International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Transfer Learning for Animal Species Identification from CCTV Image: Case Study Zakouma National Park

Oumar Hassan Djibrine[1], Daouda Ahmat[1], Moussa Mahamat Boukar[2], Usman Abubakar Bello[3], Azza Youssouf Ali[1]

Submitted: 17/08/2023 **Revised**: 07/10/2023 **Accepted**: 21/10/2023

Abstract: The precise identification of animal species is of utmost importance in comprehending biodiversity, conducting surveillance of endangered species, and examining the potential ramifications of climate change on the geographical dispersion of species in specific areas. Closed-circuit television (CCTV) cameras represent a form of passive surveillance technology that generates a substantial volume of ecological imagery. The manual evaluation of extensive datasets is a labor-intensive, time-consuming, and costly process, underscoring the importance of automated ecological analysis. In the realm of computer vision, there has been significant progress in addressing issues related to object and species recognition through the development of deep learning networks. These networks have demonstrated state-of-the-art performance, showcasing their effectiveness in this domain. In this study, we have undertaken the development and evaluation of machine learning models with the purpose of classifying animal groups based on camera trap images. Transfer learning was employed in our study to conduct experiments with VGG19, GoogLeNet (InceptionV3), ResNet50, and DenseNet121. CNN-1 achieved an accuracy of 53% in multiclassification, whereas GoogLeNet demonstrated a higher accuracy of 87%. Similarly, ResNet50 earned an accuracy of 83%, DenseNet reached 81%, and VGG achieved 79%. The findings of this study indicate that the utilization of transfer learning yields superior performance compared to that of the self-trained model. The models exhibited encouraging results in terms of species detection and classification within Zakouma National Park.

Keywords. *CCTV, species, image classification, deep learning, machine learning, Convolution Neural Network, image augmentation.*

1. Introduction

The utilization of remote cameras or video traps has emerged as a groundbreaking instrument in the realm of wildlife ecology and conservation monitoring [1], gaining increased significance in tandem with advancements in the technology. This very effective instrument takes a substantial quantity of images, with each image potentially offering a comprehensive range of data regarding the existence of an animal inside a strategically chosen research area [1], its population magnitude, and its interactions within the community [2]. researchers have the ability to collect biographical and critical evidence in a remote manner, free from any potential interference caused by human observation [2,3]. The unprocessed images obtained can be archived for subsequent analysis [3,4]. In addition, the inclusion of imagery samples offers supplementary data for detection purposes, encompassing specific details such as the exact day, time, and environmental circumstances throughout the imaging process [5].Zakouma National Park functions as a protected area that provides refuge for several species of animals originating from West and Central Africa, a region where many of these species face endangerment[6]. The park harbors a total of 66 mammal species, of which 16 are classified as large animals. Notably, the park saw the introduction of two black rhinoceros specimens in the year 2018. Zakouma National Park harbors a variety of species that are currently facing threats or are at risk of extinction. Notably, it is a significant habitat for the Kordofan giraffe population, accounting for almost 50% of the whole population found in Africa. This subspecies of giraffe is critically endangered. Additionally, the park provides a sanctuary for the Lelwel hartebeest, the north-east African cheetah (Acinonyx jubatus soemmeringii), and Buffon's kob, all of which are either vulnerable or endangered. The floodplains, rivers, marshes, and pans of Zakouma serve as significant habitats for migrating birds, providing essential resting and nesting locations. Additionally, the wetlands in the southeastern region of Zakouma are included in the Inundation Plains of Bahr Aouk and Salamat Ramsar site, which is recognized as one of the world's largest wetland areas [7].

[1] *Department of Computer Science, Université Virtuelle du Tchad*
[2] *Department of Computer Science, Nile University, Abuja, Nigeria.*
ORCID ID: 0000-002-2494-2698
[3] *Department of Computer Science, Baze University, Abuja, Nigeria.*
** Corresponding Author Email: musa.muhammed@nileuniversity.edu.ng*

Fig 1. Meta herds of elephant occupy Zakouma, indicating population growth

The buffalo population in Zakouma National Park has had a substantial growth, rising from a mere 220 individuals in 1986 to a current estimate of over 15,000 individuals. The aforementioned circumstance facilitated the first reintroduction endeavor conducted in the Siniaka-Minia Wildlife Reserve, during which a total of 900 buffalo specimens were successfully translocated. The buffalo that were reintroduced originated from a substantial herd that faced the imminent threat of encroaching into agricultural territories. Despite experiencing many setbacks, the translocated herds have successfully established a founder population in Siniaka Minia, marking the emergence of a substantial population that has not been seen for several decades. Efforts are now being made to initiate the reintroduction of rhinoceroses to the park. Notably, two female specimens of the wild black rhinoceros species have already been effectively tranquilized, subjected to thorough examination, and equipped with tracking devices. The rhinoceros population in Zakouma has positive health indicators, therefore validating the suitability of its habitat despite past setbacks encountered during relocation efforts.

Fig 2. Conservation law enforcement teams monitor and track wildlife.

2. Review of Literature

2.1. Zakouma National Park size, history, location, and map

2.1.1. History

The establishment of Zakouma Hunting Reserve in 1958 may be attributed mostly to the endeavors of Michael Anna, who demonstrated a keen awareness of the ecological perils confronting the area [9]. Following the establishment of the reserve, there was a significant surge in animal populations in Chad, thus drawing the attention of poachers. In 1963, Zakouma was officially designated as a national park, so granting it the utmost degree of safeguarding as stipulated by the legal framework of Chad. Chad has submitted an application for the official recognition of Zakouma National Park (ZNP) as a UNESCO World Heritage Site. The United Nations Educational, Scientific and Cultural Organization (UNESCO) conducted visits to the ZNP (Zoological National Park) in both 2009 and 2011, resulting in notable advancements towards the attainment of the stated objective. Nevertheless, the lingering boundary dispute in the vicinity of Kiéké, including the settlements of Bone Daoud and Bone Fakhara, has the potential to impede the timely resolution of the matter.

The southeastern region of ZNP is included inside the Plaines d'Inondation des Bahr Aouk et Salamat RAMSAR site, which has the distinction of being one of the most expansive RAMSAR sites globally, spanning an impressive area of 4,922,000 hectares. This website has significant importance because to its prominent role as a primary catchment region for the Chari River, a key tributary that ultimately feeds into Lake Chad. The 50th anniversary of ZNP was commemorated on February 21, 2014, with the esteemed presence of the President of Chad. During the incident, the whole ivory stockpile belonging to Chad, which amounted to a total of 1.1 tons, was subjected to incineration. In recognition of their contributions to conservation efforts in Chad, a number of individuals, including three members of the AP crew, were bestowed with civilian medals.

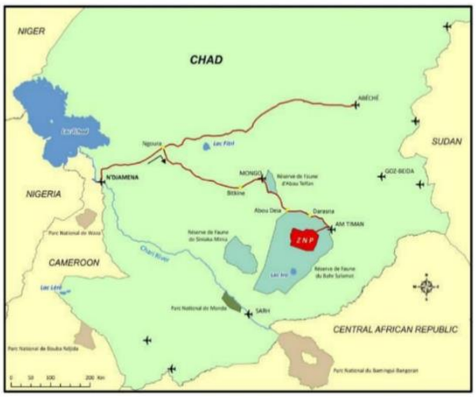

Fig 3. The location of Zakouma National Park

2.1.2. Key conservation threats and issues

The main threats currently affecting the park can be summarised as follows

Table 1: Key conservation threats.

Threat	Description
Elephant Poaching	GEFZ lost about 90% of its elephants over an eight-year period and poaching remains to be a threat. The Zakouma elephant, moving in dense herds, has resulted in a devastating poaching method where herds are ambushed and machine gunned indiscriminately resulting in substantial immediate and subsequent losses, wounded animals, and high levels of stress. The presence of calves in the population further increases the number of animals killed during a poaching incident as they are either orphaned or lost/trampled during the stampede.
Low breeding in elephant herds due to poaching stress	Up until mid-2013 there were very few elephant calves below two years of age, presumably due to the heavy poaching stress over the past years, or severe loss of calves due to poaching related stampeding of the herds, where small calves tend to get left behind. This resulted in almost 0% recruitment during the first 3 years. Fortunately, this negative trend has been reversed and new-born calves are now seen regularly. To date, at least 100 calves have been born since mid-2013; a growth rate of about 5% per annum. With the results of the aerial survey done in March 2016, the population is now officially on the increase, with calves of 3 years and younger comprising at least 16% of the population. Stability has been returned to the population, and all efforts are being made to keep them as free from poaching as possible, or at least to detect poaching as soon as it happens, and thereby ensure no ivory enters the market.
Seasonal wildlife movement within the periphery with Land Use Plan not yet in place	During the wet season large numbers of ungulates (and in the past elephants) move within an area of up to 20,000 km², most of which lies outside the protected areas and is mostly inaccessible except by foot and horseback.
Expanding sorghum fields in the corridors and encroachment of park boundaries	Sorghum is becoming an important cash crop in the Salamat Region (not only for local use) and fields are expanding into the migration corridors, wet season areas and the south-eastern corner of the park. Even though a solution has been agreed on in 2015 for the agriculture at Kiéké, the situation still remains to be resolved.
Unmanaged conversion of land and natural resource use in the park periphery	Wetlands and pans around the park are being converted for rice and other cultivation. These provide important breeding habitat for birds, in particular the threatened black crowned crane and all fish species. Farming and fishing on Gara plain have drastically increased in the last five years. The flooding of 2014 which caused serious road damage in the park was almost certainly as a result of cultivation in catchment areas and the resultant increase in run-off.
Fish harvesting inside and outside the park	The reduction in fish in the major lakes in Chad (Lakes Chad, Fitri and Iro) has resulted in a movement of commercial fisherman to the Zakouma Ecosystem which is resulting in heavy commercial fishing outside the park, as well as inside. This has negatively affected the economy of the local fishing industry.
Nomads with no formal links or agreements with the park and free movement close to the elephant herds throughout the year	Nomad groups share the migration routes with the wildlife. It is difficult to work with the leaders as the different groups are constantly on the move, however efforts are being made and relationships have greatly improved. These people also know the areas and elephant movements well. Some nomadic groups have a long tradition of hunting and ivory trade, and many carry weapons against livestock theft.
Bone village growing within the park	After a successful negotiation between the Government of Chad and Bone village, the village has agreed to come up with an alternative location for the village outside the park, at a distance of at least 5km from the current park boundary. The negotiations will continue at the end of October 2017.
Fast growing human population surrounding the park with a low education level	Looking at the number of children in the periphery villages and the quality of education in rural Chad it is clear that the demand on resources will increase dramatically in coming years, with little chance of educating these children on environmental issues through the normal school system.
Post war security issues	High numbers of firearms in circulation everywhere in the country.

2.2. Related work

Various studies have used machine learning algorithms in order to identify animal species shown in images but with a predominant emphasis on mammals and avian creatures [10]. There is a limited number of research that have used deep learning techniques for the purpose of species identification in camera trap images, with the majority of these studies focusing on images captured under controlled conditions. In contrast, the present research used a combination of primary and secondary datasets including camera trap images. There is a significant collection of animal data being conducted through citizen science initiatives[11]. Several examples of platforms that are often used for citizen science projects are Zooniverse, iNaturalist, Pl@ntNet, and Flora Incognita[12].In 2013, Yu et al. published a pioneering research that used a camera trap dataset in conjunction with machine learning techniques, marking the first comprehensive endeavor of its kind. The researchers used an advanced methodology known as improved sparse coding spatial pyramid matching (ScSPM) to extract distinctive characteristics from the data. These features were then classified using a support vector machine (SVM) algorithm [13]. An accuracy rate of 82% was attained.

The first implementation of a convolutional neural network (CNN) for species identification using camera traps was conducted by Chen et al[13]. in their study [14]. A comparison was conducted between a bag of visual words (BOW) technique and a deep convolutional neural network (DCNN) method. The researchers used a scale-invariant feature transform (SIFT) methodology to extract features for bag-of-words (BOW) representation. Subsequently, a linear support vector machine (SVM) was utilized for the classification of these features. The researchers used a dataset that was both intricate and characterized by high levels of noise. The dataset consisted of 14,346 training photos and 9,530 testing images, including a total of 20 species that are often seen. The deep convolutional neural network (DCNN) demonstrated superior performance when compared to the bag-of-words (BOW) approach, with an accuracy rate of 38.315% as opposed to 33.507%.

Since the year 2016, scholars have used pre-trained deep convolutional neural network (DCNN) architectures with openly accessible citizen science datasets for the purpose of species detection. In their study, Gomez et al. [15] used the unbalanced Snapshot Serengeti dataset to conduct a classification task on 26 separate groupings of animal species. To do this, they utilized eight different convolutional neural network (CNN) frameworks, including AlexNet, VGGNet, GoogLeNet, and ResNets. The convolutional neural network (CNN) architectures exhibited a diverse set of layer configurations, with AlexNet consisting of 8 layers and ResNet-152 boasting a significantly larger number of layers, specifically 152. The ResNet-101 model demonstrated superior performance compared to other models [15].

2.3. Objective

This study aimed to develop an animal recognition model for the identification of animals in Zakouma National Park. To achieve this, a combination of datasets obtained from Zakouma National Park and the Kaggle repository was utilized. The approach employed in this research involved transfer learning. The study investigated two models: a convolutional neural network (CNN) model called self-trained CNN model , and transfer learning using four pretrained models: VGG19, GoogLeNet (InceptionV3), ResNet50, and DenseNet121. In the context of the self-trained model, our objective was not to surpass existing state-of-the-art image classification architectures or pretrained models like VGG19, GoogLeNet (InceptionV3), ResNet50, or DenseNet121. Rather, our aim was to examine the influence of an efficient network and augmentation parameter on a comparatively limited dataset. The findings of this study indicate two key points. Firstly, they highlight the benefits of employing a robust network and utilizing pretrained weights for both feature extraction and classification tasks. Secondly, the study underscores the significance of augmentation parameters in influencing the performance of four transfer learning models and a self-train CNN model, particularly when dealing with a limited dataset.

2.4. Dataset

Data was gathered from two distinct sources for this study: primary data was obtained from Zakouma National Park in Chad, while secondary data was sourced via Kaggle. The datasets were merged in order to enhance the outcomes. The dataset used in the experiment consisted of 18 distinct classifications, each including a varied amount of images. The selection of this heterogeneous dataset was made with the intention of capturing the complexities inherent in image classification problems encountered in real-world scenarios.

Primary data refers to data that is gathered firsthand by the researcher, whereas secondary data refers to information that has been previously obtained by another individual or entity. The main dataset utilized in this research included of visual representations of fauna within Zakouma National Park, along by pertinent details pertaining to the species, age, gender, and geographical coordinates of the animals. The secondary dataset included of animal images sourced from Kaggle.

Through the integration of primary and secondary data, we successfully generated a dataset that exhibited more comprehensiveness and enhanced representativeness of the actual world. The utilization of primary data provided us with a distinct vantage point about the fauna within Zakouma National Park, but the incorporation of secondary data afforded us a more comprehensive outlook on the global animal variety.

The mixed dataset was utilized to train a machine learning model for the purpose of classifying images of various animals. The model demonstrated a notable level of accuracy when evaluated on a separate test dataset, indicating its potential to effectively generalize to unseen data.

It is posited that our work possesses the capacity to generate a substantial impact on the domain of image classification. The mixed dataset at our disposal serves as a great resource for academics who seek to create and assess image categorization techniques. Furthermore, our research indicates that the integration of primary and secondary data holds promise as a viable approach for enhancing the efficacy of machine learning models.

Fig 4. Image of Elephants from Zakouma National Park

Fig 5. Image of Giraffe from Zakouma National Park

Fig 6. Image of Lion from Zakouma National Park

3. Methodology

3.1.1. Deep Learning Networks

The convolutional neural network (CNN) was investigated as a deep learning technique within the context of an image recognition framework. Convolutional Neural Networks (CNNs) typically consist of three types of layers: convolutional (CONV) layers, subsampling or pooling layers, and one or more fully connected (FC) layers at the end of the network [16]. Convolutional layers employ a specific type of linear operation known as "convolution" [17] between the pixel values of the input image and the small trainable filters referred to as "kernels". A multidimensional matrix with dimensions of width, height, and depth is used to represent an RGB image [18]. Kernels are augmented with a weight that is randomly initialized and can be learned. These kernels perform an element-wise matrix multiplication, which produces feature maps that are relevant and have dimension [19]. The nonlinear activation function ReLU (rectified linear unit) is applied to each feature map in order to prevent the loss of significant information [20]. This is achieved by turning negative inputs into actual zero values [16]. The pooling procedure reduces the dimensions of the feature map, leading to an increase in the level of tolerance towards translation invariance in the presence of image distortion [19]. The decrease in input size has the effect of reducing both computational complexity and memory use [19]. Additionally, it aids in mitigating the tendency for overfitting [19]. Several fully connected (FC) layers are responsible for mapping the features obtained from the preceding convolutional and subsampling layers [21], thereby preparing them for the purpose of output prediction. Deep Convolutional Neural Networks (DCNNs) exhibit a hierarchical structure in which lower hidden layers are responsible for encoding low-level structures such as line segments, orientations, and edges. Intermediate layers in the network learn abstract elements such as squares and circles. The higher hidden layers, along with the output layer, integrate these discriminative characteristics and generate a probability distribution for the assigned class labels [16].

In this study, we have devised and evaluated a total of five Convolutional Neural Network (CNN) frameworks. These frameworks include a self-trained framework denoted as CNN-1, as well as four transfer learning frameworks that utilize pretrained models such as VGG19, GoogLeNet (InceptionV3), ResNet50, and DenseNet121. During the training process, the selection of image augmentation techniques and trainable parameters was carefully conducted in order to optimize the accuracy of each model. As previously stated, our objective was to investigate the influence of different design configurations and factors on the behavior and performance level of the model.

3.1.2. Self-Trained Model(CNN-1)

The CNN architecture employed in our study comprises four convolutional layers, each of which is then followed by batch normalization and max-pooling layers. The integration of batch normalization serves the purpose of stabilizing and expediting the training process, as well as enhancing the generalization power of the model [22]. Dropout layers are intentionally positioned subsequent to the fully linked layers in order to mitigate overfitting and enhance the model's resilience. The output layer of the neural network consists of 18 neurons, which aligns with the requirements of our multiclass classification objective. The SoftMax activation function is employed in order to enhance the estimation of class probabilities [23].This architectural design integrates core ideas of Convolutional Neural Networks (CNNs) with established methodologies, including batch normalization and dropout, in order to enhance the stability of model training, mitigate overfitting, and enhance the performance of classification tasks.

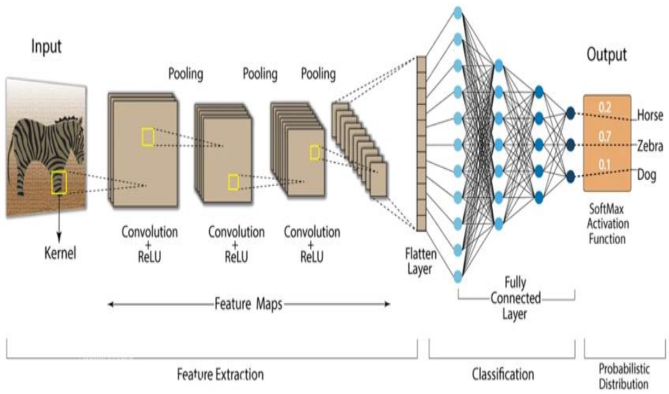

Fig: 7. CNN Architecture Model

3.1.3. VGG19Model

Oxford's Visual Geometry Group (VGG) created VGG19, a CNN architecture. In 2014, Karen Simonyan and Andrew Zisserman released "Very Deep Convolutional Networks for Large-Scale Image Recognition" to introduce it. VGG19 is a 19-layer deep CNN with 16 convolutional and 3 fully linked layers [24]. It is trained on ImageNet, which has over 1 million images and 1000 item classifications. VGG19 excels at image classification challenges like the ImageNet Large Scale Visual Recognition Challenge. Other computer vision tasks like object detection and segmentation use it extensively [25]. VGG19 extracts picture characteristics using convolutional layers. Different convolutional layers learn to extract edges, corners, and textures. The fully connected layers learn to classify images into object categories using the convolutional layers' characteristics. VGG19 is a sophisticated CNN architecture but computationally expensive to train and implement. Online pre-trained VGG19 models can be fine-tuned for specific needs [26].

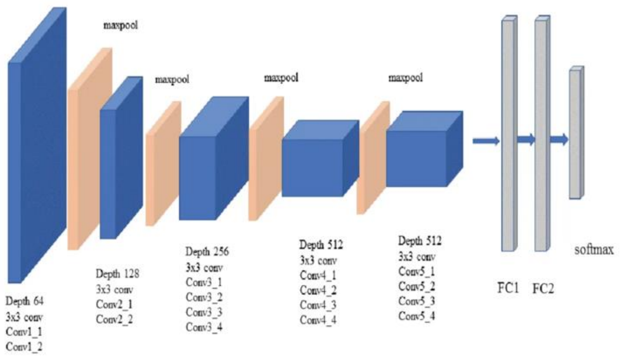

Fig 8: VGG19 Architecture model

3.1.4 GoogleNet Model

GoogLeNet, developed by Google, is an advanced deep neural network architecture. The distinguishing characteristic of this particular model lies in its innovative Inception modules, which consist of parallel convolutional layers including different filter sizes[27]. These modules facilitate the network in effectively capturing properties across various scales.The GoogLeNet architecture is characterized by its computational efficiency and parameter reduction achieved through the utilization of global average pooling[28]. The system is renowned for its exceptional precision and incorporates auxiliary classifiers to address training difficulties

3.1.5 DenseNet Model

The DenseNet architecture is a type of convolutional neural network (CNN) that establishes connections between each layer and all preceding layers within the same block[29]. In contrast to conventional convolutional neural network (CNN) architectures, which exhibit a sequential connection of layers.The substantial connectivity of DenseNet promotes the flow of gradients, facilitates feature reuse, and facilitates the learning of discriminative features[30]. DenseNet demonstrates efficacy in scenarios characterized by restricted data availability or where the control of model size is of utmost importance, as it effectively addresses the issue of vanishing gradient difficulties. DenseNet is particularly advantageous in the context of deep learning endeavors that strive to enhance feature representation [31].

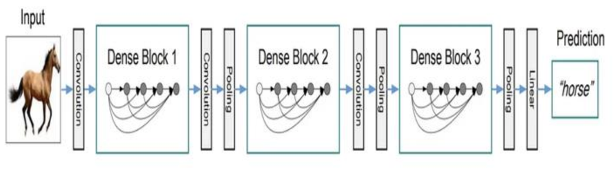

Fig: 9 Densest Architecture Model

3.1.6 ResNet50 Model

The concept of residual neural networks (ResNets) was initially introduced by He et al. in 2015 [32]. Residual Networks (ResNets) are a kind of deep neural networks that have exceptional performance and exhibit remarkable convergence properties [33]. According to recent research, it has been discovered that deeper networks may encounter a phenomenon known as the vanishing gradient problem. This problem manifests as a saturation of accuracy, leading to a subsequent degradation in performance as the depth of the network grows [32]. This problem can provide challenges in the training process of the network and may result in overfitting, a phenomenon where the network becomes excessively specialized in the training data and fails to effectively generalize to novel input[34].

Residual Neural Networks (ResNets) employ skip connections as a means to mitigate the issue of disappearing gradients. Skip connections facilitate the unimpeded transmission of information from the input or preceding layers to subsequent layers situated at greater depths within the neural network architecture. This mechanism aids in the propagation of gradients over the whole network, even in networks with a large number of layers.Skip connections, also known as identity mappings, involve the direct addition of the input from the preceding layer to the output of another layer. This enables the neural network to acquire residual functions, which represent the discrepancy between the input and the intended output. Residual Neural Networks (ResNets) have demonstrated remarkable efficacy across a range of applications, encompassing image classification, object detection, and picture segmentation.

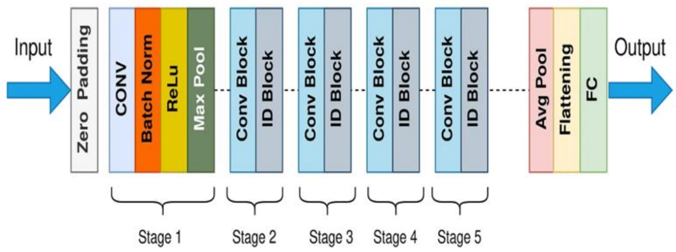

Fig: 10. ResNet50 Architecture

3.2. Image Classification Framework

The process of image classification infrastructure encompasses several steps, including preprocessing the input dataset, training and validating the deep learning architecture, and evaluating the final model using unseen test data, as depicted in Figure 7. The data that is accessible is partitioned into three distinct sets: the training set, the validation set, and the test set[35]. The Convolutional Neural Network (CNN) model acquires significant characteristics through the process of training, wherein it extracts pertinent information from the provided training data[36]. The validation dataset is employed for the purpose of evaluating the model's performance and optimizing the hyperparameters in order to attain the most optimal fit. In the process of forward propagation, the model quantifies the disparity between the observed and projected labels. During the backpropagation process, the model iteratively adjusts its learnable parameters, namely weights and biases, in order to minimize the loss function and enhance the overall performance of the model.

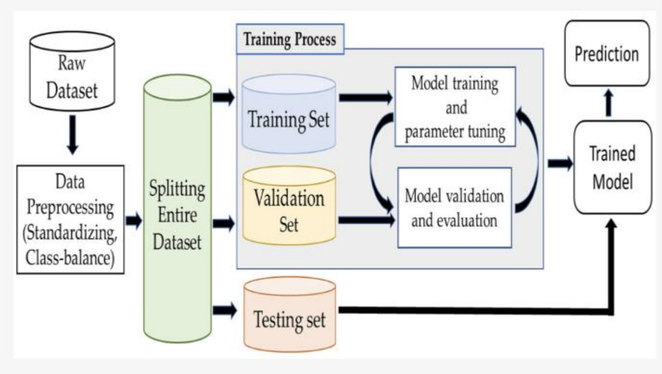

Fig 11: The workflow of the image classification pipeline involves preprocessing the raw dataset, training and validating architecture, and testing the final model with different sets of samples.

The model underwent iterative training through the adjustment of many parameters, including the number of layers, filter size, learning rate, epoch size, and batch size. Once adequate accuracy was achieved through many training and validation iterations, the model that performed the best was chosen as the final model. This final model was then assessed using test data that had not been previously seen.

3.3. Data Processing

The efficacy of machine learning models, particularly deep convolutional neural networks (CNNs), is greatly influenced by the effectiveness of data preprocessing techniques. This section outlines the preparation techniques employed to enhance data quality, mitigate noise, and assure compatibility with the selected convolutional neural network (CNN) designs [37].

3.3.1. Image Resizing

The initial stage of data preprocessing for Convolutional Neural Networks (CNNs) involves the normalization of image dimensions. For the sake of this study, all images within the dataset were subjected to resizing, resulting in a uniform width and height of 244 pixels. This process guarantees that all input images possess a consistent aspect ratio, a prerequisite for their utilization in convolutional neural network (CNN) models. Consequently, this enables the models to acquire spatial hierarchies and characteristics efficiently through the imposition of consistent dimensions.

3.3.2. Data Augmentation

The implementation of data augmentation serves as a crucial technique in enhancing the dataset's robustness and enhancing the model's ability to generalize. There are several compelling justifications for this, one of which being the mitigation of overfitting [38]. Overfitting is a phenomenon in machine learning where a model becomes excessively specialized in the training data, resulting in a lack of ability to effectively generalize to unseen data. Data augmentation is a technique that can mitigate the issue of overfitting by introducing diversity into the training data.

The dataset utilized in our study exhibited class imbalance concerns, prompting us to employ data augmentation techniques in an attempt to alleviate this issue. Class imbalance occurs when there is an unequal distribution of samples among different classes within a given dataset. The presence of imbalanced class distributions might pose challenges for machine learning models in effectively classifying minority classes. Augmentation techniques were employed to generate additional instances of underrepresented classes, hence contributing to the enhancement of the model's accuracy.

3.3.3 Pixel Normalization

In order to achieve uniformity in pixel values and maintain numerical stability during the process of model training, we performed dataset normalization by dividing each individual pixel value by 255. Normalization is a technique that mitigates the presence of sharp gradients and promotes equal contribution of each feature in the learning process, hence facilitating smoother convergence during training [39]. Furthermore, the process of normalization allows the model to effectively adapt to different lighting conditions and variations in color gradients present in the images.

3.3.4 Image Denoising

In our data pretreatment pipeline for image denoising, we employ a Gaussian blur technique with a dynamically derived standard deviation. This standard deviation is determined using the std () function from the Python NumPy library. This denoising technique aids in mitigating the adverse effects of noise present in the dataset. The application of Gaussian blur serves to effectively mitigate the presence of high-frequency noise in images, hence facilitating the preservation of crucial features[40]. This stage proves to be highly beneficial in the context of working with images that exhibit varying degrees of noise, particularly those acquired from real-world scenarios. The utilization of Gaussian blur as a denoising technique enhances the resilience of our models by mitigating the influence of noise present in the dataset, hence yielding classification outcomes that are more precise and dependable[41].

3.4. Evaluation Metrics

The evaluation of a machine learning model's performance can be assessed based on its ability to accurately classify novel or unseen data that originates from the same underlying distribution as the training samples. The training and validation curve, also known as the learning curve, illustrates the progressive enhancement of the model's performance throughout the training process. The test outcome is determined through the use of statistical measures, including accuracy, precision, recall, and F1-score. These metrics rely on the identification of true positives (TP), false negatives (FN), false positives (FP), and true negatives (TN).

$$\text{Accuracy} = \frac{\text{True Positive (TP)} + \text{True Negative (TN)}}{\text{Total number of data samples}} \quad (1)$$

The precision score determines the ratio of correctly identified target group images to all the images predicted as that particulargroup:

$$\text{Precision} = \frac{\text{True Positive (TP)}}{\text{True Positive (TP)} + \text{False Positive (FP)}} \quad (2)$$

Recall calculates the ratio of correctly identified target group images to all the images of that target group in the test data:

$$\text{Recall} = \frac{\text{True Positive (TP)}}{\text{True Positive (TP)} + \text{True Negative (FN)}} \quad (3)$$

F1-score represents a weighted average of precision and recall:

$$\text{F1} = 2 \times \frac{(\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} \quad (4)$$

4. Results and Discussion

In order to assess the performance of our model, we employed a dataset consisting of 1,100 test images that were evenly divided among 18 distinct classes within our study. In order to maintain equal evaluation of all classes, a random assignment was conducted, ensuring that each class received a nearly identical amount of test images. This methodology guarantees a fair assessment of the model's performance across all categories, hence enhancing the reliability and objectivity of the evaluation. Figure 1 presents an estimation of the distribution of images among several classifications.

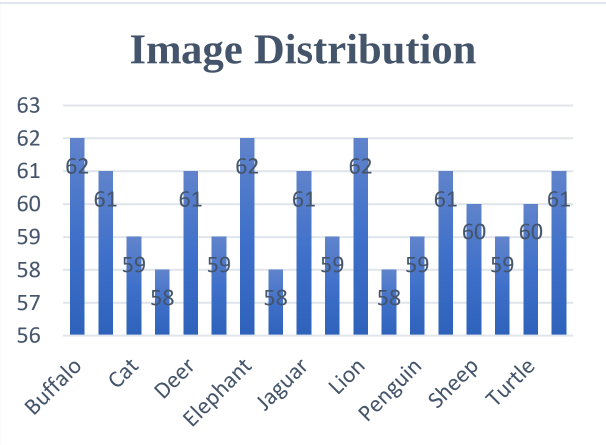

Fig12: Image Distribution

4.1 Model Metric

In this study, we provide the results and performance assessment of a machine learning classification job utilising VGG19, GoogLeNet (InceptionV3), ResNet50, and DenseNet121 models, implemented with transfer learning techniques. The analysis focuses on evaluating the attained accuracy and other performance measures, including precision, recall, and F1-score, in order to make comparisons between the models' performances. In addition, we engage in the examination and analysis of significant discoveries and perspectives.

Table 2: Performance Evaluation of the Proposed Method

Model	Accuracy in (%)	Precision in (%)	Recall in (%)	F1 in (%)
CNN	52	51	50	52
GoogLeNet	87	84	85	86
VGG19	79	74	76	77
ResNet-50	83	80	83	82
DenseNet	81	80	82	81

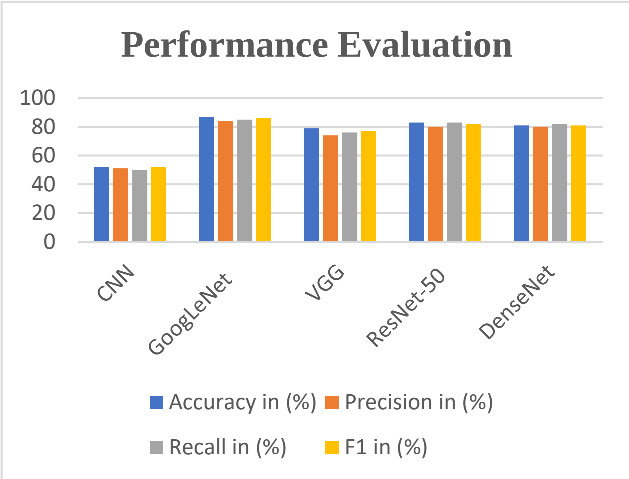

Fig 13: Graphical representation of Performance Evaluation metrics

4.2 Implications of Accuracy Result

The observed discrepancies in accuracy outcomes across different models highlight the significance of carefully choosing the most suitable architecture for a specific image categorization assignment. The GoogLeNet (InceptionV3) model demonstrates exceptional performance in capturing complicated features, rendering it the preferred option when dealing with datasets containing complex patterns. The performance of ResNet50 underscores the significance of incorporating skip connections in deep neural networks, whereas DenseNet121 exemplifies the capabilities of densely connected structures. Although VGG19 exhibits a relatively lower level of accuracy, it remains a feasible choice for certain image classification applications that do not priorities complexity as a significant factor.

4.3 Confusion Matrix

The utilization of a confusion matrix in the context of multiclass image classification tasks offers a thorough assessment of the model's efficacy in accurately detecting examples within several classes. The tracking of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) predictions for each class enables the assessment of the model's capabilities and limitations. In order to obtain a deeper understanding of the classification behaviour exhibited by the four transfer learning models, namely VGG19, GoogLeNet, ResNet50, and DenseNet121, we proceeded to create confusion matrices for each individual model.

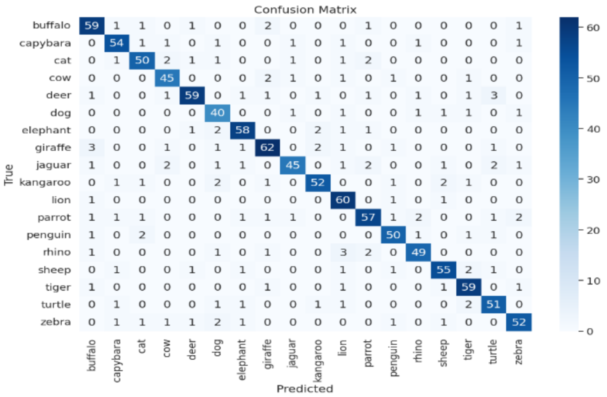

Fig14: Confusion Matrix for GoogLeNet

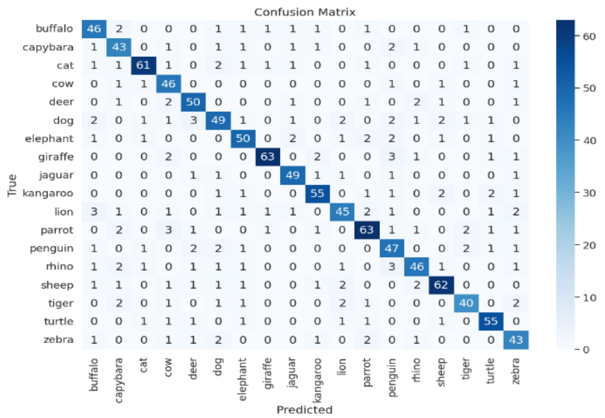

Fig15 : Confusion Matrix for ResNet-50

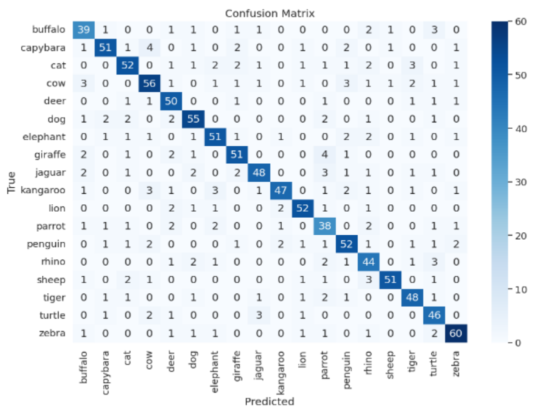

Fig16: Confusion Matrix for Densenet

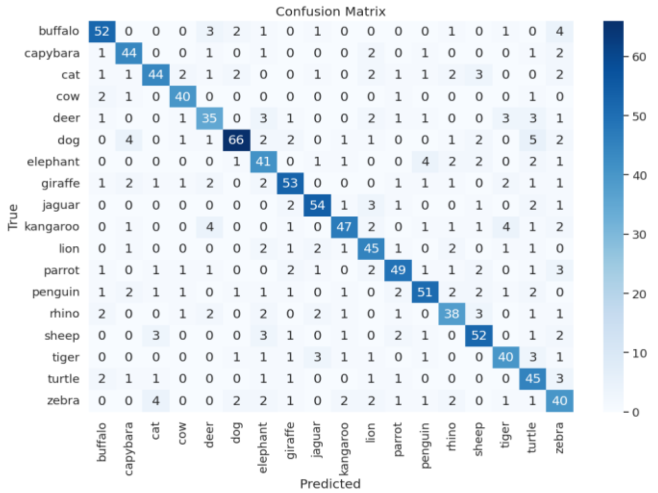

Fig 17 : Confusion Matrix for VGG19

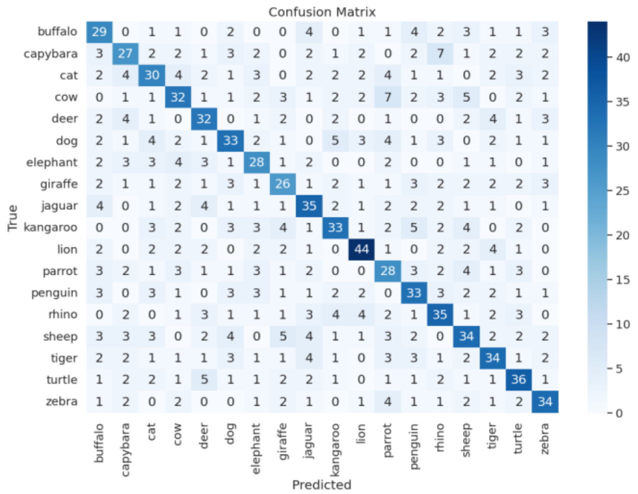

Fig18: Confusion Matrix for CNN

4.4 Accuracy Epoch Graph

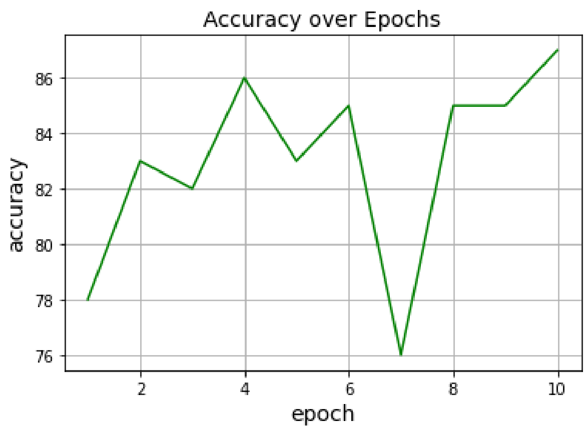

Fig 19: Accuracy over Epoch for GoogLeNet

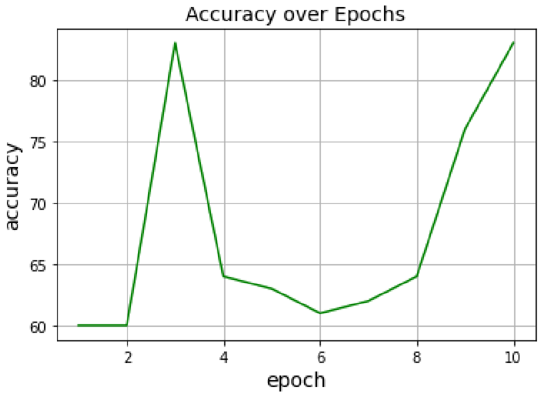

Fig 20: Accuracy over Epoch for ResNet-50

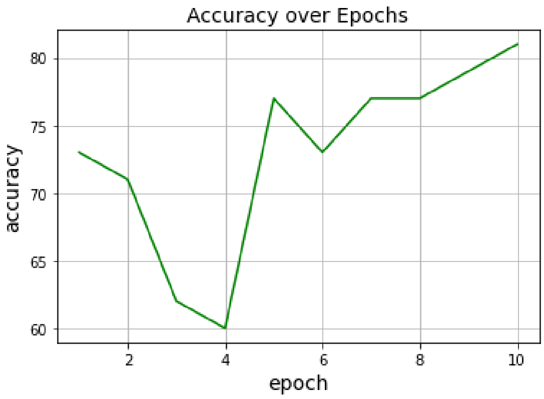

Fig 21: Accuracy over Epoch for Densenet

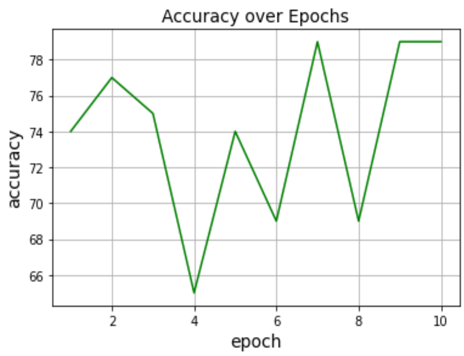

Fig 22: Accuracy over Epoch for VGG19

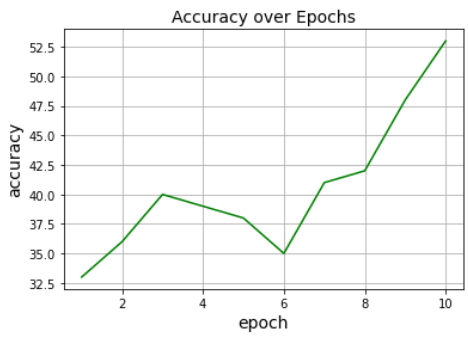

Fig 23: Accuracy over Epoch for CNN

Within this particular section, an assessment is conducted on the efficacy of a bespoke Convolutional Neural Network (CNN) model in the context of image classification. Furthermore, a comparative analysis is undertaken to juxtapose the performance of this custom CNN model against that of transfer learning approaches. The models were assessed using a dataset consisting of images of several species of animals within the Zakouma National Park. The accuracy of our customized Convolutional Neural Network (CNN) model was found to be 53%. The very low performance seen in this study can be attributed to a combination of variables, such as the scarcity of training data and the utilization of a poor model architecture.

On the other hand, transfer learning models demonstrated notable levels of accuracy, ranging from 79% to 87%. The aforementioned findings underscore the benefits of utilizing pre-trained models in the context of image categorization endeavors. Transfer learning has the potential to surpass the performance of custom convolutional neural network (CNN) models, even when trained with a limited amount of data. This characteristic renders transfer learning a highly advantageous tool for practitioners in the field of machine learning.

Transfer learning has numerous benefits for image categorization tasks, encompassing:

The utilization of transfer learning models enables the extraction of characteristics from images, thus facilitating the training of a novel classifier. This approach can prove advantageous in scenarios when the available training data is scarce, since it facilitates the reduction of trainable parameters.

Fine-tuning refers to the process of enhancing the performance of transfer learning models by training them on a novel dataset. This can prove advantageous in scenarios when the newly acquired dataset exhibits dissimilar characteristics compared to the dataset utilized for pre-training the model.

Transfer learning models have the advantage of requiring a smaller quantity of data for training compared to custom Convolutional Neural Network (CNN) models, hence enhancing data efficiency. This approach can prove advantageous in situations where the quantity of accessible training data is restricted.

In summary, the results of our study indicate that transfer learning is a great asset for machine learning professionals engaged in the development of image classification models when confronted with a scarcity of training data.

5. Conclusion

This study examines the efficacy of bespoke convolutional neural network (CNN) models and transfer learning models in the context of species classification. The proprietary convolutional neural network (CNN) model demonstrated a modest level of accuracy at 53%. In contrast, the transfer learning models, namely VGG19, GoogLeNet, DenseNet, and ResNet50, exhibited significantly higher accuracies ranging from 79% to 87%. The findings of this study provide evidence supporting the benefits of employing transfer learning in species categorization tasks, particularly in scenarios when the available training data is scarce.

Additionally, we conducted fine-tuning on the pre-trained models by unfreezing select top layers and subsequently re-training them using our species categorization dataset. The performance of the models was further enhanced, resulting in GoogLeNet achieving the greatest accuracy rate of 87%.

The results of our study indicate that the utilization of transfer learning and fine-tuning pre-trained models can yield considerable levels of accuracy in the classification of species. Nevertheless, there exist certain constraints in our research. For instance, our evaluation was limited to a small subset of deep learning architectures, and the potential for enhancing our study exists through the acquisition of additional training data. Future study should aim to investigate other deep learning architectures, ascertain the ideal parameters, and amass a larger corpus of sample data in order to get a more precise and automated solution for species recognition.

In its entirety, our study offers a significant contribution to the domain of species classification through the utilization of deep learning techniques. This study showcases the efficacy of transfer learning as a robust methodology for enhancing the efficacy of

species classification models, particularly in scenarios when the availability of training data is constrained. Additionally, we offer a methodology for optimizing the performance of pre-trained models through a process known as fine-tuning. It is our aspiration that the outcomes of this study will serve as a catalyst for further investigation into the application of deep learning techniques in the context of species categorization endeavors. Additionally, we anticipate that this will inspire the development of more precise and automated solutions in this domain.

Reference

[1] C. J. Torney *et al.*, "A comparison of deep learning and citizen science techniques for counting wildlife in aerial survey images," *Methods in Ecology and Evolution*, vol. 10, no. 6, pp. 779–787, Mar. 2019, doi: https://doi.org/10.1111/2041-210x.13165.

[2] Z. He *et al.*, "Visual Informatics Tools for Supporting Large-Scale Collaborative Wildlife Monitoring with Citizen Scientists," *IEEE Circuits and Systems Magazine*, vol. 16, no. 1, pp. 73–86, 2016, doi: https://doi.org/10.1109/MCAS.2015.2510200.

[3] [3]M. Bessone *et al.*, "Drawn out of the shadows: Surveying secretive forest species with camera trap distance sampling," *Journal of Applied Ecology*, vol. 57, no. 5, pp. 963–974, Mar. 2020, doi: https://doi.org/10.1111/1365-2664.13602.

[4] S. Islam, D. Valles, M. Forstner, and W. Stapleton, "HERPETOFAUNA SPECIES CLASSIFICATION FROM CAMERA TRAP IMAGES USING DEEP NEURAL NETWORK FOR CONSERVATION MONITORING," 2020. Accessed: Sep. 23, 2023. [Online]. Available: https://digital.library.txst.edu/server/api/core/bitstreams/a8660318-cc76-4a72-bbc6-1cbef6ecfaee/content

[5] D. J. Welbourne, A. W. Claridge, D. J. Paull, and F. Ford, "Improving Terrestrial Squamate Surveys with Camera-Trap Programming and Hardware Modifications," *Animals*, vol. 9, no. 6, p. 388, Jun. 2019, doi: https://doi.org/10.3390/ani9060388.

[6] R. Travel, "ZAKOUMA NATIONAL PARK," *Medium*, Jun. 21, 2023. https://medium.com/@responsible_travel/zakouma-national-park-1aeb6cf50548 (accessed Sep. 22, 2023).

[7] mjsnairobdev, "ADVENTURE TO ZAKOUMA NATIONAL PARK – 'AN ABUNDANT AND UNTAMED AFRICAN WILDERNESS'- WITH ORIGINS SAFARIS," *Origins Safaris*, Nov. 28, 2019. https://originsafaris.com/adventure-to-zakouma-national-park-with-origins-safaris/

[8] "Zakouma Biodiversity Conservation," *www.africanparks.org*. https://www.africanparks.org/the-parks/zakouma/biodiversity-conservation#:~:text=Buffalo%20numbers%20have%20increased%20exponentially (accessed Sep. 22, 2023).

[9] "Zakouma National Park Five-year Business Plan," 2018. Accessed: Sep. 22, 2023. [Online]. Available: https://rris.biopama.org/sites/default/files/2019-03/zakouma_business_plan2018-2022.pdf

[10] J. F. Moore *et al.*, "The potential and practice of arboreal camera trapping," *Methods in Ecology and Evolution*, vol. 12, no. 10, pp. 1768–1779, Jul. 2021, doi: https://doi.org/10.1111/2041-210x.13666.

[11] Z. He *et al.*, "Visual Informatics Tools for Supporting Large-Scale Collaborative Wildlife Monitoring with Citizen Scientists," *IEEE Circuits and Systems Magazine*, vol. 16, no. 1, pp. 73–86, 2016, doi: https://doi.org/10.1109/MCAS.2015.2510200.

[12] R. van Klink *et al.*, "Emerging technologies revolutionise insect ecology and monitoring," *Trends in Ecology & Evolution*, vol. 37, no. 10, pp. 872–885, Oct. 2022, doi: https://doi.org/10.1016/j.tree.2022.06.001.

[13] X. Yu, J. Wang, R. Kays, P. A. Jansen, T. Wang, and T. Huang, "Automated identification of animal species in camera trap images," *EURASIP Journal on Image and Video Processing*, vol. 2013, no. 1, Sep. 2013, doi: https://doi.org/10.1186/1687-5281-2013-52.

[14] J. L. Dinerman, C. J. Lowenstein, and S. H. Snyder, "Molecular mechanisms of nitric oxide regulation. Potential relevance to cardiovascular disease.," *Circulation Research*, vol. 73, no. 2, pp. 217–222, Aug. 1993, doi: https://doi.org/10.1161/01.res.73.2.217.

[15] A. Gomez Villa, A. Salazar, and F. Vargas, "Towards automatic wild animal monitoring: Identification of animal species in camera-trap images using very deep convolutional neural networks," *Ecological Informatics*, vol. 41, pp. 24–32, Sep. 2017, doi: https://doi.org/10.1016/j.ecoinf.2017.07.004.

[16] Federica Di Michele *et al.*, "Comparison of machine learning tools for damage classification: the case of L'Aquila 2009 earthquake," *Natural Hazards*, vol. 116, no. 3, pp. 3521–3546, Jan. 2023, doi: https://doi.org/10.1007/s11069-023-05822-4.

[17] A. Mathis, S. Schneider, J. Lauer, and M. W. Mathis, "A Primer on Motion Capture with Deep Learning: Principles, Pitfalls, and Perspectives," *Neuron*, vol. 108, no. 1, pp. 44–65, Oct. 2020, doi: https://doi.org/10.1016/j.neuron.2020.09.017.

[18] E. Sadeqi Azer, M. Haghir Ebrahimabadi, S. Malikić, R. Khardon, and S. C. Sahinalp, "Tumor Phylogeny Topology Inference via Deep Learning," *iScience*, vol. 23, no. 11, p. 101655, Nov. 2020, doi: https://doi.org/10.1016/j.isci.2020.101655.

[19] K. W. Ahmed, O. Chanda, N. Mohammed, and Y. Wang, "Obfuscated image classification for secure image-centric friend recommendation," *Sustainable Cities and Society*, vol. 41, pp. 940–948, Aug. 2018, doi: https://doi.org/10.1016/j.scs.2017.10.001.

[20] D. Liu, "A Practical Guide to ReLU," *Medium*, Nov. 30, 2017. Available: https://medium.com/@danqing/a-practical-guide-to-relu-b83ca804f1f7

[21] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights into Imaging*, vol. 9, no. 4, pp. 611–629, Jun. 2018, doi: https://doi.org/10.1007/s13244-018-0639-9.

[22] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *Journal of Big Data*, vol. 6, no. 1, Jul. 2019, doi: https://doi.org/10.1186/s40537-019-0197-0.

[23] F. Wang, W. Liu, H. Liu, and J. Cheng, "Additive Margin Softmax for Face Verification," *IEEE Signal Processing Letters*, vol. 25, no. 7, pp. 926–930, Jul. 2018, doi: https://doi.org/10.1109/LSP.2018.2822810.

[24] M. Shaha and M. Pawar, "Transfer Learning for Image Classification," *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, Mar. 2018, doi: https://doi.org/10.1109/iceca.2018.8474802.

[25] Animesh Seemendra, R. Singh, and S. Singh, "Breast Cancer Classification Using Transfer Learning," *Lecture notes in electrical engineering*, pp. 425–436, Nov. 2020, doi: https://doi.org/10.1007/978-981-15-7804-5_32.

[26] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv:1409.1556 [cs]*, Apr. 2015, Available: http://arxiv.org/abs/1409.1556

[27] L. Yang *et al.*, "GoogLeNet based on residual network and attention mechanism identification of rice leaf diseases," *Computers and Electronics in Agriculture*, vol. 204, pp. 107543–107543, Jan. 2023, doi: https://doi.org/10.1016/j.compag.2022.107543.

[28] J. Chen, J. Chen, D. Zhang, Y. Sun, and Y. A. Nanehkaran, "Using deep transfer learning for image-based plant disease identification," *Computers and Electronics in Agriculture*, vol. 173, p. 105393, Jun. 2020, doi: https://doi.org/10.1016/j.compag.2020.105393.

[29] C. Zhang, P. Benz, A. Karjauv, and I. S. Kweon, "Universal Adversarial Perturbations Through the Lens of Deep Steganography: Towards a Fourier Perspective," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 4, pp. 3296–3304, May 2021, doi: https://doi.org/10.1609/aaai.v35i4.16441.

[30] R. Muhammad, M. Mahamat Boukar, S. Adeshina, and S. Dane, "Deep Learning-Based OCT for Epilepsy: A Review," *Journal of Research in Medical and Dental Science* |, vol. 10, no. 8, 2022, Accessed: Sep. 23, 2023. [Online]. Available: https://www.jrmds.in/articles/deep-learningbased-oct-for-epilepsy-a-review.pdf

[31] U. B. Abubakar, M. M. Boukar, S. Adeshina, and S. Dane, "Transfer Learning Model Training Time Comparison for Osteoporosis Classification on Knee Radiograph of RGB and Grayscale Images," *WSEAS*

TRANSACTIONS ON ELECTRONICS, vol. 13, pp. 45–51, Sep. 2022, doi: https://doi.org/10.37394/232017.2022.13.7.

[32] "Report of the 1995 World Health Organization/International Society and Federation of Cardiology Task Force on the Definition and Classification of Cardiomyopathies," *Circulation*, vol. 93, no. 5, pp. 841–842, Mar. 1996, doi: https://doi.org/10.1161/01.cir.93.5.841.

[33] J. Astrup, B. K. Siesjö, and L. Symon, "Thresholds in cerebral ischemia - the ischemic penumbra.," *Stroke*, vol. 12, no. 6, pp. 723–725, Nov. 1981, doi: https://doi.org/10.1161/01.str.12.6.723.

[34] Assia Aboubakar Mahamat *et al.*, "Machine Learning Approaches for Prediction of the Compressive Strength of Alkali Activated Termite Mound Soil," *Applied sciences*, vol. 11, no. 11, pp. 4754–4754, May 2021, doi: https://doi.org/10.3390/app11114754.

[35] S. Loussaief and A. Abdelkrim, "Machine learning framework for image classification," *2016 7th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT)*, Dec. 2016, doi: https://doi.org/10.1109/setit.2016.7939841.

[36] W. Rawat and Z. Wang, "Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review," *Neural Computation*, vol. 29, no. 9, pp. 2352–2449, Sep. 2017, doi: https://doi.org/10.1162/neco_a_00990.

[37] Q. Bai, "Big Data Research: Database and Computing," *Journal of Big Data Research*, vol. 1, no. 1, pp. 1–4, Apr. 2018, doi: https://doi.org/10.14302/issn.2768-0207.jbr-17-1925.

[38] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, May 2012, doi: https://doi.org/10.1145/3065386.

[39] Umar Adam Ibrahim, Moussa Mahamat Boukar, and M. A. Suleiman, "Development of Hausa Acoustic Model for Speech Recognition," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 5, Jan. 2022, doi: https://doi.org/10.14569/ijacsa.2022.0130559.

[40] Aishat Salau, N. Agwu, and Moussa Prof. Moussa, "Deep Learning for Fraud Prediction in Preauthorization for Health Insurance," *International journal of engineering and advanced technology*, vol. 12, no. 2, pp. 75–81, Dec. 2022, doi: https://doi.org/10.35940/ijeat.b3915.1212222.

[41] G. George, S. Adeshina, and M. M. Boukar, "Development of Android Application for Facial Age Group Classification Using TensorFlow Lite," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 11, no. 4, pp. 11–17, Sep. 2023, Accessed: Sep. 24, 2023. [Online]. Available: https://ijisae.org/index.php/IJISAE/article/view/3449

[42] Badiy, M. ., & Amounas, F. . (2023). Embedding-based Method for the Supervised Link Prediction in Social Networks . International Journal on Recent and Innovation Trends in Computing and Communication, 11(3), 105–116. https://doi.org/10.17762/ijritcc.v11i3.6327

[43] Esposito, M., Kowalska, A., Hansen, A., Rodríguez, M., & Santos, M. Optimizing Resource Allocation in Engineering Management with Machine Learning. Kuwait Journal of Machine Learning, 1(2). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/115

[44] Kathole, A. B., Katti, J., Dhabliya, D., Deshpande, V., Rajawat, A. S., Goyal, S. B., Suciu, G. (2022). Energy-aware UAV based on blockchain model using IoE application in 6G network-driven cybertwin. Energies, 15(21) doi:10.3390/en15218304 Kawale, S., Dhabliya, D., & Yenurkar, G. (2022).