Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 04 Issue 02
2023
Research Article
https://doi.org/10.56979/402/2023

Multi-Animal Recognition in Zoo Cluster using Deep Learning Technique

Sana Naeem[1], Sadaqat Ali Ramay[1], Muhammad Ashad Baloch[1], Ghulam Jillani Ansari[2], Abdul Majid Soomro[1] and Asif Hussain[1]

[1]Department of Computer Science, National College of Business Administration & Economics, Multan, Pakistan.
[2]Department of Information Sciences University of Education, Lahore (Multan Campus), Pakistan.
[*]Corresponding Author: Muhammad Ashad Baloch. Email: ashad5765@gmail.com

Received: October 12, 2022 **Accepted:** February 19, 2023 **Published:** March 29, 2023.

Abstract: The accelerated development of new methodological approaches and techniques for quantifying real-time animal recognition can be credited to the concern with the question of how the brain originates and organizes animal recognition. Even while the latest development in deep learning and computer vision have made it possible to estimate a single animal's pose, expanding this capability to several animals creates a new set of challenges for researchers who study animals in their natural environments. In this article, we will discuss the Radial Basis Function Networks (RBFNs) technique, which is a deep learning system designed for the recognition of several animals. This system makes it possible to execute a broad range of workflows for the labeling of data, the training of models, and the generation of conclusions based on data that has been seen executed several times before. RBFNs are preloaded with a graphical user interface that is easily accessible, a uniform data model, and a reproducible setup system. We deployed RBFNs to datasets spanning a wide range of zoo animals to thoroughly analyze each strategy and architecture, and we compared it to other current approaches. RBFNs can achieve both improved accuracy and increased speed. This enables the utilization of RBFNs in real-time applications, such as those in which we demonstrate the face pose of one animal based on the tracking and detection of another animal.

Keywords: Animal recognition; Deep Learning; Radia Basis Function Network; Zoo Cluster.

1. Introduction

Object detection is a computer technology that is connected with computer vision and image processing that detects meaningful things in digital photos and movies. Face detection and pedestrian detection are well-researched. Object detection is used in image retrieval and video monitoring. According to the Code of Hammurabi, animal identification by body marks dates back 3,800 years. 18th-century records show the earliest formal identity systems. Uruguay kept a hot brand list during that time. Different animals use different strategies, yet all rely on senses (After all, this is the way the organism learns about its circumstances). Chemical signatures (smell), shape or colour (sight), sounds (hearing), or behavior patterns can be used for recognition. Often, they're combined. To identify other members of their species, humans use their senses of sight and smell. Facial analysis consumes a significant amount of the processing capacity in the human brain. We can differentiate practically all humans from every other (barring look-alikes) and anthropomorphic apes with a fast glance. Intra-species identification is subtle. By eye, chiffchaffs and willow thrushes are difficult to distinguish, and there is no evidence that the birds can do so without the male's song. In several frog species, males copulate with the inappropriate species females or even inanimate stuff.

From raw input, deep learning uses many layers to extract higher-level features. In image processing, lower layers may retrieve edges while higher layers detect human concepts like numerals, faces, or letters.

The latest deep learning models are based on convolutional neural networks (CNNs), but they can also incorporate propositional formulas or latent variables grouped layer-wise in deep generative models like deep belief networks and deep Boltzmann machines. Each level of deep learning abstracts and combines its incoming data. A matrix of pixels may be the first input for an image recognition program. The first layer may encode edges and extract the pixels. The second layer may encode and assemble arrangements of edges. The third layer may capture the eyes and nose. The fourth layer may recognize a face.

Deep learning can figure out which characteristics fit at which level. Different levels of abstraction can be produced by varying the number of layers and layer widths. Deep learning transforms data across multiple levels. Deep learning systems have a deep CAP. The CAP is an input-to-output transformation. CAP depth is equal to the network depth plus one hidden layer for a feed-forward neural network (as the output layer is also mapped). CAP depth is potentially infinite in reoccurring neural networks since signals can repeatedly pass through layers in these networks. Deep learning involves CAP depths above 2, according to most studies. CAP of depth 2 can pretend any function. More layers don't improve the network's function approximation. Extra layers assist deep models (CAP > 2) learn enhanced features than shallow models.

Deep learning architectures are built layer-by-layer. Deep learning disentangles abstractions and identifies performance-boosting characteristics. Deep learning techniques minimize feature engineering by compressing data into intermediary representations that are layered to reduce representation redundancy and resemble principal components. Deep learning is capable of occurring unsupervised. Data without labels are more prevalent than data with labels. Unsupervised deep structures include deep belief networks.

1.1 Deep Learning Revolution

George E. Dahl's team won the "Merck Molecular Activity Challenge" in 2012 by employing deep neural networks with several tasks to forecast a drug's bimolecular target. Hoch Reiter's group won the "Tox21 Data Challenge" in 2014 by using deep learning to find unintended and adverse effects of environmental pollutants in food and household products, and pharmaceuticals. 2011-2012 saw significant improvements in picture or object identification. CNNs trained using back-propagation have existed for decades, as have GPU implementations of NNs, including CNNs. However, to develop computer vision, quick CNN implementations on GPUs were required with superhuman performance; this method triumphed in a visual pattern recognition competition in 2011. It won the ISBI image segmentation and ICDAR Chinese handwriting competitions in 2011. CNNs were underutilized at computer vision conferences before 2011, however, Ciresan demonstrated in June 2012 how max-pooling CNNs on GPU may break multiple benchmark records for vision. An analogous system developed by Krizhevsky won the major Image Net competition in October 2012. Ciresan won the ICPR challenge on large medical image analysis for cancer diagnosis in November 2012 and the MICCAI Grand Challenge the later year. Deep learning decreased Image Net error rates in 2013 and 2014, paralleling a development in widespread voice recognition. The production of captions (descriptions) was later added to image categorization, frequently utilizing CNNs and LSM (LSTMs).

2. Literature Review

Many biological studies require identifying individual animals. In response to present identification constraints, new automated computer vision algorithms have emerged. Here, we highlight current improvements in computer vision identification approaches for computer scientists and biologists. We conclude with tips for launching an animal identification project, constraints, and future solutions [1].

Deep learning increases speech and picture identification, autonomous driving and natural language processing. In biology, it's utilized for identification of automated species, monitoring of environmental, ecological modeling, signs investigations, population genetics ,sequencing of DNA and phylogenetic [2]. Deep learning uses neural networks to anticipate and recognize complicated patterns. Affordable sensors speed up animal ecological data collection. Current processing methodologies inefficiently distill data into usable information, limiting the potential of these technologies for large-scale ecological understanding. We claim that animal ecologists can use big sensor-generated datasets by integrating machine learning with domain expertise. Machine learning could improve ecological model inputs and lead to hybrid modeling tools. This method will demand tight interdisciplinary collaboration to assure novel approach quality and train the latest generation of ecologists and conservationists [3].

Biology and biomedicine must understand primate behavior. Our capacity to rigorously quantify behavior has been confined to low-information measurements like predilection, gazing duration, and reaction time, or non-scalable metrics like echograms. New technologies have revolutionized behavioral measurement. Automated posture tracking software and digital video cameras can offer high-resolution measurements of numerous monkeys' entire body posture (i.e., pose) over time (i.e., behavior). Pose tracking can identify eating, sleeping, and mating. Such data has spurred data analysis approaches. These improvements will lead to considerable advances in scientific domains that rely on behavioral data. In this analysis, we place the tracking revolution in the record of the study of behavior, call for investment in and development of systematic and methodologies of research, and predict those zoos will play a major role in the epoch of big behavior [4].

It's tricky to get non-invasive behavioral metrics from video. Deep learning has greatly improved our capacity to predict posture from videos, impacting neurology and biology. This article introduce deep-learning motion capture. We'll examine the concepts of these unique algorithms, emphasize their promise and limitations, and look ahead [5].

Using DL for wildlife presents the added challenge of highly varying background, causing models that classifies objects based on full images to often predict well on training data but poorly on new test data. As new images still contain the same species, but in a different background, models that classify based on the entire photo seem to experience distraction from the background. DL object detection (OD) methods have come up with an innovative way to reduce or overcome background distortion by classifying based on the localization of the animal. Two OD methods have grown in popularity especially in recent years; YOLOv5 and Pose estimation. How these models perform on determining wildlife actions is not much explored [6].

We demonstrate a technique to reconstruct 3D quadruped structure and motion from video. Machine learning predicts 2D joint positions, discrete optimization finds kinematically reasonable joint correspondences, and energy minimization fits a comprehensive 3D model to the image. The technique uses silhouette data to get over the lack of animal training data for motion capture and the difficulties of creating convincing semisynthetic training photos. At the time of test, deep learning or traditional video segmentation technologies are utilized to extract silhouettes from actual data. The algorithm accurately reconstructs 3D shapes and position from many animal recordings [7].

Recent advances in machine vision approaches for autonomous, quantitative analysis of social behavior have improved the scope and resolution with which we can dissect interactions between members of the same species. This study reviews various strategies, focusing on how biologists can use them. We explore recording high-quality video for automated analysis, video-based tracking techniques for estimating animal placements, and machine learning methods for recognizing interaction patterns. We evaluate the successful applications of these approaches to biological concerns in numerous model systems with different social behaviors [8].

The quantification of behavior in neuroscience and ethology is changing as a result of recently developed visual analysis techniques, particularly models for position estimation and behavior classification. These innovations go beyond traditional "center of mass" tracking techniques and manual video frame scoring to offer scalable video analysis. Open-source video capture and processing tools have led to novel behavioral experiments. Here, we discuss open-source video analytical techniques, how to set them up in a lab that is just starting to record video, and some concerns advanced users and developers should have. These concerns include the demand for documentation and community-wide standards, the need to freely share datasets and code, and how to compare algorithms and their parameters. We want to use and develop tools more. They can accelerate brain and behavior research [9].

Deep learning has reshaped bioimage analysis by transforming how big, complex picture collections are treated. As bioimaging data complexity and size rise, so does this new analysis paradigm. This Review introduces deep learning topics for beginners. Furthermore, we investigate how deep learning has improved bio-image analysis and open-source resources for integrating it into a research endeavor. Eventually, we examine deep learning in cell and developmental biology. We evaluate how modern image-based analysis and modeling can revolutionize our understanding of biological systems [10, 11].

To understand psychiatric and neurodegenerative illnesses, animal models must accurately measure social interactions. Callithrix jacchus is a good model for these illnesses. However, sociality has not been assessed in free-moving group animals. We built a 3D animal behavioral analysis system for free-moving individuals [12].

Neuroscientists use real-time animal behavior to control or trigger brain activity. The best tools are non-intrusive, low-latency, and external devices based on position. The latest breakthroughs in deep learning posture estimation allow researchers to properly measure animal behaviors. New Deep Lab Cut-Live! The package performs low latency actual time posture estimation (inside 15 ms, >100 FPS), with a zero-latency onward prediction module and a dynamic-cropping mode [13].

Wild feline behavior analysis helps protect grassland ecosystems. Comparatively few researchers have studied feline behavior. In this study, a two-stream architecture for recognizing wild feline activity is presented. The spatial part describes the object region that R-CNN found and creates a VGG network for recognizing static actions. Tiny VGG reduces network parameters and prevents overfitting compared to VGG16. The temporal component introduces a skeleton-based action identification model based on knee joint bending angle fluctuation amplitude. Due to its temporal properties, the model can discriminate between standing, ambling, and galloping, even when the felines are obstructed by vegetation, fallen trees, etc. A skeleton-based movement identification model based on knee joint bending angle variation amplitude is introduced in the temporal component. The model can distinguish between standing, ambling, and galloping because of its temporal characteristics, even when the felines are obscured by foliage, downed trees, etc. The proposed two-stream network strategy can more accurately identify wild feline action in gathered photographs because of its spatial and temporal features [14].

Zoos have helped humans acquire, maintain, and learn about native and exotic animals throughout history. Zoos save species, educate the public, and preserve endangered animals. Modern zoos balance animal care and visitor education to benefit the environment. This endeavor promotes the study, conservation, education, and entertainment while conserving animals and their habitats more accurately. "Zoo and aquarium design shouldn't just create creative ways to shelter and see animals; it should engage visitors in conservation and animal welfare missions [15]."

Neuroscientists use real-time animal behavior to control or trigger brain activity. Noninvasive, low-latency instruments that activate external equipment based on posture are ideal. Recent breakthroughs in deep learning posture estimation allow researchers to properly measure animal behaviors. New Deep Lab Cut-Live! The package performs low-latency actual-time posture estimation (within 15 ms, >100 FPS), with a zero-latency onward prediction module and a dynamic-cropping mode. This utility has three easy-to-use options: DLC-Live! GUI and connection with Bonsai and Auto Pilot. Finally, we benchmarked a extensive range of systems so experimentalists can readily choose hardware [16].

Despite advancements, it is still difficult to study weakly electric fish, especially pulse-type species, because each individual must be allocated short signal epochs at different frequencies ranging from a small number of hertz to more than 100 Hz. Here, we demonstrate that supervised learning algorithms may automate or partially automate the workflow, enabling the study of longer behavioral episodes in a sensible amount of time. We give an open-source process. We show the approach's efficacy by analyzing Gnathonemus petersii dyadic interactions. By combining the presented methods with boundary element modelmodelingcan simulate agonistic encounter information. Before participating in agonistic behavior, fish did not use or depend on the theoretically obtainable sensory knowledge of the contest achievement difference size amongst competitors. [17].

3. Methodology

The data engine utilizes a Fused Deep Learning architecture for the purpose of recognizing and categorizing different species of animals. The retrained model and the pyramid histogram of orienteered gradient feature extraction technique are incorporated into the design. Both approaches are able to extract features from the photos contained in the Zoo dataset. After the features have been retrieved, they are concatenated in a sequential fashion, and then the most crucial traits are chosen using the DL approach. A flow chart of the Animals classification process is shown in Figure 1.

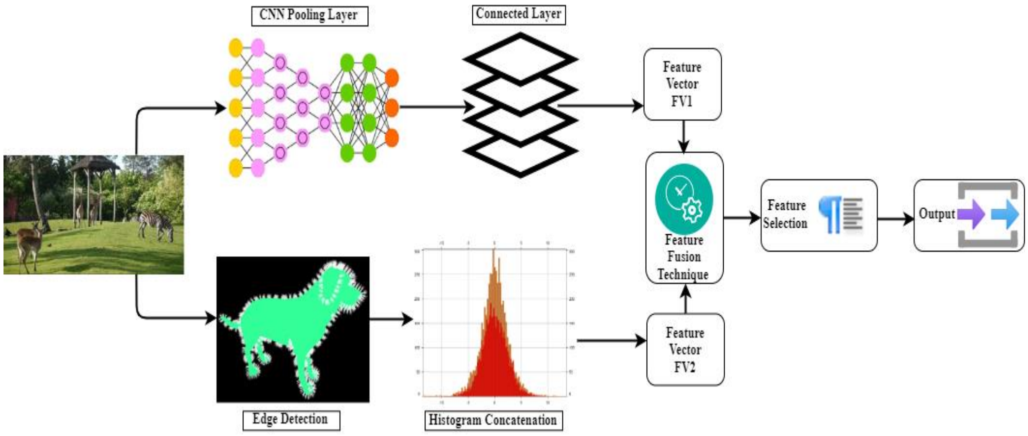

Figure 1. Proposed Architecture of Zoo Animal Classification

3.1 Pre-processing and Data Acquisition

In any process involving Machine Learning, the stage known as "Data Preprocessing" is the one in which the data undergoes a transformation, also known as "Encoding," to bring it to a state in which it can be easily parsed by the machine. To put it another way, the algorithm is now in a position to readily interpret the characteristics of the data. The data transformation, normalization, and management of missing values are handled by this layer. The application layer receives the data that has been processed.

3.1.1 Deep Neural Network Techniques

Artificial intelligence (AI) refers to the creation of computer systems that are capable to adapt and learn without following explicit instructions. This is accomplished via the use of statistical models and algorithms that examine patterns and take out conclusions based on those analyses. Artificial deep neural networks algorithms are used in this layer to help predict patterns. The application layer's output will serve as the input for the performance layer.

3.1.2 Validation of Cloud Data

The process of verifying that the results are accurate and of high quality is referred to as data validation. Putting it into practice involves incorporating a number of tests into a system or report in order to assure the data that is input and which is stored are logically consistent. Precision and error rates are assessed using the performance layer. If the data meet to learning criteria save to cloud database, else update in application layer. Cloud data validate in the validation phase with real time analysis and confirm certainty and uncertainty.

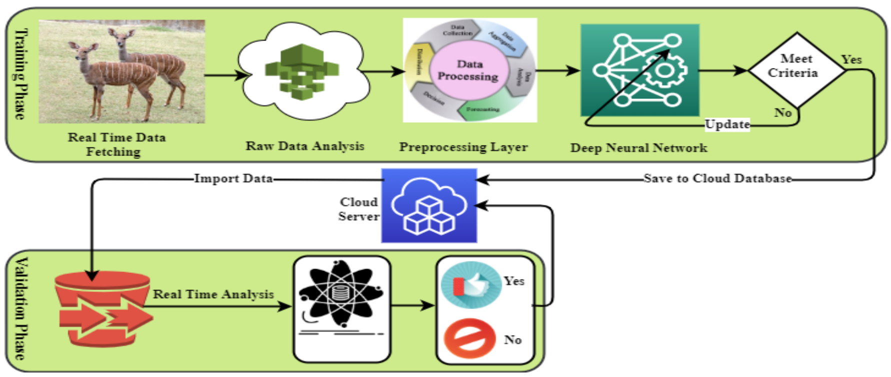

Figure 2. Proposed Methodology for Real Time Analysis

3.2 Algorithm Overview

RBF networks utilized in mathematical modeling. This network activates with radial basis functions. The result of the network is a linear combination of neuron parameters and input RBFs. RBF networks can approximate functions, predict time series, classify data, and control systems. Broom head and Lowe, Royal Signals and Radar Establishment researchers, initially formulated them in 1988.

RBF networks, which stand for radial basis function, usually contain three layers: the first input layer, a second hidden layer that contains a non-linear RBF activation function, and third is linear output layer. It is possible to represent the input as a vector of real values. Therefore, the outcome of the network is a scalar function of the input vector, and the equation for this is as follows:

Algorithm 1: Predicting Multi-animal Recognition in zoo cluster by using deep learning Network(M, λ, C)

Input: Z = ((X_1, Y_1)… (Xi, Yi)), is the order of labeled training patterns

M represents the centers of RBF

λ represents the constant of Regularization

C represents iterations in the network

Initialization of network:

Here, M- represents the cluster to gain preliminary values for μ_K & find out σ_M, M = 1, M, represents the interval between μ_M and $\mu_{i.}$ where *(i is not equal to M).*

foreach c= 1: C,

1. Calculate optimal output weights w = $(G^T G + 2_\lambda/\mathrm{i})^{-1}\, G^T Y$.

2a. Calculate gradients $\frac{\partial}{\partial \mu M} E$ and $\frac{\partial}{\partial \sigma M} E$, with optimum w & v as vector gradient.

2b. Evaluate the direction of conjugate $\bar{v}$ by using CG-Method of Fletcher Reeves Polak Ribiere

3a. To get the reducing step size by using line search δ direction $\bar{v}$; in every perception of E to re-calculate the optimum weights as output that was represent in line one.

3b. remake σM & μM with $\bar{v}$ & δ

Desired output :get perfect Radial basis network

3.2. 1 Network Architecture

Networks with radial basis functions, which are also known as RBF networks, usually consist of 3 layers: first is input layer second is hidden layer with a non-linear activation function, and third is linear output layer. RBF, short for radial basis function, is intended as an abbreviation. Modeling the input as a vector of actual values is possible $x \in R^n$. The network's output is afterwards a scalar function from the input vector, $\varphi : R^n \rightarrow R$ and given as,

$$\varphi(x) = \sum_{i=1}^{N} a_i \mathrm{P}\,(\|x - c_i\|) \tag{1}$$

N represent the amount of neurons in the network hidden layer, Ci represent the weight of neuron I and I represent the centre vector for neuron I in the linear output neuron. Is the amount of neurons in the hidden layer? The term "radial basis function" refers to functions that are radially symmetric about a given vector and whose behaviour is determined solely by the distance from a certain centre vector. In its most fundamental form, each hidden neuron receives input from all of the other hidden neurons. The Euclidean distance is generally considered to be the standard, despite the fact that When it comes to pattern recognition, the Mahalanobis distance seems to work superior, and the radial basis function is often thought of as the standard Gaussian,

$$\mathrm{P}\,(\|x - c_i\|) = \exp\,[\beta_i \|x - c_i\|^2 \tag{2}$$

The Gaussian basis functions are considered to be local to the center vector in the sense that modifying the parameters of a single neuron has a negligible impact on the input values that are situated at the periphery of that neuron.

$$\lim_{\|x\|\to\infty} \rho(\|x - c_i\|) = 0 \tag{3}$$

RBF networks are universal approximations on a compact subset of Rn under some lax requirements on the form of the activation function. This indicates that every continuous function on a closed, constrained set can be approximated by an RBF network with sufficient hidden neurons with arbitrary precision.

3.2.2 Normalized Architecture

RBF networks can be stabilized in addition to the aforementioned unnormalized architecture. The mapping in this instance is denoted as a normalized radial basis function.

$$(\varphi)X \stackrel{\text{def}}{=} \frac{\sum_{i-1}^{N} \alpha i\rho(\|X - ci\|)}{\sum_{i-1}^{N} \rho(\|X - ci\|)} = \sum_{i-1}^{N} \alpha iu(\|X - ci\|) \tag{4}$$

Where,

$$u(\|X - ci\|) \stackrel{\text{def}}{=} \frac{\rho(\|x - c_i\|)}{\sum_{j-1}^{N} \rho(\|X - ci\|)} \tag{5}$$

3.2.3 Training RBFNs

RBF networks are usually trained from combinations of input and combination of target values X(t), Y(t), t-1,… by 2 step algorithm.

$$M(W) \stackrel{\text{def}}{=} \sum_{t-1}^{T} M_t(W) \tag{6}$$

Where,

The first thing that is done is picking out the centre vectors Ci for the RBF functions that are in the hidden layer. This step can be perform in a number of different methods; for instance, the centers can be selected at random from a given set of samples, or they can be established through the utilization of k-means clustering. This phase does not have any supervision..

In the second stage, all that is done is to apply a linear model with coefficients Wi to the outputs of the hidden layer while taking into account several objective function. The least squares function is a typical example of an objective function, at the very least for regression and function estimation.

$$M_t(W) \stackrel{\text{def}}{=} [y(t) - \varphi(X(t), W)]^2 \tag{7}$$

They specifically incorporated the model's dependence on the weights. The precision of the fit can be improved by selecting the appropriate weights so that the objective function of least

squares is minimized to its maximum. There are times when it is necessary to maximize not just one but some different goals, such as smoothness in addition to accuracy. In such a scenario, it would be beneficial to maximize a regularised objective function such as

$$\mathrm{H(W)} \stackrel{\text{def}}{=} M(W) + \lambda S(W) = def \sum_{t=1}^{T} H_t(W) \tag{8}$$

Where,

$$S(W) \stackrel{\text{def}}{=} \sum_{t=1}^{T} S_t(W) \tag{9}$$

And

$$H_t(W) \stackrel{\text{def}}{=} M_t(W) + \lambda S_t(W) \tag{10}$$

4. Result and Discussion

4.1 Confusion Metrics Expression

A confusion matrix, also called an error matrix, is a particular table layout that, when applied to the area of machine learning and more explicitly the problem of statistical classification, enables visualization of the performance of an algorithm, most often a supervised learning one. While every row of the matrix epitomizes the cases that belong to an actual class, every column of the matrix epitomizes the instances that belong to a predicted class or vice versa the literature refers to both of these permutations. The name comes from the fact that it is clear to decide either the system is conflating any two classes or not, which means that it is frequently misidentifying one as being the other).

Table 1. Mathematical Equation of Confusion Metrics

Mathematical Equation	Equation
Sensitivity	$\frac{TP}{TP+FN}$
Specificity	$\frac{TN}{TN+FP}$
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
Miss Rate	$\frac{FN}{FN+TP}$
Fall out	$\frac{FP}{FP+TN}$
LR +	$\frac{TPR}{FPR}$
LR –	$\frac{FNR}{TNR}$
Precision	$\frac{TP}{TP+FP}$
Negative Predicted Value	$\frac{TN}{TN+FN}$

4.2 Graphical Representation

The figure 3 (a), (b), and (c) show the accuracy rate with animal classes.

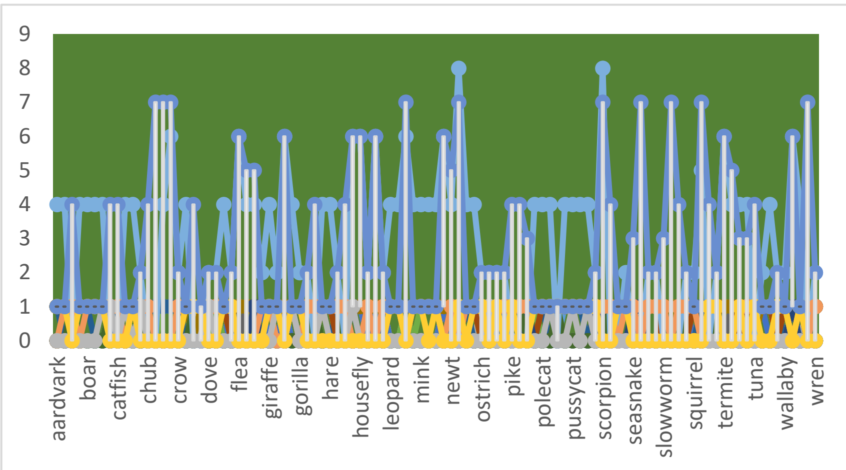

Figure 3 (a) Accuracy Collaboration

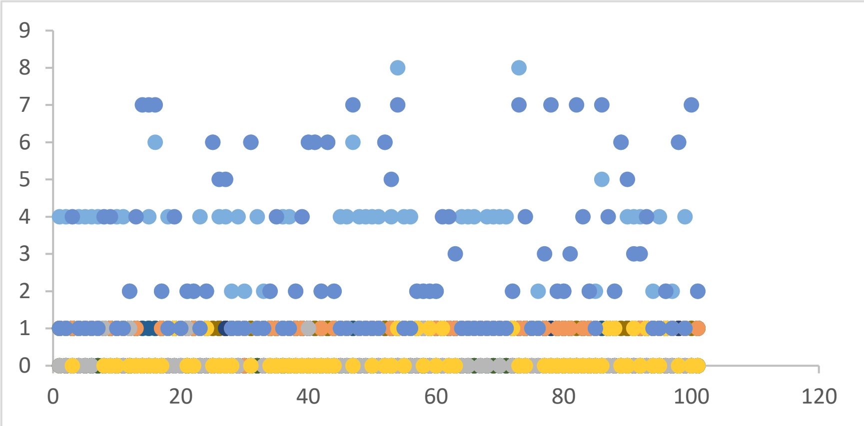

Figure 3 (b) Accuracy rate of proposed model

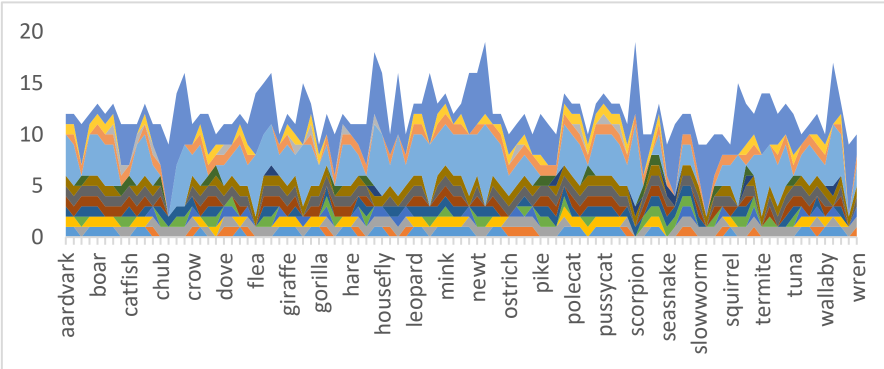

Figure 3 (c) Accuracy rate with animal classes

4.3 Training and Validation

Evaluation of the Efficiency of the Proposed Model concerning Training and Validation Utilizing a Wide range of statistical Measures.

This dataset is the benchmark test set that is utilized most frequently for animal classifications (kaggle.com). This data set contains information that has been separated into two distinct sections: the first is training dataset and other is validation dataset. The data used for training have a specific identifier, whereas the data used for testing do not have any identification. Other forms of attacks were not included in the training samples that are included in the test data. Because of this, the identification provided by the system is both more precise and trustworthy.

In Table 2, there is 70% dataset utilized in the training stage, as well as 30% dataset utilized in the Validation stage.

Table 2. Training and validation Stages

Phases	Accuracy	TPR	TNR	FNR	FPR	LR+	LR-	PPV	NPV
Training	97.8	0.983	0.965	2.4	0.035	28.08	0.025	0.981	0.967
Validation	95.7	0.968	0.937	4.3	0.063	15.36	0.045	0.966	0.940

Table 2. Represent the performance of the proposed RBFNs model in terms of "accuracy, Sensitivity TPR, TNR, miss rate & PPV (precision)" throughout the training stage and validation stage. Radial basis network model provides 97.8%, 0.965, 0.983, 2.4%, and 0.981 accuracy of detection, specificity, sensitivity, miss rate, and precision at the time of training phase. At the time of validation, Radial basis network model

provides 95.7%, 0.937, 0.968, 4.3%, & 0.966 accuracies of detection, specificity, sensitivity, miss rate, and PPV (precision), appropriately. Moreover, several statistical measures of the proposed RBFNs model are encompasses to calculate the values during training phase, such as fall-out FPR, LR+, LR-, and NPV provide the result 0.035, 28.08, 0.025, & 0.967 during validation 0.063, 15.36, 0.045, and 0.94 appropriately.

4.4 Comparison Analysis

Table 3. Represents the comparison of the performance of proposed Radial basis network model with other existing models. RBFNs is better than other existing models like ResNet18, VGG16, Novel 5L CNN, AlexNet, DenseNet201, InceptionV3, MixtureNet, ResNet18, Xception. The accuracy of the proposed model is 97.8%, which is better than other models.

Table3. Comparison analysis of the proposed model.

Sr.	Article	Model	Accuracy
1	Willi et al [18].	ResNet18	96.60
2	J. K. Teto & Y. Xie [19]	VGG16	96.40
3	Rathi et al [20].	Novel 5L CNN	96.29
4	Yousif et al [21].	AlexNet	95.60
5	Schneider et al [22].	DenseNet201	95.60
6	Allken et al [23].	InceptionV3	94
7	Zuluaga et al [24].	MixtureNet	92.65
8	Hu M. & You F [25].	ResNet18	92
9	Lai et al [26].	Xception	91.29
10	Proposed	RBFNs	97.8

5. Conclusion

In this article, we presented RBFN, a multi-animal pose tracking data classification system with generic deep learning capabilities. This technique improves the state of the art for both solitary-animal as well as multi-animal pose estimates, and it incorporates these advancements inside a versatile and effective open-source framework that was created and proven by non-technical practitioners. The RBFN's modular design builds it straightforward to recognize the root of inaccuracies or subpar performance, which can then guide changes to the technique of data gathering and the overall experimental design. Moreover, this model exposes modular functionality through defined APIs, making it simple for other frameworks to adopt, and we offer data-export formats to facilitate the portability of outputs so that they may be used in downstream analytic frameworks. This allows for greater flexibility.

References

1. Vidal, M., Wolf, N., Rosenberg, B., Harris, B. P., & Mathis, A. (2021). Perspectives on individual animal identification from biology and computer vision. Integrative and Comparative Biology, 61(3), 900-916.
2. Borowiec, M. L., Dikow, R. B., Frandsen, P. B., McKeeken, A., Valentini, G., & White, A. E. (2022). Deep learning as a tool for ecology and evolution. Methods in Ecology and Evolution, 13(8), 1640-1660.
3. Tuia, D., Kellenberger, B., Beery, S., Costelloe, B. R., Zuffi, S., Risse, B., . . . Burghardt, T. (2022). Perspectives in machine learning for wildlife conservation. Nature communications, 13(1), 1-15.
4. Hayden, B., Park, H. S., & Zimmermann, J. (2021). Automated tracking of primate behavior. arXiv preprint arXiv:2108.13486.
5. Mathis, A., Schneider, S., Lauer, J., & Mathis, M. W. (2020). A primer on motion capture with deep learning: principles, pitfalls, and perspectives. Neuron, 108(1), 44-65.
6. van Gils, J. (2022). Recognition of wildlife behaviour in camera-trap photographs using machine learning.
7. Biggs, B., Roddick, T., Fitzgibbon, A., & Cipolla, R. (2018). Creatures great and smal: Recovering the shape and motion of animals from video. Paper presented at the Asian Conference on Computer Vision.
8. Panadeiro, V., Rodriguez, A., Henry, J., Wlodkowic, D., & Andersson, M. (2021). A review of 28 free animal-tracking software applications: current features and limitations. Lab animal, 50(9), 246-254.
9. Luxem, K., Sun, J. J., Bradley, S. P., Krishnan, K., Pereira, T. D., Yttri, E. A., . . . Laubach, M. (2022). Open-Source Tools for Behavioral Video Analysis: Setup, Methods, and Development. arXiv preprint arXiv:2204.02842.
10. Hallou, A., Yevick, H. G., Dumitrascu, B., & Uhlmann, V. (2021). Deep learning for bioimage analysis in developmental biology. Development, 148(18), dev199616. doi:10.1242/dev.199616.
11. Hallou, A., Yevick, H., Dumitrascu, B., & Uhlmann, V. (2021). Deep learning for bioimage analysis. arXiv preprint arXiv:2107.02584.
12. Yurimoto, T., Kumita, W., Sato, K., Kikuchi, R., Shibuki, Y., Hashimoto, R., . . . Kurotaki, Y. (2022). Development of a new 3D tracking system for multiple marmosets under free-moving conditions. bioRxiv.
13. Kane, G. A., Lopes, G., Saunders, J. L., Mathis, A., & Mathis, M. W. (2020). Real-time, low-latency closed-loop feedback using markerless posture tracking. Elife, 9, e61909.
14. Feng, L., Zhao, Y., Sun, Y., Zhao, W., & Tang, J. (2021). Action Recognition Using a Spatial-Temporal Network for Wild Felines. Animals 2021, 11, 485: s Note: MDPI stays neutral with regard to jurisdictional claims in published ….
15. Whipple, A. E. (2019). Creating Authentic Experiences in Zoos: Exploring How Design Enhances Visitor Experience and Animal Welfare. The Florida State University.
16. Mathis, W. Gary Kane1, Gonçalo Lopes2, Jonny L. Saunders3, Alexander Mathis1, 4, Mackenzie.
17. Pedraja, F., Herzog, H., Engelmann, J., & Jung, S. N. (2021). The Use of Supervised Learning Models in Studying Agonistic Behavior and Communication in Weakly Electric Fish. Frontiers in Behavioral Neuroscience, 15.
18. Willi, M., Pitman, R. T., Cardoso, A. W., Locke, C., Swanson, A., Boyer, A., . . . Fortson, L. (2019). Identifying animal species in camera trap images using deep learning and citizen science. Methods in Ecology and Evolution, 10(1), 80-91.
19. Kamdem Teto, J., & Xie, Y. (2018). AUTOMATIC IDENTIFICATION OF ANIMALS IN THE WILD: A COMPARATIVE STUDY BETWEEN C-CAPSULE NETWORKS AND DEEP CONVOLUTIONAL NEURAL NETWORKS.
20. Rathi, D., Jain, S., & Indu, S. (2017). Underwater fish species classification using convolutional neural network and deep learning. Paper presented at the 2017 Ninth international conference on advances in pattern recognition (ICAPR).
21. Yousif, H., Yuan, J., Kays, R., & He, Z. (2019). Animal Scanner: Software for classifying humans, animals, and empty frames in camera trap images. Ecology and evolution, 9(4), 1578-1589.
22. Schneider, S., Greenberg, S., Taylor, G. W., & Kremer, S. C. (2020). Three critical factors affecting automated image species recognition performance for camera traps. Ecology and evolution, 10(7), 3503-3517.
23. Allken, V., Handegard, N. O., Rosen, S., Schreyeck, T., Mahiout, T., & Malde, K. (2019). Fish species identification using a convolutional neural network trained on synthetic data. ICES Journal of Marine Science, 76(1), 342-349.
24. Zualkernan, I., Dhou, S., Judas, J., Sajun, A. R., Gomez, B. R., & Hussain, L. A. (2022). An IoT System Using Deep Learning to Classify Camera Trap Images on the Edge. Computers, 11(1), 13.
25. Hu, M., & You, F. (2020). Research on animal image classification based on transfer learning. Paper presented at the Proceedings of the 2020 4th International Conference on Electronic Information Technology and Computer Engineering..
26. Lai, K., Tu, X., & Yanushkevich, S. (2019). Dog identification using soft biometrics and neural networks. Paper presented at the 2019 International Joint Conference on Neural Networks (IJCNN).