

# The Taxonomy of Living Organisms Using Self-organizing Map

**Adebayo Rotimi Philip**

Department of Computer Science, University of Lagos, Akoka, Yaba, Lagos, Nigeria

**Email address:**

rotphilipadeb@yahoo.com

**To cite this article:**

Adebayo Rotimi Philip. The Taxonomy of Living Organisms Using Self-organizing Map. *American Journal of Artificial Intelligence*. Vol. 4, No. 2, 2020, pp. 50-61. doi: 10.11648/j.ajai.20200402.12

**Received:** May 30, 2020; **Accepted:** June 15, 2020; **Published:** September 7, 2020

---

**Abstract:** The Self Organizing Map (SOM) is an unsupervised network algorithm that projects high dimensional data into low dimensional maps. The projection preserves the topology of the data so that similar data items are mapped to nearby locations on the map. The algorithm has been so popular because of its application in Computer Science and other areas; it has been applied in speech recognition, pattern identification, control engineering, earthquake detection et al. This research aimed to apply the SOM in the taxonomy of living organisms using 46 attributes. 68 animals from 6 phyla were considered and 46 attributes were used detailing their physical features, physiological features, evolution, adaptation, habitat et al. The features extracted were converted to 0s and 1s for the SOM algorithm to process. The result shows 96.569% accuracy of the SOM's classification but better accuracy can be obtained if the SOM had processed the data for about 1000 iterations. This research revealed that SOM is a veritable tool or algorithm that can be used to classify living organisms. This research will help taxonomists, biologists and students who spend much time in classifying living organism and it will be of help to researchers who want to explore the SOM algorithm as a solution to taxonomy of living organisms. The SOM will ease taxonomy and will help to minimize the stress and time involved in classifying thousands of living organisms.

**Keywords:** Self-organizing Map, Taxonomy, Unsupervised Neural Network, Classification

---

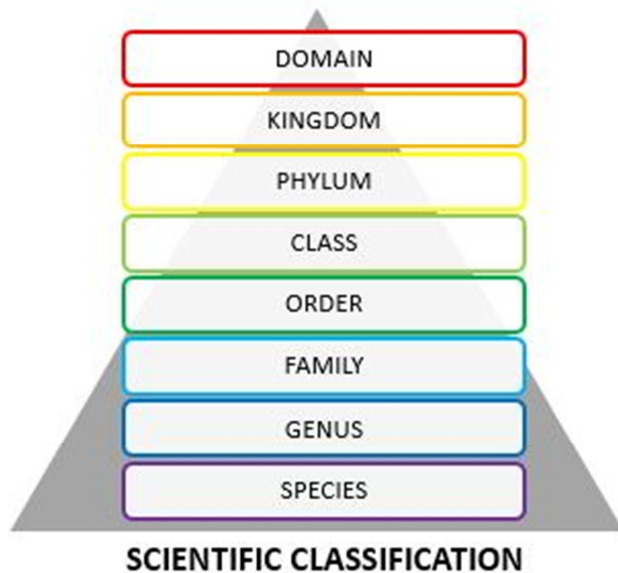
## 1. Introduction

Taxonomy is the biological area that recognizes reports, classifies and names the discovered and undiscovered species and other taxa [1] and classification is one of the data mining techniques that is used to categorize samples into different classes based on supervised learning [2]. Classification, identification, naming, and grouping of organisms into a formal system are based on similarities such as internal and external anatomy, physiological functions, genetic makeup, and evolutionary history. With an estimated 10 million to 13 million species on Earth, out of which 6.5 million [3] species are on land and 2.2 million in the oceans, the diversity of life is immense. Determining an underlying order in the complex web of life is a difficult undertaking that encompasses advanced scientific methods as well as fundamental philosophical issues about how to view the living world. There is currently no consensus among the world's taxonomists concerning which classification scheme to use for the overall hierarchy of life [4].

A lot of researches had been done on the classification of living organisms. Classification is dated back to the Latin philosopher: Aristotle (more than 2300 years ago), he was the first who classified animals into red-blooded animals and bloodless animals, plants into trees, shrubs and weeds [5]. Swedish scientist Carl Linnaeus created and published in 1758 the system which he used to formally name and describe species [3] and he established the traditional classification system that classify living organisms into two kingdoms only, which are animal kingdom and plant kingdom [5]. The scientist: Robert Whittaker was the first to propose the five kingdom taxonomic classification of the world's biota into the Animalia, Plantae, Fungi, Protista, and Monera in 1969 [6].

The modern day hierarchy of classification has classified organisms into Domain (3), Kingdom (6), Phylum (2), class (4), Order, Family, Genus, and Species [7]. The higher classifications are by domain and kingdom levels; however, the deepest and the most detailed classification are the genus and species levels [1]. The hierarchical levels in-between

include phylum, subphylum, class, family and order [8]. With all the problems of classification of living organisms, I propose an unsupervised artificial neural network model-Self-Organizing map- to classifying living organisms into their Phyla. This idea of classifying organisms using Self-organizing map is robust.



**Figure 1.** Modern Classification of Living organisms (Divya Pochimcherla (2019), scientific classification).

## 2. Literature Review

Self-organizing map (SOM), sometimes also called a Kohonen map, is an unsupervised neural network algorithm, which learns competitively to produce low dimensional, discretized representation of high dimensional data, while simultaneously preserving [9] similarity relations between the presented data items. A lot of research has been done about Self-organizing map and its applications. Self-organizing map was discovered by Professor T Kohonen in 1980. Since its introduction by T. Kohonen in his seminal 1982 articles (Kohonen (1982)), the self-organizing map (SOM) algorithm has been widely use. This is due to the easiness of its practical development, to its clustering properties as well as its visualization ability [10]. He published his first book on SOM in March 2 1995. T Kohonen (1990) delivered in a conference: [11] the Self-organizing map algorithm (an algorithm which order responses spatially) review, focusing on best matching cell selection and adaptation of the weight vectors and the demonstrations of the ordering process. He addressed the Fine tuning of the map by learning vector quantization and discussed the practical speech recognition and a simulation experiment on semantic mapping. Teuvo Kohonen [12] demonstrated in one of his papers that an input item can be more accurately represented by a linear mixture of a few best-matching models. This becomes possible by a least-squares fitting procedure where the coefficients in the linear

mixture of models are constrained to nonnegative values. There have been extensive researches since the introduction of SOM.

Umut A and Secil E [13] provided a general introduction to the structure, algorithm and quality of Self-organizing Maps and presents industrial engineering related applications. Merja O and Kaski S did an extensively bibliography on SOM papers from 1998 to 2001. They collected a comprehensive list of 5384 scientific papers that use and analyze SOM algorithms [14]. Dubravko M, Hrvatska E, and Zagreb C [9] did a brief review paper basic tenet, including motivation, architecture, math description and applications of SOM.

Self-organizing map has been applied in various fields: Pollock R, Toby L, and Michael W delivered a conference paper in January 2002. They described [15] the use of a Kohonen Self-organizing Map (SOM) to categorize proteins based on secondary structure, and they relate this information to functional data. Lars B and Bjorn G [16] argued that the Self-Organizing Map is well suited to classify a document collection in which many documents simultaneously belong to several categories. Marie Cottrell, Madalina O, Fabrice R, and Nathalie V [10] reviewed some of the many SOM algorithm variants which were defined to overcome the theoretical difficulties and adapt the algorithm to the processing of complex data such as time series, missing values in the data, nominal data, textual data, et al.

Derajad H and Saruni D [17] used the SOM algorithm to determine the earthquake pattern criteria and which areas it often occurred in order to be able to mitigate earthquake that causes fatal impacts. Faigl J [18] applied SOM for the Multiple Traveling Salesman Problem (MTSP) with minmax objective to unravel the robotic problem of multi-goal path planning in the polygonal domain. The [19] factors influencing soil fertility in Shiraz plain, southern Iran was classified by Marzieh M, Mahdi N and Abdol R. The relationships among soil features were studied using the SOM in which the clustering tendency of soil fertility was investigated using seven parameters (N, P, K, Fe, Zn, Mn, and Cu). Subana S, Sallis P and Buckeridge J [20] researched on methods to analyze environmental effects integrated with human activities. Yuan-Chao L, Ming L and Xiao-Long W use [21] SOM to create a text clustering. Text clustering is important means of effective organization and navigation of text information.

## 3. Architecture Self-Organizing Map

SOM is an efficient algorithm in visualizing data by reducing its dimensions from n-dimensional input to a lower dimension while maintaining its original topology relationship [18]. The SOM architecture consists of two layers [13] of nodes, namely the input layer and the output layer. Each neuron in the input layer is connected to each neuron in the output layer [18]. Then, each neuron in the output layer represents a class (cluster) of the input given.

The input layer is  $X = \{x_{1,1}, x_{1,2}, x_{1,3}, x_{1,4}, \dots, x_{1,j}\}$

$$\{x_{p,1}, x_{p,2}, x_{p,3}, x_{p,4}, \dots, x_{p,j}\}$$

Where  $j$  is the length of the vector  $X$  or the number of attributes and  $p$  represents the number of animals. In this architecture  $p = 68$  while  $j = 46$ . Each input is attached to neurons and the neurons are connected together and each neuron has a weight which determine the special location of the neuron. Overtime, the

weights  $W = \{w_{1,1}, w_{2,2}, \dots, w_{1,j}\} w_{n,1}, w_{n,2}, \dots, w_{n,j}\}$  get trained and updated to change the position of the neurons into clusters, where  $W \in R$ .  $W$  is a set of real numbers that are randomly selected. In this architecture  $i=6$  and  $j=46$ . The architecture of our SOM is shown below. Each circle represents a dimension and we have a total of 46 dimensions. The output of our architecture is defined as  $\{y_1, y_2, y_3, \dots, y_i\}$ , where  $i=6$

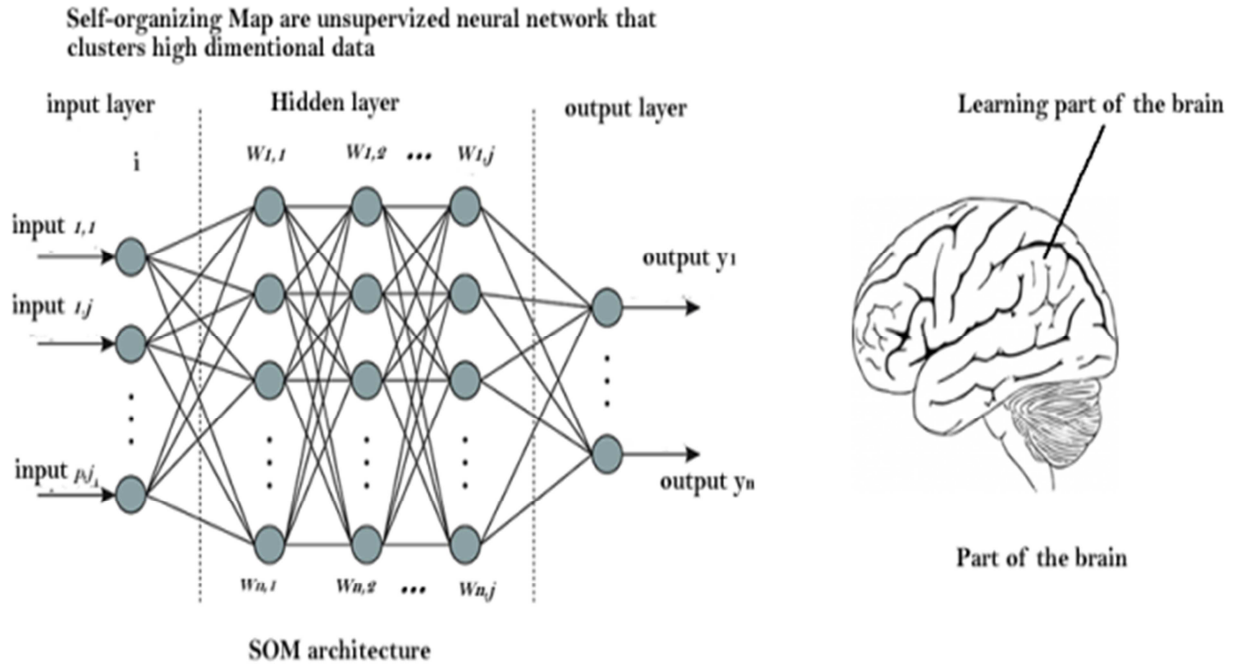


Figure 2. SOM Architecture.

An example of SOM in nature is the brain. Our brains take input in form of smell, noise et al and map them topological distinct areas. The training process is as follow

[Initialize weight]: initialize Weight  $w_{ij}$  ( $1 < i < N$ ) to small random values. Set the initial radius of the neighbour around  $j$  to  $N_j(t)$

[present input]: the vector  $x_0, x_1, \dots, x_{N-1}$ , where  $x_i$  is the input to node  $j$  at time  $t$

[Calculate the distance  $d_j$ ]: calculate the distance between input and each node  $j$ , given by  $d_j = \sum_{i=1}^N [x_i(t) - w_{ij}(t)]^2$

[determine  $D_j$ ]:  $D_j$  is the minimum value of  $d_j$

[Update Weight]: update weight for node  $D_j$  and its neighbour define by  $N_j(t)$ . New weight are  $\Delta w_{ij} = \eta(t) * T_{j,i(x)}(t) * [x_i - w_{ij}(t)]$  for  $j$  in  $N_j(t)$ , where  $t$ =epoch,  $i$  = a neuron and  $j$  =another neuron,  $i(x)$  =the winning neuron.

$\eta(t)$  =Learning rate given as  $\eta(t) = \eta_0 \exp(-t/\tau_\eta)$  and  $T_{j,i(x)}(t)$  is the

Topological neighborhood, given by  $T_{j,i(x)}(t) = \exp(-S_{j,i}^2 / 2\sigma(t)^2)$ .

$S_{j,i}$  is a lateral distance given as  $S_{j,i} = ||w_j - w_i||$ .

Neighborhood size  $\sigma = \sigma_0 \exp(-t/\tau_\sigma)$

Repeat by going to step 2 if the stopping condition  $t = n$  is not satisfied, where  $n$  is the number of iterations

In the topological neighborhood formula  $T_{j,i(x)}(t) =$

$\exp(-S_{j,i(x)}^2 / 2\sigma(t)^2)$ ,  $j$  and  $i$  are winning neurons. So the distance between the neurons is 0. So the topological neighborhood  $T_{j,i(x)}(t)$  is 1. The topological neighborhood is important for magnifying the learning rate and its value ranges from 0 to 1.

## 4. Research Methodology

Living organisms are classified into six kingdoms: Archaeobacteria, Eubacteria, Plantae, Animalia, Fungi, and Prokista. However, this project only considered the common six phyla in the kingdom Animalia. These phyla are invertebrates, fishes, amphibians, birds, reptiles, and mammals. I collected about 65 animals from available animals and then I studied extensively the characteristics of each animal using 45 attributes.

Data Collection: I collected 65 animals and extract the features of the animals. The attributes are considered first from the general features of each phylum. I later comprehensively considered the characteristics of each animals used in the SOM algorithm.

Features extraction: There are general characteristics and natural behavior of animals from which other basic features are extracted. I considered the features from these 9 categories.

Physical characteristics: these include sizes, feather, nails,

hair, tails, claws, beaks, mane etc.

**Physical adaptation:** these features enable the organisms to live successfully in their habitat. Examples in birds are light weight and light bone (furculum), toothless, claws, streamlined body for easy flight, sharp sight etc.

**Physiology:** the study of the internal systems and organs of living organisms.

**Habitat:** animals can be aquatic, terrestrial, or boreal.

**Mode of feeding:** carnivores, herbivores, omnivores.

**Reproduction:** some animals lay eggs (oviparous) while some give birth to their young ones (viviparous).

**Evolution:** the scientific origin of organisms.

**Locomotion:** some animals fly, some swim, some walk, some crawl etc.

**Sizes:** the sizes of living organisms vary. For examples the Hummingbird measures about 2.25 inches and weighs 1.6g, while the Ostrich is about 2.7m tall and weighs 127kg.

In this reasearch I used 45 basic characteristic to classify animals into phyla. I took a sample of 65 common animals making a total of 2,925 inputs. The common attributes of most animals are highlighted below.

**Vertebrates:** animals with backbones.

**Invertebrates:** animals without backbones.

**Ectothermy:** these are animals in which their body temperature changes with the environment.

**Endothermy:** these are animals in which their body temperature remains fairly the same irrespective of their environment.

**Terrestrial animals:** animals that live on land.

**Aquatic animals:** animals that live in water.

**Arboreal animals:** animals that live on trees

**Gill:** a gaseous exchange organ in some animals, namely aquatic animals.

**Lungs:** gaseous exchange organ in terrestrial animals.

**Tetrapods:** these are animals with four limbs. Some amphibians and snakes are believed to be the descendant of four limbs ancestors.

**Possession of scales and scrutes**

**Amniotes:** animals that produce egg with amnions (an elastic sac within which the embryo develops).

**Oviparous animals:** animals that lay eggs

**Viviparous:** animals that give birth to their young ones.

**Nocturnal:** animals that is more active in the night than day time.

**Carnivores:** flesh eater

**Herbivores:** plant eater

**Omnivores:** animals that eat fresh and herbs.

**Possession of hair.**

**Mammary gland:** a milk producing organ possessed by mammals.

**Diphyodonty:** is a pattern of tooth replacement in animals in which the tooth is replace only once in a life time. This features is possessed by only mammals

**Polyphydonty:** is a pattern of tooth replacement in animals in which the tooth is continuously replaced throughout the life time of the animals.

**Diaphragm:** a layer of muscles located at the base of the

rib cage that separates the thoracic cavity from the abdominal cavity. Mammals are the only animals that possess diaphragm.

**Four Heart Chambers:** found only in mammals.

**Three Heart Chambers:** found in birds.

**Two Heart Chambers:** found in fishes and reptiles.

**Lower jaw:** possessed by only mammals.

**Feather**

**Beak or Bill**

**Furcula:** a forked bone formed by the fusion of the clavicles in birds.

**Tracheae:** the respiratory system of insects.

**Insectivores:** insect-eating animals

**Webs:** a fold of tissue connecting the toes of some certain birds and amphibians.

**Antenna:** a feeler organ on the head of an insect, crab, or other animal.

**Wings:** appendages of animals' body (insect, bat, birds) that enables them to fly.

**Exoskeleton:** skeleton outside the body. They are basically possessed by insects.

**Proboscis:** an elongated tube from the head or connected to the mouth of insects for sucking fluid.

**Crop:** a pouch-like part of the alimentary tract of some birds and insects used to store food before digestion or for regurgitation.

**Tails**

**Jacobson's organ:** this is a sense of smell and taste in amphibians.

**Cocoon:** a silky protective case spun by the larvae of some insects and moths in which they metamorphose into the pupa.

**Fins:** one of the appendages of a fish use to propel in water.

**Swim bladder:** an organ in which that enables them to float at any desired depth without spending extra energy to swim in place.

**Thorax:** the middle of the three distinct divisions in an insect, crustacean and arachnid body

**Metamorphosis:** a change that occur during the life cycle of an insect. E.g. the change from egg to larva, larva to pupa, pupa to adult insect.

**Feature Conversion**

This section illustrate the conversion of the 46 features alighted above into 1s and 0s for SOM Algorithm to process. We took a sample of 68 different animals from different phyla. Each row is a schematic description of the features of an animal, based on the presence (=1) or absence (=0) of some of the 46 different attributes. The attributes are the features of the animal described; they are presented in the table 6 (see page by 7). Each row represent a pattern vector  $x$  that act as a set of signal values at the inputs of the network. The animal name itself does not belong to the vector but instead specifies the label of the animal in the calibration of the map. These categories are the values of the weight obtained after some iteration. In our algorithm, we limit our iteration to 172.

## 5. Analysis of the Result

The output for the architecture is 6 and I did 172 numbers of iterations or epochs. The SOM algorithm was run four

times and different category numbers which represent similar pattern are displayed. The outputs with the category numbers are displayed below:

*Table 1. Outcome of three classifications after the SOM process.*

PATTERNS	ANIMAL SAMPLES	R1 (CATS)	R2 (CATS)	R3 (CATS)
PATTERN 0	CATERPILLA	19	33	15
PATTERN 1	CRAYFISH	19	33	15
PATTERN 2	FLIES	19	33	15
PATTERN 3	GRASSHOPPER	19	33	15
PATTERN 4	MOSQUITO	19	33	15
PATTERN 5	SCORPION	19	33	15
PATTERN 6	SPIDER	19	33	15
PATTERN 7	FLEA	19	33	15
PATTERN 8	JELLYFISH	9	20	7
PATTERN 9	SALMON	9	20	7
PATTERN 10	SARDINE	9	20	7
PATTERN 11	SAWFISH	9	20	7
PATTERN 12	WHALE	9	20	7
PATTERN 13	EEL	9	20	7
PATTERN 14	TILAPIA	9	20	7
PATTERN 15	ADDER	14	29	1
PATTERN 16	ALLIGATOR	14	29	1
PATTERN 17	ANACONDA	14	29	1
PATTERN 18	RAY	9	20	7
PATTERN 19	ANT	19	33	15
PATTERN 20	CHIMAERA	9	20	7
PATTERN 21	DOGFISH	9	20	7
PATTERN 22	CATFISH	9	20	7
PATTERN 23	MACKEREL	9	20	7
PATTERN 24	TUNA	9	20	7
PATTERN 25	PYTHON	14	29	1
PATTERN 26	CHAMELEON	14	29	1
PATTERN 27	TORTOISE	14	29	1
PATTERN 28	TURTLE	9	29	11
PATTERN 29	LIZARD	14	29	1
PATTERN 30	GECKO	14	29	1
PATTERN 31	ASP	14	29	1
PATTERN 32	SHARK	9	20	7
PATTERN 33	TOAD	13	19	11
PATTERN 34	SALAMANDA	13	19	11
PATTERN 35	CAECILIAN	14	29	1
PATTERN 36	NEWT	13	19	11
PATTERN 37	MUDPUPPY	13	19	11
PATTERN 38	AMPHIUMA	13	19	11
PATTERN 39	BEE	19	33	15
PATTERN 40	BEEBLE	19	33	15
PATTERN 41	OSTRICH	3	21	0
PATTERN 42	COW	18	26	10
PATTERN 43	PENGUIN	3	21	0
PATTERN 44	BUSTARD	3	21	0
PATTERN 45	PIGEON	3	21	0
PATTERN 46	COCKROACH	19	33	15
PATTERN 47	VULTURE	3	21	0
PATTERN 48	SPARROW	3	21	0
PATTERN 49	COBRA	14	29	1
PATTERN 50	COCRODILE	14	29	1
PATTERN 51	PEACOCK	3	21	0
PATTERN 52	DOG	18	26	10

PATTERNS	ANIMAL SAMPLES	R1 (CATS)	R2 (CATS)	R3 (CATS)
PATTERN 53	DOLPHIN	9	20	7
PATTERN 54	ZEBRA	18	26	10
PATTERN 55	RABBIT	18	26	10
PATTERN 56	HUMAN	18	26	10
PATTERN 57	DUCK	3	21	0
PATTERN 58	ELEPHANT	18	26	10
PATTERN 59	EAGLE	3	21	0
PATTERN 60	FROG	13	19	11
PATTERN 61	PIG	18	26	10
PATTERN 62	GORILLA	18	26	10
PATTERN 63	GOOSE	3	23	0
PATTERN 64	HAWK	3	21	0
PATTERN 65	HORSE	18	26	10
PATTERN 66	LION	18	26	10
PATTERN 67	OWL	3	21	0

Output of the Self-organizing map algorithm

Tables 2, 3 and 4 show the detailed output of the SOM algorithm. R1 (CATS) represents the output of the first run of the SOM algorithm with their pattern numbers after 172 epochs. R2 (CATS) represents the outputs, with their pattern

number after 172 epochs and so on.

The tables below classify the patterns into their clusters results or output. The three tables represent the output for the first, second, and third run of the SOM algorithm respectively.

**Table 2.** Result of the classification after the first process by the SOM algorithm

CAT 19	CATERPILLA, FLIES, GRASSHOPPER, MOSQUITO, SCORPION, SPIDER, FLEA, ANT, BEE, BEETLE, COCKROACH, CRAYFISH
CAT 9	JELLYFISH, EEL, TILAPIA, RAY, CHIMAERA, DOGFISH, CATFISH, MACKEREL, TUNA, TURTLE, SHARK, SALMON, SARDINE, SAWFISH, WHALE, DOLPHIN
CAT 13	MUDPUPPY, AMPHIUMA, FROG, TOAD, SALAMANDA, NEWTS
CAT 14	ADDER, ALLIGATOR, ANACONDA, PYTHON, CHAMELEON, TORTOISE, LIZARD, GECKO, ASP, CAECILIAN, COBRA, COCRODILE
CAT 3	OSTRICH, PENGUIN, BUSTARD, PIGEON, VULTURE, SPARROW, PEACOCK, , DUCK, HAWK, OWL, EAGLE, GOOSE
CAT 18	COW, DOG, ZEBRA, RABBIT, HUMAN, ELEPHANT, PIG, GORILLA
	HORSE, LION

FIRST RESULT/OUTPUT.

**Table 3.** Result of the second classification after the second process by the SOM algorithm.

CAT 33	CATERPILLA, FLIES, GRASSHOPPER, MOSQUITO, SCORPION, SPIDER, FLEA, ANT, BEE, BEETLE, COCKROACH, CRAYFISH
CAT 20	JELLYFISH, SALMON, SARDINE, SAWFISH, WHALE, EEL, TILAPIA, RAY, CHIMAERA, DOGFISH, CATFISH, MACKEREL, TUNA, SHARK, DOLPHIN,
CAT 29	ADDER, ALLIGATOR, ANACONDA, PYTHON, CHAMELEON, TORTOISE, TURTLE, LIZARD, GECKO, ASP, CAECILIAN, COBRA, COCRODILE
CAT 21	OSTRICH, PENGUIN, BUSTARD, PIGEON, VULTURE, SPARROW, PEACOCK, DUCK, HAWK, OWL, EAGLE
CAT 26	COW, DOG, ZEBRA, RABBIT, HUMAN, ELEPHANT, PIG, GORILLA, HORSE, LION
CAT 19	FROG, TOAD, MUDPUPPY, SALAMANDA, NEWTS, AMPHIUMA

SECOND RESULT.

**Table 4.** Outcome of the third classification after the third process by the SOM algorithm.

CAT 15	CATERPILLA, FLIES, GRASSHOPPER, MOSQUITO, SCORPION, SPIDER, FLEA, ANT, BEE, BEETLE, COCKROACH, CRAYFISH.
CAT 7	JELLYFISH, SALMON, SARDINE, SAWFISH, WHALE, EEL, TILAPIA, RAY, CHIMAERA, DOGFISH, CATFISH, MACKEREL, TUNA, SHARK, DOLPHIN,
CAT 1	ADDER, ANACONDA, PYTHON, CHAMELEON, TORTOISE, LIZARD, ALLIGATOR, GECKO, ASP, CAECILIAN, COBRA, COCRODILE
CAT 11	TURTLE, MUDPUPPY, AMPHIUMA, FROG, TOAD, SALAMANDA, NEWTS
CAT 0	OSTRICH, PENGUIN, BUSTARD, PIGEON, VULTURE, SPARROW, PEACOCK, DUCK, GOOSE, HAWK, OWL, EAGLE
CAT 10	COW, DOG, ZEBRA, RABBIT, HUMAN, ELEPHANT, PIG, GORILLA, HORSE, LION

THIRD RESULT.

### 5.1. Interpretation of the Result

We compare the result of the SOM with the appropriate classification in order to determine the accuracy of the SOM's

**Table 5.** Shows the appropriate classification of the organisms. The SOM results are compared with this table to determine the percentage of accuracy.

**Table 5.** Shows the appropriate classification of the organisms. The SOM results are compared with this table to determine the percentage of accuracy.

INVERTEBRATE	FISH	AMPHIBIANS	REPTILES	BIRDS	MAMMALS
	Dolphin				
Bee	Jellyfish		Python	Ostrich	
Beetle	Salmon		Asp	Penguin	Cow
Caterpillar	Sardine		Adder	Eagle	Dog
Ant	Sawfish	Frog	Anaconda	Bustard	Elephant
Cockroach	Shark	Toad			Gorilla
Crayfish	Eel	Salamanda	Chameleon	Pigeon	Zebra
Grasshopper	Tilapia	Newts	Cobra	Duck	Horse
Flies	Ray	Caecilian	Crocodiles	Goose	Whale
Mosquito	Chimaeras	Mudpuppy	Alligator	Hawk	Rabbit
Scorpion	Dogfish	Amphiuma	Tortoise	Vulture	Human
Spider	Catfish		Turtle	Sparrow	Pig
Flea	Mackerel		Lizard	Owl	Lion
	Tuna		Gecko	Peacock	

Percentage error=100% - 97.06%=2.94%

$$\text{Average percentage accuracy} = (97.06 + 95.588 + 97.06) / 3 = 96.569\%$$

$$\text{Percentage accuracy} = \frac{66}{68} \times 100\% = 97.06\%$$

Percentage error=100% - 97.06%=2.94%

Average percentage error=(2.94 + 4.412 + 2.94)/3=3.431%

### 5.2. Limitation of the Research Work

The output of the SOM algorithm varies slightly. This is because the input samples have closely related attributes or characteristics.

$$\text{Percentage accuracy} = \frac{65}{68} \times 100\% = 95.588\%$$

Percentage error=100% - 95.588%=4.412%

For accurate clustering to occur, the number of iterations must be considerably high, about 1000 iterations, especially when the input patterns have very similar attributes.

We only select 68 input samples. There are millions of living organisms on the earth. Super computer will be required to clusters these kinds of samples.

We only select 68 input samples. There are millions of living organisms on the earth. Super computer will be required to clusters these kinds of samples.

$$\text{Percentage accuracy} = \frac{66}{68} \times 100\% = 97.06\%$$

**Table 6.** Shows the outcome of the feature extraction of the 68 animals processed by the SOM.

[illegible]

	Bee	Beetle	Caterpilla	Ant	Cockroach	Crayfish	Flies	Grasshoper	Mosquito	Scorppion	Spider	Flea
Feather	0	0	0	0	0	0	0	0	0	0	0	0
Beak/ Bill	0	0	0	0	0	0	0	0	0	0	0	0
Furcula	0	0	0	0	0	0	0	0	0	0	0	0
Tracheae	1	1	1	1	1	1	1	1	1	1	1	1
Insectivores	0	0	0	0	0	0	0	0	0	1	1	0
Webs	0	0	0	0	0	0	0	0	0	0	0	0
Antenna	1	1	1	1	1	1	1	1	1	1	1	1
Wings	1	1	0	0	1	0	1	1	1	0	0	0
Exoskeleton	1	1	1	1	1	1	1	1	1	0	1	1
Proboscis	1	1	1	1	1	0	1	1	1	0	0	1
Crops	1	1	1	1	1	1	1	1	1	1	1	1
Tails	0	0	0	0	0	0	0	0	0	1	0	0
Jacobson's org	0	0	0	0	0	0	0	0	0	0	0	0
Cocoon	1	1	1	1	1	0	1	1	0	0	0	1
Fins	0	0	0	0	0	0	0	0	0	0	0	0
Swim bladder	0	0	0	0	0	0	0	0	0	0	0	0
thorax	1	1	1	1	1	1	1	1	1	1	1	1
metamorphosis	1	1	1	1	1	1	1	1	1	1	1	1
Cloaca	0	0	0	0	0	0	0	0	0	0	0	0

	Dolphin	Jellyfish	Asp	Salmon	Sardine	Sawfihs	Horse	Shark	Whale	Eel	Tilapia	Ray
Vertebrates	1	1	1	1	1	1	1	1	1	1	1	1
Invertebrates	0	0	0	0	0	0	0	0	0	0	0	0
Ectothermy	0	1	1	0	0	0	0	1	0	1	1	1
Endothermy	1	0	0	1	1	1	1	0	1	0	0	0
Terrestrial	0	0	1	0	0	0	1	0	0	0	0	0
Aquatic	1	1	0	1	1	1	0	1	1	1	1	1
Arboreal	0	0	0	0	0	0	0	0	0	0	0	0
Gills	0	1	0	0	1	1	0	1	0	1	1	1
Lungs	1	0	1	1	0	0	1	0	1	0	0	0
Tetrapods	0	0	1	0	0	0	1	0	0	0	0	0
Scales	0	0	1	1	1	1	0	1	0	0	1	1
Amniotes	0	0	0	1	0	0	0	0	0	0	0	0
Oviparous	0	0	0	1	1	0	0	1	0	1	1	1
Viviparous	1	0	1	0	0	1	1	1	1	0	0	0
Nocturnal	0	0	0	0	0	0	0	0	0	0	0	0
Carnivores	1	0	1	1	1	1	0	1	1	0	1	1
Herbivores	0	0	0	0	0	0	1	0	0	0	0	0
Omnivores	0	0	0	0	0	0	0	0	0	0	0	0
Hair	0	1	0	0	0	0	1	0	0	0	0	0
Mammaygland	1	0	0	0	0	0	1	0	1	0	0	0
Diphyodonty	0	0	0	0	0	0	1	0	1	0	0	0
Polyphydnty	1	0	0	1	1	1	0	1	0	1	1	1
Diaphragm	1	0	0	0	0	0	1	0	1	0	0	0
4 Heart Cham	1	0	0	0	0	0	1	0	1	0	0	0
3 Heart Cham	0	0	0	0	0	0	0	0	0	0	0	0
2 Heart Cham	0	0	1	1	1	1	0	1	0	1	1	0
Lower jaw	1	0	1	1	1	1	1	1	1	1	1	1
Feather	0	0	0	0	0	0	0	0	0	0	0	0
Beak/ Bill	1	0	0	0	0	1	0	0	0	0	0	1
Furcula	0	0	0	0	0	0	0	0	0	0	0	0
Tracheae	0	0	0	0	0	0	0	0	0	0	0	0
Insectivores	0	0	0	0	0	0	0	0	0	1	1	0
Webs	0	0	0	0	0	0	0	0	0	0	0	0
Antenna	0	1	0	0	0	0	0	0	0	0	0	0
Wings	0	0	0	0	0	0	0	0	0	0	0	0
Exoskeleton	0	1	0	0	0	0	0	0	0	0	0	0
Proboscis	0	0	0	0	0	0	0	0	0	0	0	0
Crops	0	0	0	0	0	0	0	0	0	0	0	0
Tails	1	0	1	1	1	1	1	1	1	1	1	1
Jacobson's org	0	0	0	0	0	0	0	0	0	0	0	0
Cocoon	0	0	0	0	0	0	0	1	0	0	0	0
Fins	1	1	0	1	1	1	0	1	0	1	1	1
Swim bladder	1	0	0	1	1	0	0	1	1	1	1	1
thorax	0	0	0	0	0	0	0	0	0	0	0	0
metamorphosis	0	0	0	0	0	0	0	0	0	1	0	0
Cloaca	1	1	1	1	1	1	1	1	0	1	1	1







	Amphiuma	Ostrich	Penguin	Eagle	Bustard	Pigeon	Duck	Goose	Hawk	Vulture	Sparrow	Owl
Tails	1	1	1	1	1	1	1	1	1	1	1	1
Jacobson's org	1	0	0	0	0	0	0	0	0	0	0	0
Cocoon	0	0	0	0	0	0	0	0	0	0	0	0
Fins	0	0	0	0	0	0	0	0	0	0	0	0
Swim bladder	0	0	1	0	0	0	0	0	0	0	0	0
thorax	0	0	0	0	0	0	0	0	0	0	0	0
metamorphosis	1	0	0	0	0	0	0	0	0	0	0	0
Cloaca	1	1	1	1		1	1	1	1	1	1	1

	Peacock	Cow	Dog	Elephant	Gorilla	Zebra	Rabbit	Human	Pig	Lion
Vertebrates	1	1	1	1	1	1	1	1	1	1
Invertebrates	0	0	0	0	0	0	0	0	0	0
Ectothermy	0	0	0	0	0	0	0	0	0	0
Endothermy	1	1	1	1	1	1	1	1	1	1
Terrestrial	1	1	1	1	1	1	1	1	1	1
Aquatic	0	0	0	0	0	0	0	0	0	0
Arboreal	0	0	0	0	1	0	0	0	0	0
Gills	0	0	0	0	0	0	0	0	0	0
Lungs	1	1	1	1	1	1	1	1	1	1
Tetrapods	1	1	1	1	1	1	1	1	1	1
Scales	0	0	0	0	0	0	0	0	0	0
Amniotes	1	0	0	0	0	0	0	0	0	0
Oviparous	1	0	0	0	0	0	0	0	0	0
Viviparous	0	1	1	1	1	1	1	1	1	1
Nocturnal	0	0	0	0	0	0	0	0	0	0
Carnivores	1	0	0	0	0	0	0	0	0	1
Herbivores	1	1	0	1	1	1	1	0	0	0
Omnivores	1	0	1	0	0	0	0	1	1	0
Hair	0	1	1	1	1	1	1	1	1	1
Mammary gland	0	1	1	1	1	1	1	1	1	1
Diphyodonty	0	1	1	1	1	1	1	1	1	1
Polyphyodonty	1	0	0	0	0	0	0	0	0	0
Diaphragm	0	1	1	1	1	1	1	1	1	1
4 Heart Cham	1	1	1	1	1	1	1	1	1	1
3 Heart Cham	0	0	0	0	0	0	0	0	0	0
2 Heart Cham	0	0	0	0	0	0	0	0	0	0
Lower jaw	1	1	1	1	1	1	1	1	1	1
Feather	1	0	0	0	0	0	0	0	0	0
Beak/ Bill	1	0	0	0	0	0	0	0	0	0
Furcula	1	0	0	0	0	0	0	0	0	0
Tracheae	0	0	0	0	0	0	0	0	0	0
Insectivores	1	0	0	0	0	0	0	0	0	0
Webs	0	0	0	0	0	0	0	0	0	0
Antenna	0	0	0	0	0	0	0	0	0	0
Wings	1	0	0	0	0	0	0	0	0	0
Exoskeleton	0	0	0	0	0	0	0	0	0	0
Proboscis	1	0	0	0	0	0	0	0	0	0
Crops	1	0	0	0	0	0	0	0	0	0
Tails	1	1	1	1	0	1	1	0	1	1
Jacobson's org	0	0	0	0	0	0	0	0	0	0
Cocoon	0	0	0	0	0	0	0	0	0	0
Fins	0	0	0	0	0	0	0	0	0	0
Swim bladder	0	0	0	0	0	0	0	0	0	0
thorax	0	0	0	0	0	0	0	0	0	0
metamorphosis	0	0	0	0	0	0	0	0	0	0
Cloaca	1	0	0	0	0	0	0	0	0	0

## 6. Conclusion

With the problems of classification of living organisms, SOM algorithm can actually be applied to classify living organisms though more work needs to be done in this area. We used the SOM algorithm in this research to classifying organisms only in the animal kingdom. We selected 68

animals from different Phyla with 46 attributes. The SOM algorithm gave an average of 96.596% accuracy. This result can be improved on by increasing the number of iterations (about 1000) and the number of attributes.

## Acknowledgements

I want to give thanks to God for the grace and enablement

to complete this research work despite many delay and constraints encountered. I want to appreciate Dr. Oladipupo Sennaike, University of Lagos, for his help and support. Lastly I want to appreciate my wife for his help and concern.

## References

- [1] Saed K, Usama P, Mohammed E, Alaa-Eddin A, Nayel A. Taxonomic Classification for Living Organisms Using Convolutional Neural Networks.
- [2] Fernando, S, Perera, S. Empirical Analysis of Data Mining Techniques for Social Network Websites. Compusoft, 2014, 3, 582.
- [3] Science daily, <https://www.sciencedaily.com/releases/2011/08/110823180459.htm>.
- [4] Michael A, Dennis P, Thomas M, Nicolas, Thierry B, Richard C, Thomas C, Michael D, Paul M. a higher level of classification of all living organisms. Integrated Taxonomic Information System.
- [5] <https://www.online-science.com/biology/modern-classification-of-living-organisms-kingdom-monera-and-protista>.
- [6] Wikipedia, Robert Whittaker (ecology), [https://en.wikipedia.org/wiki/Robert\\_Whittaker\\_\(ecologist\)](https://en.wikipedia.org/wiki/Robert_Whittaker_(ecologist)).
- [7] Wikibook, general biology/classification of living things/classification and domains of life. [https://en.wikibooks.org/wiki/General\\_Biology/Classification\\_of\\_Living\\_Things/Classification\\_and\\_Domains\\_of\\_Life](https://en.wikibooks.org/wiki/General_Biology/Classification_of_Living_Things/Classification_and_Domains_of_Life).
- [8] Huston M. Biological Diversity: The Coexistence of Species; Cambridge University Press: Cambridge, UK, 1994.
- [9] Dubravko M, Hrvatska E, Zagreb C. Brief Review of Self-Organizing Maps.
- [10] Marie C, Madalina O, Fabrice R, Nathalie V (2016). Theoretical and applied aspect of Self-organizing map. <https://hal.archives-ouvertes.fr/hal-01270701>.
- [11] Kohonen T (1990). The self-organizing map, Vol 78, issue 9.
- [12] Kohonen T (2012). Essentials of Self-organizing map, National Library of Medicine. doi: 10.1016/j.neunet.2012.09.018.
- [13] Umut A, Secil E (2012). An introduction to Self-organizing map: <https://www.researchgate.net/publication/263084866>.
- [14] Merja O, Samuel K, Kohonen T (2003), Bibliography of Self-Organizing Map (SOM) Papers: 1998-2001 Addendum, Helsinki University of Technology, Neural Networks Research Centre, P. O. Box 5400, FIN-02015 HUT, FINLAND.
- [15] Pollock R, Toby L, Michael W (2002). A Kohonen Self-Organizing Map for the functional classification of proteins based on one-dimensional sequence information, conference paper 2002. DOI: 10.1109/IJCNN.2002.1005467.
- [16] Lars B, Bjorn G (2016). Self-Organizing Maps for Classification of a Multi-Labeled Corpu. Department of Computer and Information Science Norwegian University of Science and Technology.
- [17] Herry-Derajad W, Saruni D (2019), clustering of earth quake data using Kohonen Self-organizing (SOM) algoeithm, DOI:10.21276/sb.2019.5.7.11.
- [18] Faigl J (2016). An Application of Self-Organizing Map for Multirobot Multigoal Path Planning with Minmax Objective, vol Volume 2016, Article ID 2720630, 15 pages <http://dx.doi.org/10.1155/2016/2720630>.
- [19] Marzieh M, Mahdi and Abdol Rassoul Z (2018). Using Self-Organizing Maps for Determination of Soil Fertility (Case Study: Shiraz Plain. Soil & Water Res., 13, 2018 (1): 11–17 doi: 10.17221/139/2016-SWR.
- [20] Subana S, Sallis P, Buckeridge J. Self-organizing map for integrating data across multiple scales, e Auckland University of Technology, New Zealand, [www.aut.ac.nz](http://www.aut.ac.nz).
- [21] Yuan-Chao L, Ming L and Xiao-Long W. Application of Self-Organizing maps in text clustering: A review. Doi: 10.5772/50618.