International Journal of Academic Engineering Research (IJAER)
ISSN: 2643-9085
Vol. 4 Issue 10, October - 2020, Pages: 23-31

Classification of Animal Species Using Neural Network

Rand Suhail Abu Al-Araj, Shaima Khalil Abed, Ahmed Nabil Al-Ghoul, Samy S. Abu-Naser

Department of Information Technology,
Faculty of Engineering and Information Technology,
Al-Azhar University, Gaza, Palestine

***Abstract**: Over 1.5 million living animal species have been described—of which around 1 million are insects—but it has been estimated there are over 7 million animal species in total. Animals range in length from 8.5 micrometres to 33.6 metres. In this paper an Artificial Neural Network (ANN) model, was developed and tested to predict animal species. There are a number of features that influence the classification of animal species. Such as the existence of hair/ feather, if the animal gives birth or spawns, it is airborne, aquatic, predator, toothed, backboned, venomous, has –fins, has-tail, cat-sized, and domestic. A model based on the Multilayer Perceptron Topology was proposed and trained, using data set what was collected from UCI Machine Learning Repository. Evaluation of the proposed model shows that the ANN model is able to correctly predict the animal category with 100% accuracy.*

Keywords: Artificial Neural Networks, animal, classification, JNN.

Introduction

Animals are multicellular eukaryotic organisms that form the biological kingdom Animalia. With few exceptions, animals consume organic material, breathe oxygen, are able to move, can reproduce , and grow from a hollow sphere of cells, the blastula, during embryonic development. Over 1.5 million living animal species have been described—of which around 1 million are insects—but it has been estimated there are over 7 million animal species in total. Animals range in length from 8.5 micrometres to 33.6 metres. They have complex interactions with each other and their environments, forming intricate food webs. The kingdom Animalia includes humans but in colloquial use the term *animal* often refers only to non-human animals. The scientific study of animals is known as zoology[1].

Most living animal species are in Bilateria, a clade whose members have a bilaterally symmetric body plan. The Bilateria include the protostomes—in which many groups of invertebrates are found, such as nematodes, arthropods, and molluscs—and the deuterostomes, containing both the echinoderms as well as the chordates, the latter containing the vertebrates. Life forms interpreted as early animals were present in the Ediacaran biota of the late Precambrian. Many modern animal phyla became clearly established in the fossil record as marine species during the Cambrian explosion, which began around 542 million years ago. 6,331 groups of genes common to all living animals have been identified; these may have arisen from a single common ancestor that lived 650 million years ago.[2]

Historically, Aristotle divided animals into those with blood and those without. Carl Linnaeus created the first hierarchical biological classification for animals in 1758 with his *Systema Naturae*, which Jean-Baptiste Lamarck expanded into 14 phyla by 1809. In 1874, Ernst Haeckel divided the animal kingdom into the multicellular Metazoa (now synonymous for Animalia) and the Protozoa, single-celled organisms no longer considered animals. In modern times, the biological classification of animals relies on advanced techniques, such as molecular phylogenetics, which are effective at demonstrating the evolutionary relationships between animal taxa[3].

Humans make use of many other animal species, such as for food (including meat, milk, and eggs), for materials (such as leather and wool), and also as pets, and for transports, as working animals. Dogs have been used in hunting, while many terrestrial and aquatic animals were hunted for sports. Non-human animals have appeared in art from the earliest times and are featured in mythology and religion [4].

The purpose of this research is to study the feasibility of classification animal species using neural networks. An animal class is made up of animal that are all alike in important ways. So we need to train a neural network to make it able to predict which species belong to a particular animal.

An Artificial Neural Network (ANN) is a mathematical model that is driven by the functional feature of biological neural networks. A neural network contains an interconnected set of artificial neurons, and it processes information using a connectionist form to computation. As a rule, an ANN is an adaptive system that adjusts its structure based on external or internal data that runs over the network during the learning process. Current neural networks are non-linear numerical data modeling tools. They are usually used to model tricky relationships among inputs and outputs or to uncover patterns in data. ANN has been applied in several applications with significant accomplishment [5]. For example, ANN has been effectively applied in the area of prediction, and handwritten character recognition [6].

Neurons are often come together into layers. Layers are groups of neurons that perform similar functions. There are three kinds of layers. The input layer is the layer of neurons that take input from the user program. The layer of neurons that send data to the user program is the output layer. Between the input layer and output layer there are hidden layers. Hidden layer neurons are connected only to other neurons and never directly interact with the user program. The input and output layers are not just there as interface points. Every neuron in a neural network has the opportunity to affect processing. Processing can occur at any layer in the neural network. Not every neural network has this many layers. The hidden layer is optional. The input and output layers are required, but it is possible to have a layer that act as both an input and output layer [7-8].

ANN learning can be either supervised or unsupervised. Supervised training is accomplished by giving the neural network a set of sample data along with the expected outputs from each of these samples. Supervised training is the most common form of neural network training. As supervised training proceeds, the neural network is taken through several iterations, or epochs, until the actual output of the neural network matches the expected output, with a reasonably small error. Each epoch is one pass through the training samples. Unsupervised training is similar to supervised training except that no expected outputs are provided. Unsupervised training usually occurs when the neural network is to classify the inputs into several groups[9-11].

The training progresses through many epochs, just as in supervised training. As training progresses, the classification groups are "discovered" by the neural network [12-15]. Training is the process by which these connection weights are assigned. Most training algorithms begin by assigning random numbers to the weight matrix. Then the validity of the neural network is tested. Next, the weights are adjusted based on validation results. This process is repeated until the validation error is within an acceptable limit [41-50]. Validation of the system is done once a neural network has been trained and it must be evaluated to see if it is ready for actual use. This final step is important so that it can be determined if additional training is required. To properly validate a neural network validation data must be set aside that is completely separate from the training data [16-22].

4. METHODOLOGY

A data set refer to Richard Forsyth [23] was used; it contains a number of factors that are considered to have an effect on the classification of an animal. These factors were classified as input variables. The output variable represents the predicted animal classification based on those inputs.

4.1 The Input Variables

This database includes 101 cases. Each case is the name of animal. It was found that each of these animals belonged to one of seven classes.

Table 1: name of Animal and its class that it belongs to

Class#	Count	Set of animals
1	41	aardvark, antelope, bear, boar, buffalo, calf, cavy, cheetah, deer, dolphin, elephant, fruitbat, giraffe, girl, goat, gorilla, hamster, hare, leopard, lion, lynx, mink, mole, mongoose, opossum, oryx, platypus, polecat, pony, porpoise, puma, pussycat, raccoon, reindeer, seal, sealion, squirrel, vampire, vole, wallaby, wolf
2	20	chicken, crow, dove, duck, flamingo, gull, hawk, kiwi, lark, ostrich, parakeet, penguin, pheasant, rhea, skimmer, skua, sparrow, swan, vulture, wren
3	5	pitviper, seasnake, slowworm, tortoise, tuatara
4	13	bass, carp, catfish, chub, dogfish, haddock, herring, pike, piranha, seahorse, sole, stingray, tuna
5	4	frog, frog, newt, toad
6	8	flea, gnat, honeybee, housefly, ladybird, moth, termite, wasp
7	10	clam, crab, crayfish, lobster, octopus, scorpion, seawasp, slug, starfish, worm

This variable, named *type*, represents the output variable. Except the output variable, there were 17 input variables for each animal species. Information of input variables:

Table2: Input attributes and its types

#	Hair	Boolean;

1	Feathers	Boolean;
2	Eggs	Boolean;
3	Milk	Boolean;
4	Airborne	Boolean;
5	Aquatic	Boolean ;
6	Predator	Boolean;
7	Toothed	Boolean;
8	Backbone	Boolean;
9	Breathes	Boolean;
10	Venomous	Boolean;
11	Fins	Boolean;
12	Legs	Numeric (set of values {0, 2, 4, 5, 6, 8});
13	Tail	Boolean;
14	Domestic	Boolean;
15	Catsize	Boolean;

Each variable is type of Boolean, except variable animal name which is nominal variable and variable legs is a numeric variable (set of values: {0, 2, 4, 6, 8}).

Handling non-numeric data, such as Boolean = {true, false}, is more difficult. However, nominal-valued variables can be represented numerically. Value **true** will be replaced with value 1, and value **false** will be replaced with value 0. We wil not use variable animal name in experiment, because this variable is unique for each case.

Once the most appropriate raw input data has been selected, it must be preprocessed; otherwise, the neural network will not produce accurate forecasts.

Transformation and normalization are two widely used preprocessing methods. Transformation involves manipulating raw data inputs to create a single input to a network, while normalization is a transformation performed on a single data input to distribute the data evenly and scale it into an acceptable range for the network. Acceptable rang of value, in *JNN tool*, varies in the range between zero and one. In this study we will use normalization, as preprocessing method, to normalized variable legs, because this variable have set of values 0, 2, 4, 6, 8. Set of values of other input variables are zero or one, so we will not apply this method over them. However, we will have to normalize the values of output variable type, because this variable we have set of values between 1 and 7.

In order to train a neural network to predict the class of animal species using this data set, there are procedure that has to be followed.

4.2 Data Normalization

Linear scaling of data is one of the methods of data normalization. Linear scaling requires that a minimum and maximum values associated with the facts for a single data input be found. Let's call these values *Xmin* and *Xmax*, respectively. The formula for transforming each data value to an input value *X* is:

$$X_i = (X_i - X_{min}) / (X_{man} - X_{min})$$

Our desired range varies in the interval between zero and one. The original data set indicate that the variable legs is associated with a set of values {0, 2, 4, 6, 8}, so the constant value of X_{min} is 0 and the constant value of X_{max} is 8. This method of normalization will scale input data into the appropriate range 0 to 1.

Variable *type* is the output value and it represents the seven different classes of animals. The output were normalized in the same manner between 0 and 1.

4.4 Building the ANN Model

We have used Just Neural Network (JNN) tool [24] to build a multilayer ANN model. The proposed model consists of 5 Layers: Input Layer with 17 nodes, First Hidden Layer with 3 nodes, Second Hidden Layer with 1 node, the third Hidden Layer with 4 nodes, and Output Layer with one node as can be seen in Figure 1.

We have sat the parameters of the proposed model as follows: Learning Rate 0.37 and the Momentum to be 0.46, and Average Error rate to be 0.01 (as shown in Figure 2).

4.5 Evaluating the ANN model

The Animal Species dataset consists of 101 samples with 17 attributes as in Table 1 and Table 2. We imported the CSV file of the Animal Species dataset into the JNN environment (as seen in Figure 3). We divided the imported dataset into two groups (Training and Validation) randomly using the JNN tool. The Training consists of approximately 67% (68 samples) and the validation set consists of 33% of the dataset (33 samples). After making sure that the parameter control was sat properly, we started training the ANN model and kept eye on the learning curve, loss error and validation accuracy. We trained the ANN model for 801 cycles. The best accuracy we got was 100% (as seen in Figure 4). We determined the most influential factors in the animal species dataset as in Figure 5. Figure 6 shows the summary of the proposed model.

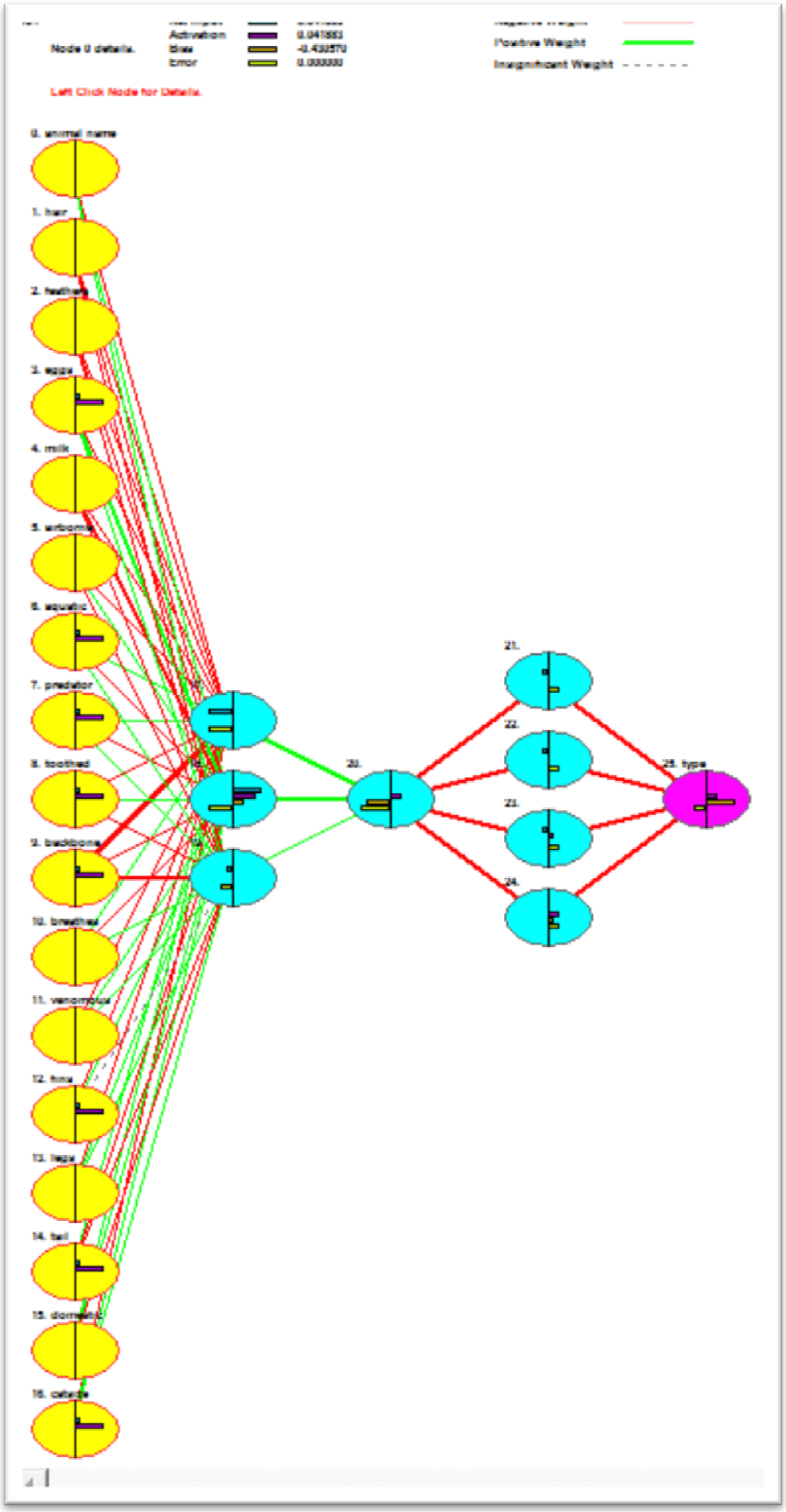


Figure 1: Proposed model architecture.

International Journal of Academic Engineering Research (IJAER)
ISSN: 2643-9085
Vol. 4 Issue 10, October - 2020, Pages: 23-31

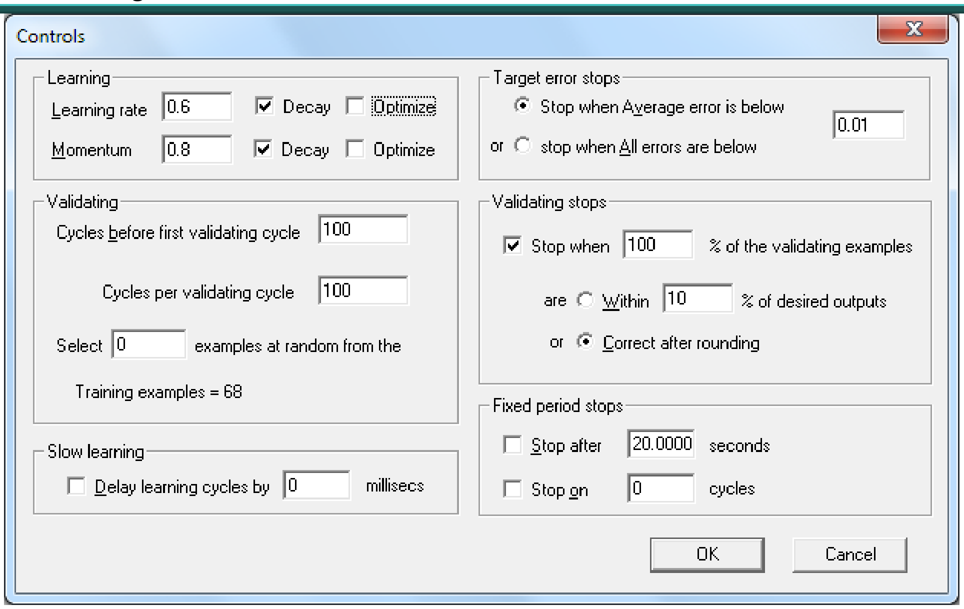
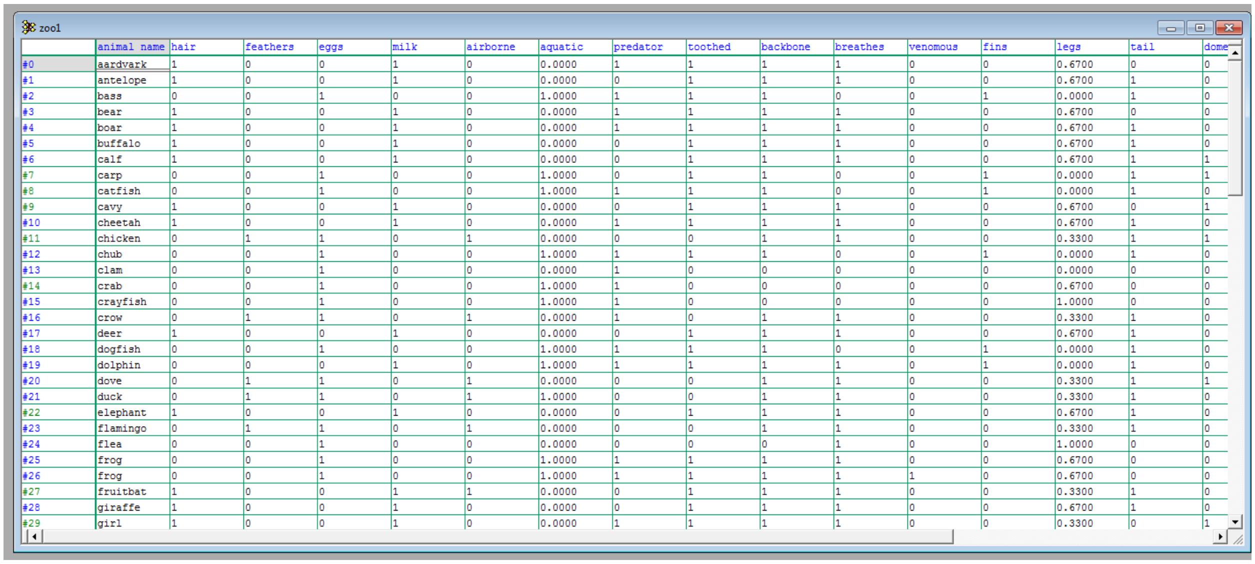


Figure 2: Setting Parameters of the proposed model



zoo1

	animal name	hair	feathers	eggs	milk	airborne	aquatic	predator	toothed	backbone	breathes	venomous	fins	legs	tail	dome
#0	aardvark	1	0	0	1	0	0.0000	1	1	1	1	0	0	0.6700	0	0
#1	antelope	1	0	0	1	0	0.0000	0	1	1	1	0	0	0.6700	1	0
#2	bass	0	0	1	0	0	1.0000	1	1	1	0	0	1	0.0000	1	0
#3	bear	1	0	0	1	0	0.0000	1	1	1	1	0	0	0.6700	0	0
#4	boar	1	0	0	1	0	0.0000	1	1	1	1	0	0	0.6700	1	0
#5	buffalo	1	0	0	1	0	0.0000	0	1	1	1	0	0	0.6700	1	0
#6	calf	1	0	0	1	0	0.0000	0	1	1	1	0	0	0.6700	1	1
#7	carp	0	0	1	0	0	1.0000	0	1	1	0	0	1	0.0000	1	1
#8	catfish	0	0	1	0	0	1.0000	1	1	1	0	0	1	0.0000	1	0
#9	cavy	1	0	0	1	0	0.0000	0	1	1	1	0	0	0.6700	0	1
#10	cheetah	1	0	0	1	0	0.0000	1	1	1	1	0	0	0.6700	1	0
#11	chicken	0	1	1	0	1	0.0000	0	0	1	1	0	0	0.3300	1	1
#12	chub	0	0	1	0	0	1.0000	1	1	1	0	0	1	0.0000	1	0
#13	clam	0	0	1	0	0	0.0000	1	0	0	0	0	0	0.0000	0	0
#14	crab	0	0	1	0	0	1.0000	1	0	0	0	0	0	0.6700	0	0
#15	crayfish	0	0	1	0	0	1.0000	1	0	0	0	0	0	1.0000	0	0
#16	crow	0	1	1	0	1	0.0000	1	0	1	1	0	0	0.3300	1	0
#17	deer	1	0	0	1	0	0.0000	0	1	1	1	0	0	0.6700	1	0
#18	dogfish	0	0	1	0	0	1.0000	1	1	1	0	0	1	0.0000	1	0
#19	dolphin	0	0	0	1	0	1.0000	1	1	1	1	0	1	0.0000	1	0
#20	dove	0	1	1	0	1	0.0000	0	0	1	1	0	0	0.3300	1	1
#21	duck	0	1	1	0	1	1.0000	0	0	1	1	0	0	0.3300	1	0
#22	elephant	1	0	0	1	0	0.0000	0	1	1	1	0	0	0.6700	1	0
#23	flamingo	0	1	1	0	1	0.0000	0	0	1	1	0	0	0.3300	1	0
#24	flea	0	0	1	0	0	0.0000	0	0	0	1	0	0	1.0000	0	0
#25	frog	0	0	1	0	0	1.0000	1	1	1	1	0	0	0.6700	0	0
#26	frog	0	0	1	0	0	1.0000	1	1	1	1	1	0	0.6700	0	0
#27	fruitbat	1	0	0	1	1	0.0000	0	1	1	1	0	0	0.3300	1	0
#28	giraffe	1	0	0	1	0	0.0000	0	1	1	1	0	0	0.6700	1	0
#29	girl	1	0	0	1	0	0.0000	1	1	1	1	0	0	0.3300	0	1

Figure 3: Imported Normalized dataset to JNN environment

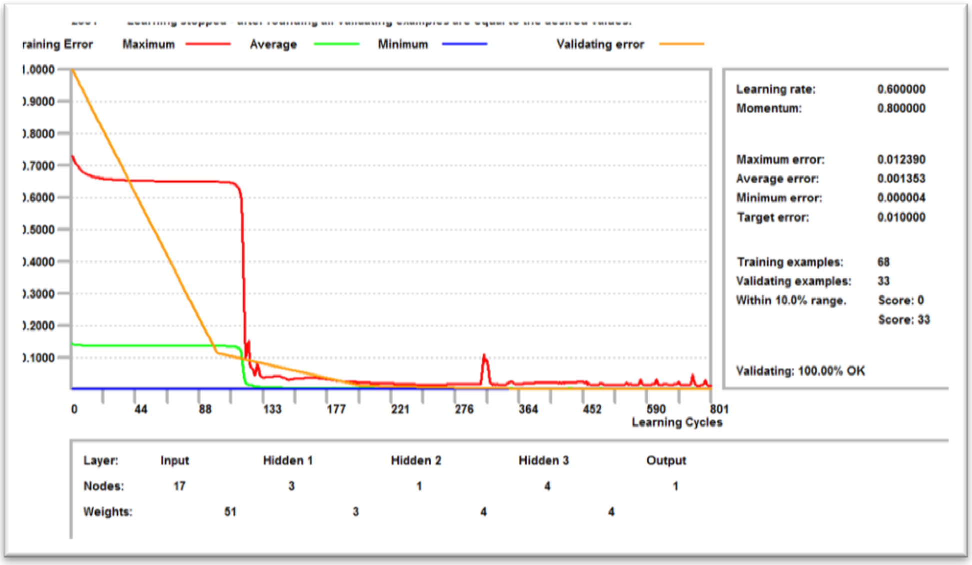
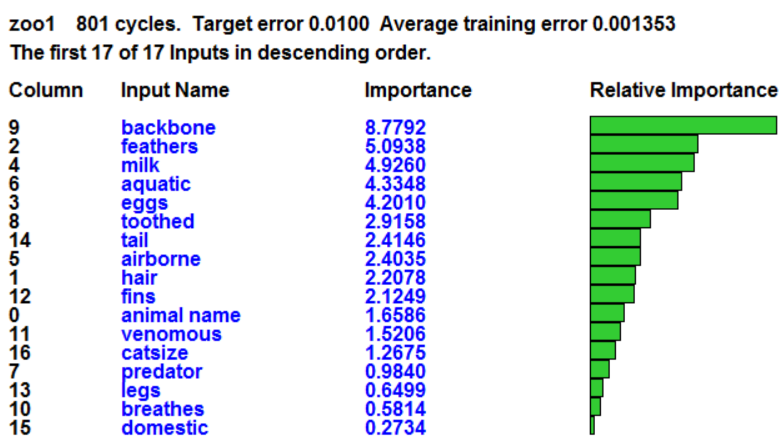


Figure 4: Training and validation of the proposed model.



zoo1 801 cycles. Target error 0.0100 Average training error 0.001353
The first 17 of 17 Inputs in descending order.

Column	Input Name	Importance	Relative Importance
9	backbone	8.7792	
2	feathers	5.0938	
4	milk	4.9260	
6	aquatic	4.3348	
3	eggs	4.2010	
8	toothed	2.9158	
14	tail	2.4146	
5	airborne	2.4035	
1	hair	2.2078	
12	fins	2.1249	
0	animal name	1.6586	
11	venomous	1.5206	
16	catsize	1.2675	
7	predator	0.9840	
13	legs	0.6499	
10	breathes	0.5814	
15	domestic	0.2734	

Figure 5: Most influential filed in the dataset

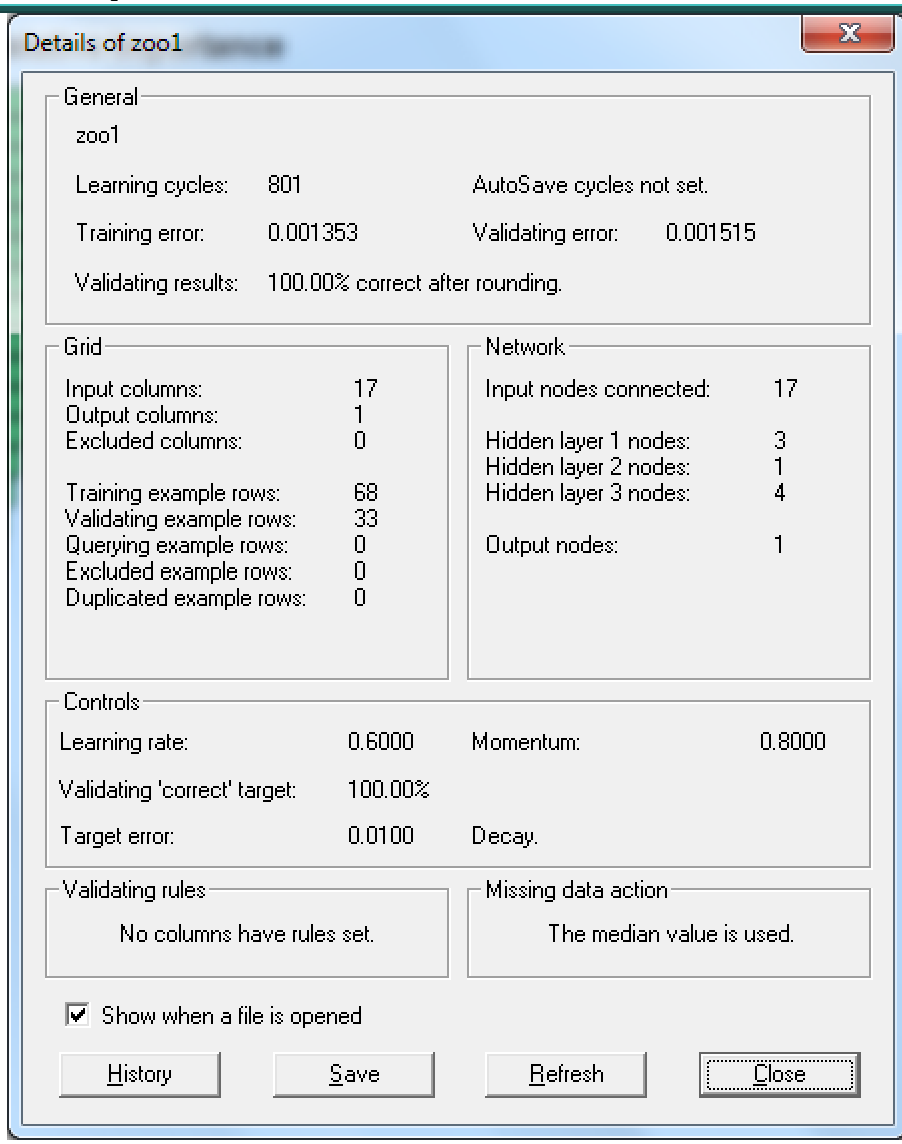


Figure 6: Detail of the proposed Model

CONCLUSION

An artificial Neural Network model for predicating animal species was presented. The model used feed forward backpropagation algorithm for training the proposed ANN model using JNN tool. The factors for the model were obtained from dataset which represents animal features of each animal species. The model was tested and the accuracy rate was 100%. This study showed that artificial neural network is capable of predicating animal species accurately.

References

1. Keith Laidler, “Animals: A Visual Guide to the Animal Kingdom”.
2. Richard Freeman, “Classification of the Animal Kingdom”.
3. Kathryn Whyman , “The Animal Kingdom: A Guide to Vertebrate Classification and Biodiversity” .
4. Sherman Hollar, “A Closer Look at the Animal Kingdom (Introduction to Biology)” .
5. Ting Qin, et al. "Continuous CMAC-QRLS and its systolic array." Neural Processing Letters 22.1 (2005): 1–16.
6. LeCun et al., (2019). “Backpropagation Applied to Handwritten Zip Code Recognition," Neural Computation, 1, pp. 541–551.
7. Yann LeCun (2016). Slides on Deep Learning Online
8. Hochreiter, Sepp; Schmidhuber, Jürgen (1997). "Long Short-Term Memory". Neural Computation. 9 (8): 1735–1780.
9. Sak, et al. (2014). "Long Short-Term Memory recurrent neural network architectures for large scale acoustic modeling" .

10. Li, Xiangang; Wu, Xihong (2014). "Constructing Long Short-Term Memory based Deep Recurrent Neural Networks for Large Vocabulary Speech Recognition".
11. Fan, Y.; Qian, Y.; Xie, F.; Soong, F. K. (2014). "TTS synthesis with bidirectional LSTM based Recurrent Neural Networks". Proceedings of the Annual Conference of the International Speech Communication Association, Interspeech: 1964–1968.
12. Zen, Heiga; Sak, Hasim (2015). "Unidirectional Long Short-Term Memory Recurrent Neural Network with Recurrent Output Layer for Low-Latency Speech Synthesis" .
13. Fan, et al. (2015). "Photo-Real Talking Head with Deep Bidirectional LSTM" . Proceedings of ICASSP.
14. Silver, et al. (2017). "Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm".
15. Goodfellow, et al. (2014). Generative Adversarial Networks . Proceedings of the International Conference on Neural Information Processing Systems (NIPS 2014). pp. 2672–2680.
16. Zoph, Barret; Le, Quoc V. (2016). "Neural Architecture Search with Reinforcement Learning".
17. "AutoKeras". autokeras.com. Retrieved 21 August 2019.
18. "Claesen, et al. (2015). "Hyperparameter Search in Machine Learning." arXiv preprint arXiv:1502.02127".
19. Zissis, Dimitrios (2015). "A cloud based architecture capable of perceiving and predicting multiple vessel behaviour". Applied Soft Computing. 35: 652–661.
20. Roman M. et. al. (2009). "Neural network approach to quantum-chemistry data: Accurate prediction of density functional theory energies". J. Chem. Phys. 131 (7): 074104.
21. Silver, David; et al. (2016). "Mastering the game of Go with deep neural networks and tree search" (PDF). Nature. 529 (7587):
22. Sengupta, et al.(2016). "Lung sound classification using cepstral-based statistical features". Computers in Biology and Medicine. 75 (1): 118–129.
23. UCI Machine Learning Repository
24. EasyNN Tool