analysis indicated that only 15 features are sufficient for snake identification. In a real-life situation, the snake feature database reflects a situation when the bite victim has seen the snake, and based on the observed features it is required to identify the class of the snake. In addition to the venom detection research required for treating the bite victims, the proposed automatic snake recognition method could provide valuable information to administer correct medication and treatment in life threatening situation. Survey of snakes in wild is another major activity in the process to ensure the preservation of snake population and diversity. This is however a very challenging task and require prohibitive investments in manpower. The automatic classification using snake image database can be extended to the analysis of snake images captured remotely with minimal human intervention. The progress in snake taxonomy research is in the decline for the last 60 years, and has resulted in lack of expertise for environmental surveys and help required for medical practitioners in emergency situations. With a computerized analysis on the images of snakes using the proposed database and classification approach, we hope that more studies would come out to generate interest on this topic.

Additional files

Additional file 1: Snake Feature Database. <http://www.hcis-journal.com/imedia/2229930941022190/supp1.doc>.

Additional file 2: Supplementary info. <http://www.hcis-journal.com/imedia/1032655649102219/supp2.xls>.

Competing interests

We declare that there are no competing interests.

Authors' contributions

AJ carried out the problem formulation, algorithm development and drafted the manuscript. BM organized the dataset and performed the feature analysis. SS helped in the implementation of algorithm. DR collected the original snake images. All authors read and approved the final manuscript.

Acknowledgements

The authors thank the snake catchers in Trivandrum for the assistance with the creation of the database. The assistance of Balaji Balasubramaniam (TRDDC) and Anaswara Krishnan (Department of Zoology, Kerala University) is also acknowledged. Dileep Kumar R would like to acknowledge the support of Prof Ommen V Ommen for the encouragement and support for this research.

Author details

¹School of Engineering, Nazarbayev University, Astana, Kazakhstan. ²Enview R&D Labs, Thiruvananthapuram, India. ³Department of Computational Biology and Bioinformatics, University of Kerala, Trivandrum, India.

Received: 20 June 2013 Accepted: 17 February 2014

Published online: 28 March 2014

References

1. Smith MA (1981) Reptilia and Amphibia, vol 3, Serpentes. Today & Tomorrow's Printers & Publishers, India
2. Whitaker R, Captain A, Ahmed F (2004) Snakes of India: the field guide. Draco Books, Chengalpattu
3. Mattison C (1999) Snake. Dorling Kindersley, New York, USA
4. Firth SMJWJR (2002) Snake. Scholastic, India
5. Weidensaul S (1996) Snakes of the World. Grange Books Ltd, Chartwell House, London
6. Mertens T (1995) Deadly & Dangerous Snakes. Magic Bean. Era Publications, Flinders Park, South Australia
7. Backshall S (2007) Venomous Animals of the World. Johns Hopkins University Press, Maryland, USA

8. Stevens D (2011) The Big Four Snakes: The Indian Cobra, the Common Krait, the Russell's Viper, and the Saw-Scaled Viper. Webster's Digital Services, USA
9. Premawardhena A, De Silva C, Fonseka M, Gunatilake S, De Silva H (1999) Low dose subcutaneous adrenaline to prevent acute adverse reactions to antivenom serum in people bitten by snakes: randomised, placebo controlled trial. *BMJ: Brit Med J* 318(7190):1041
10. Warrell DA (1999) The clinical management of snake bites in the Southeast Asian region. *Southeast Asian J Trop Med Public Health* 1(Suppl 1):1–89
11. Calvete JJ, Ju'arez P, Sanz L (2007) Snake venomics. Strategy and applications. *J Mass Spectrom* 42(11):1405–1414
12. Sorower MS, Yeasin M (2007) Robust Classification of Dialog Acts from the Transcription of Utterances. In *ICSC 2007. IEEE International Conference on Semantic Computing*, 3-10
13. Chanda P, Cho YR, Zhang A, Ramanathan M (2009) Mining of attribute interactions using information theoretic metrics. In: *Data mining workshops, ICDMW'09. IEEE International Conference on Data Mining, Florida, USA*, pp 350–355
14. Devi MI, Rajaram R, Selvakuberan K (2008) Generating best features for web page classification. *Webology* 5, <http://www.webology.org/2008/v5n1/a52.html>
15. Marquez-Vera C, Romero C (2011) Ventura S: Predicting school failure using data mining. In *Proceedings of the 4th International Conference on Educational Data Mining* 271–276
16. John GH, Kohavi R, Pfleger K (1994) Irrelevant features and the subset selection problem. In *Proceedings of the eleventh international conference on machine learning*, Volume 129, San Francisco 121–129
17. Jensen R, Shen Q (2007) Fuzzy-rough sets assisted attribute selection. *Fuzzy Systems, IEEE Transactions on* 15:73–89
18. Meng YX (2011) The practice on using machine learning for network anomaly intrusion detection. In *IEEE International Conference on Machine Learning and Cybernetics (ICMLC)*, 2011, Vol. 2, 576-581
19. Indra Devi M, Rajaram R, Selvakuberan K (2007) Automatic web page classification by combining feature selection techniques and lazy learners. In *conference on computational intelligence and multimedia applications, 2007. Int Conference on* 2:33–37
20. Koonsanit K, Jaruskulchai C (2011) Band selection for hyperspectral image using principal components analysis and maxima-minima functional. In: *Knowledge, Information, and Creativity Support Systems. Thailand, Springer*, pp 103–112
21. Frank E, Hall M, Holmes G, Kirkby R, Pfahringer B, Witten IH, Trigg L (2005) Weka. In: *Data Mining and Knowledge Discovery Handbook. Springer, USA*, pp 1305–1314
22. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH (2009) The WEKA data mining software: an update. *ACM SIGKDD Explorations Newsletter* 11:10–18
23. James AP, Dimitrijević S (2012) Ranked selection of nearest discriminating features. *Hum-centric Comput Inform Sci* 2:12
24. Milacic M, James AP, Dimitrijević S (2013) Biologically inspired features used for robust phoneme recognition. *International Journal of Machine Intelligence and Sensory Signal Processing* 1(1):46–54
25. James AP, Maan AK (2011) Improving feature selection algorithms using normalised feature histograms. *Electron Lett* 47(8):490–491
26. Longstaff ID, Cross JF (1987) A pattern recognition approach to understanding the multi-layer perception. *Pattern Recogn Lett* 5(5):315–319
27. Kim SB, Han KS, Rim HC, Myaeng SH (2006) Some effective techniques for naive bayes text classification. *Knowledge and Data Engineering, IEEE Transactions on* 18(11):1457–1466
28. Freund Y, Schapire RE (1995) A decision-theoretic generalization of on-line learning and an application to boosting. In *Computational learning theory, Springer* 23–37
29. Benbouzid D, Busa-Fekete R, Casagrande N, Collin FD, Kégl B (2012) MultiBoost: a multi-purpose boosting package. *J Mach Learn Res* 13:549–553
30. Buhmann MD (2003) Radial basis functions: theory and implementations, Volume 12. Cambridge university press
31. Aha DW, Kibler D, Albert MK (1991) Instance-based learning algorithms. *Machine learning, Boston, USA*, pp 37–66, 6
32. Atkeson CG, Moore AW, Schaal S (1997) Locally weighted learning for control. *Artif Intell Rev* 11(1–5):75–113
33. Kohavi R (2001) Bayes rule based and decision tree hybrid classifier. [US Patent 6,182,058]
34. Kotsiantis SB, Zaharakis ID, Pintelas PE (2006) Machine learning: a review of classification and combining techniques. *Artif Intell Rev* 26(3):159–190
35. Ho TK (1998) The random subspace method for constructing decision forests. *Pattern Anal Mach Intel, IEEE Transactions on* 20(8):832–844
36. Breiman L (1996) Bagging predictors. *Mach Learn* 24(2):123–140
37. Singhal A, Brown C (1997) Dynamic Bayes net approach to multimodal sensor fusion. In *Proceedings of the SPIE-The International Society for Optical Engineering, Volume 3209*, 2–10

doi:10.1186/s13673-014-0003-0

Cite this article as: James et al.: Discriminative histogram taxonomy features for snake species identification. *Human-centric Computing and Information Sciences* 2014 4:3.