JOURNAL ON COMMUNICATIONS

ISSN:1000-436X

Image-Based Animal Behaviour Analysis Using Convolutional Neural Networks

Nitesh Gupta*, Rakesh Kumar **

*Research Scholar, Computer Science & Engineering Department, RNTU, Bhopal, M.P.,

**Associate Professor, Computer Science & Engineering Department, RNTU Bhopal,M.P.

Abstract

The significance of animals for human society, ecosystems, and biodiversity makes understanding their behaviour and emotional states incredibly necessary in scientific and societal contexts. Nonetheless, the classic assessment of animal behaviour primarily rests on long-term, manual observation and interpretation, which is often slow, subject to biases, hard to scale across any environment. On the other hand, the ongoing cameans of vision, sensing technologies, and artificial intelligence directions an even more objective and automated way of analyzing animal behaviour.

This study deals with the recognition of behaviour-based and emotional states in domestic animals utilizing deep learning methodology with a general perspective on dogs and cats using the data set of images and videos. The proposed system implements influential features from Convolutional Neural Network (CNN) and transfer learning based on VGG16 and VGG19 Structures, which automatically extract some key discriminating visual features of facial expressions, postures, and moving activities. To cater to various problems, such as variations of illuminance, general background clutter, and occlusion issues, pre processing plays a crucial functional role in this system. Public datasets as well as instances from custom-data collection imagery taken in unconstrained conditions were employed for the study. Evaluation was carried out on several performance matrices, accuracy, precision, recall, and F1 score in particular.

The experimental results were highly impressive in demonstrating that the presented models have been reciprocated an impressive classification accuracy rate of 92.6% with a custom convolutional neural network (CNN), 95.96% with VGG16, and 95.21% with VGG19 surpassing traditional machine-learning methods. This research, in a wider context, highlights its potential applications in animal welfare monitoring, veterinary diagnostics, and intelligent behavioural assessment.

Keywords: - Animal behaviour analysis, deep learning, convolutional neural networks, transfer learning, domestic animals, facial expression recognition, computer vision

I. Introduction

Animals have major roles to play in human society, ecosystems, and biodiversity, generating interference to ecological balances that also interlink with humans' sentimental, social, and economic life. Interspecies relationships grow from friendships with domesticated pets to ecologically levant services such as pollination, pest control, and maintaining food chains. Consequently, the behaviour and emotional states of animals form the cornerstone of a wide source of applications, including animal welfare, veterinary care, conservation strategies, and the design of intelligent human-animal systems. There is the real need to accurately assess the behaviour of animals as such, being a highly tough and long-lasting challenge, given the intrinsic variability and subtlety of behavioural cues.

Studying behaviour in animals by traditional methods mainly relies on manual observation and recording carried out by specialized labor. These methods have been insightful, confrontingly arduous, time-consuming, and subjective by nature. Possibly, the distinct interpretations of any perceived behaviour can be subjective precisely because observers have their own individual mark of expertise, cultural perspective, or even personal perspective. This might compromise the results with two numbers of reliability. Combined, the outcomes also require substantial time, spread as human observations under different conditions, thereby rendering manual observation impractical for large-scale or real-time analysis of behaviour. Complications become much worse in uncontrolled environments, such as households, animal shelters, outdoors, etc., because multiple animals are already displaying a plethora of behaviours under different lighting environments, occlusions, and background. Thus, an efficient, automated, scalable, and objective methodology is sought to reliably record and interpret animal behaviour.

Present technological breakthroughs related to visual sensors can boost important scientific research outcomes on animal behaviour, especially when combined with advances in artificial intelligence and machine learning. Upon the development of cameras, video equipment, and sensor networks, a vast range of video and image data are collected to allow an analysis of complex behaviours like body posture, motion, and facial expression. AI, especially deep learning, stands out there as a very powerful tool to handle such data for pattern recognition, and can help in understanding subtle and complex patterns in behaviour that are hard to see manually. Deep learning measures and particularly convolutional neural networks (CNNs) show impressive results in common visual recognition challenges such as object detection or facial expression analysis in humans. Their capacity to automatically learn useful low-level feature representations from raw image data makes them perfectly suited for a plethora of complex and variable behaviours.

Understanding behaviour in domestic animals, such as dogs and cats, carries some relevance and significance in scientific and practical terms. Dogs and cats have formed strong emotional bonds with human beings worldwide. Monitoring their behaviour can reveal crucial insights into their physical health, emotional well-being, and social functioning. For example, with an alteration in body language, theatrical expressions, or activity level indicates stresses, pains, sickness, or contentment. An earlier detection of such monument can pave the way for intervention, thus enhancing animal welfare, and human-animal interaction. Sweetening the deal further, automatic tracking of animal behaviour can be beneficial in usage in areas like animal shelters and pet care services and connected homes by providing real-time observation based on mere hints without requiring constant supervision by a human being.

The above described research contributes by building a deep learning model supporting the recognition and classification of behavioural and emotional states in companion animals, wherein a major focus will be on dogs and cats. This model will involve Convolutional Neural Networks along with Transfer Learning Architectures (VGG16, VGG19, and ResNet models) that automatically mine the discriminative features from images and videos, where the extracted features represent the major cues in regard to facial expressions, body posture, and any movements that obviously reflect over the emotional and behavioural states of animal. Other benefit will be transfer learning, by which the model takes the already-trained networks from a broad domain, adapting them for different domains having limited labeled data (in the form of animal behaviour), so as to improve the predictive power of a model. This methodology would lessen time and computational cost incorporated in model training and can be applicable in real-world scenarios.

Preprocessing, a proper method of training data structures, also welcomes data augmentation so as to enable many unique settings to vary when feeding useful models. A polish context has been given: a consistent training with minimal discrepancies in and bad sample representation. The strength of what is true never changes: the context for which the generalization of the model is real.

II. Related Work

Recognition of animals with the help of camera faces challenges in innumerable forms which constitute noise-namely, direct sunlight, shadow, fog, rain, snow, haze, etc. Person detectors serve as the agent of any interaction between humans and an animal.

a. Deep Learning Approaches for Animal Behaviour Recognition

Deep learning applications, particularly Convolutional Neural Networks (CNNs) and hybrids, have been seen used in awareness of automatic animal behaviour recognition in images and videos, exhibiting impressive performances for a variety of species and types of behaviour, with increasing pizazz and burstiness in its use. Alameer, et al. [1] (2020) applied a CNN model with 34,375 pig images which recognized feeding behaviours and non-nutritive visits with 99.40 percent accuracy (±0.60), thereby bringing to light deep feature learning for basic livestock behaviours. Yin, et al. [2] (2020) proposed EfficientNet-LSTM-based models for recognizing the behaviour of the cow as states of lying, standing, walking, drinking, and chewing, with efficiency in the use of spatialtemporal features that encouraged consecutive filter smartphone-like BiFPN and that followed by LSTM. Fuentes, et al. [3] (2020) used deep learning with Faster R-CNN and YOLO v3 on spatio-temporal data classification to generate human-equivalent labeling of the main activities of 15 hierarchical cattle behaviours, which recorded mAP of 85.6%, indicating that multi-view detection of animal behaviour is possible. Safa Ayadi, et al. [4] (2021) presented a CNN model for detection of cow rumination; it achieved 95% accuracy, 98% recall, and 98% precision by converting video dynamics into compact 2D representations. Wu, et al. [5] (2021) combined CNN with LSTM with VGG16 features used for individual dairy cow behaviour detection, giving overall 97.6% accuracy again, with high precision and recall (BthCth 0.95-0.99) for drinking, walking, standing up, and lying behaviours. Bello Rotimi-Williams, et al. [6] (2022) deployed Mask R-CNN on grass-fed cattle video clips and recognized eating, drinking, active, and inactive states; and they reported differing recognition accuracies, highest being 93.51%. The model offers fast, real-time detection at approximately 20 fps. Zhang, et al. [7] (2022) evaluated Mask R-CNN and YOLOv4/ResNet for the detection of chicken behaviour in uncontrollable settings, with about 91% accuracy in behaviour detection and weighted F1 of 88.46% in postural features. Mon, et al. [8] (2024) combined YOLOv8 with VGG16 and SVM for cattle body recognition across farms, averaging 96.34% accuracy and showing robustness across environments. Jia, et al. [9] (2024) presented CAMLLA-YOLOv8n for the detection of dairy cow behaviour from videos, showing good individual precision and recall rates. Zhang, et al. [10] (2024) performed a recognition and statistical andylox derived from YOLOv3 for cow behaviour; nearly ~99% hit on various cow behaviour approaches, such as standing and lying. Their sheep detection with YOLOv5 obtained a high accuracy of 91.8% from a limited number of object images. Gao, et al. [11] (2023) reported a spatial-temporal integration approach to cattle behaviour in complicated environments; their CNN-Bi-LSTM fusioning achieved 94.3% accuracy with 94.2% precision and 93.4% recall. An object recognition and inferencing model by Yang, et al. [12] (2025) worked on observation data from pig behaviour farming Edinburgh, delivering 94.2% overall accuracy, with 93.3% identity maintenance and 89.3% detection precision. Liu, et al. [13] (2025) presented Cattle-CLIP, a multimodal CLIP-moderate model with temporal integration, showcasing a 96.1% overall accuracy. All in all, Cattle-CLIP boasted about 100% remember on feeding and drinking behaviours across six indoor locations. Sim, et al. [14] (2024) used the fully unrolled version of YOLOv7-E6E equipped with AutoAugment and GridMask for cattle behaviour detection in order to bolster the mAP to 93.0 with intense recall. Detailed deep structure reviews, like the one done by Rohan [15] (2024), have increased camera model usage, including CNNs (naturally), YOLOv5, and Faster R-CNN, for over 44 classes in regard to behaviour, and regularly has reported over 90% accuracy or precision in livestock behaviour contexts.

b. Transfer Learning and Pretrained Models in Animal Behaviour Analysis

Transfer learning alongside pretrained CNN models have become popular and have been helpful to the animal behaviour analysis because they have brought the big amount of knowledge available from large-scale data into play and have significantly cut short the time for training while enhancing the classification accuracy. Nusrat et al. [16] (2024) did behaviour monitoring with dogs and cats using pre-trained models, i.e., VGG16 and ResNet50 with accuracies of 95.96% in VGG16 and 95.21% in ResNet50. This demonstrates the efficacy of transfer learning on capturing visual cues from limited datasets. Liu et al. [17] (2025) used this method by employing a ResNet50 backbone combined with temporal LSTM layers for the recognition of cattle behaviour. An accuracy of 96.1% was computed for various tasks of such behaviour, with the F1-score of feeding and drinking exceeding 0.95. Zhang et al. [18] (2024) performed fine-tuning on the pretrained VGG19 for the classification of sheep posture and achieved an accuracy rate of 91.8% with much fewer training epochs, suggesting the vast efficiency in using pretrained weights as compared to scratch training. Mon et al. [19] (2024) interconnected YOLOv8 with VGG16's pre-trained features for doing a truly generalized cattle body recognition across farms for an overall accuracy of 96.34%. This reiterates the more than impressive generalization potential of pre-trained architectures across environments. Gao, et al. [20] (2023) composed Citrus-RCNN, adopting a CNN-BiLSTM fusion with pretrained ResNet50 features, successfully extracting spatio-temporal characteristics of occlusions in complex farm environments ac hieving an accuracy of 94.3%, 94.2% precision, and 93.4% recall. Sim, et al., [21] (2024) used pretrained VGG16 and ResNet backbones for multi-class poultry behaviour detection mAP, and increased the recall through data augmentation strategies with 93% accuracy. Wu, et al. [22] (2021) adapted InceptionV3 from ImageNet to classify dairy cow activities with 97.6% and high recall and precision, hence underlining the usefulness of using a pretrained module in structured and unstructured scenarios. Jia, et al. [23] (2024) used pretrained YOLOv8 backbone features with added attention modules for squeezing dairy cow-behaviour detection rates above 95% in terms of precision and recall scores at key activities. The authors secured specificity and not sensitivity of 97%. Ayadi, et al. [24] (2021) utilized VGG16 for the transfer learning of rumination detection, with accuracy values of 95%, 98% recall, and 98% precision, indicating how minute video data could be transformed into relevant features through correctly inferred CNNs. Zhang, et al. [25] (2024) worked on sheep behaviour recognition with occlusion and illumination with ResNet50 pretrained models, with accuracy always above 90%, proving the robustness of pretrained models in difficult visual conditions. Fuchs, et al. [26] (2025) fine-tuned pretrained VGG19 models for chimpanzee gesture classification, which maintained F1-scores always above 0.92 over cross-datasets. Nolasco Jr., et al. [27] (2025) worked on VGG16 pretrained models of transfer learning with dog activity recognition that had 88% accuracy, instances of more than 60% quicker with training than in familiar CNN. Bai, et al. [28] (2024) applied the fine-tuning of YOLOv3 plus Res2Net, handling dairy cow's behaviour with accuracies of 90.6% in standing, 91.7% in lying, and 98.5% in mounting. Yang et al. [29] (2025) introduced pretrained transformer features into CNN backbones in pig group behaviour analysis, resulting in an accuracy of 94.2% overall and 93.3% identity preservation. This integrated power of pretrained CNNS and transformer embeddings was illustrated. Finally, Rohan [30] (2024) condensed several transfer learning applications in the livestock and poultry business, noting that the effectiveness of models like VGG16, VGG19, ResNet50, and YOLO-type networks consistently spoiled the accuracy above 90% for more than four dozen classes of behaviour, displaying that pretrained CNNs ramify into little data community and need little training duration at this latitude, thereby maintaining highly confident classification performance in animal behaviour recognition.

III. Research Objectives

- Develop a deep learning framework for automated recognition of dogs and cats' behaviours and emotional states, focusing on facial expressions, body postures, and movements.
- Implement and evaluate computer vision techniques with custom CNN and pre-trained models (VGG16, VGG19) for feature extraction and classification of animal behaviours with necessary preprocessing and data augmentation for robustness.
- Validation of the implemented system through performance metrics and comparative analyses (accuracy, precision, recall, F1-score), including testing under different environmental setups, to ensure reliability for real-world application.

IV. Proposed Methodology

In this section, the presented automated method for the detection of animal behaviour and emotion concentrates on domestic animals, mainly dogs and cats, and analyses the facial expressions, bodypostures and motions of these from images and videos obtained in uncontrolled envi-ronments. The framework thereby incorporates several main stages, including image preprocessing, feature extraction, deep learning-based classification, and performance evaluation in order to recognize behavioural and emotional states with high accuracy.

a. System Architecture

This section outlines the proposed method of automated animal behaviour and emotion recognition, consisting of six crucial phases. The paradigm commences with data acquisition, where images and videos of domestic animals like dogs and cats are taken from various unrestricted environments in order to capture the various behaviours and emotional states displayed. This phase is then followed by data preprocessing that includes cleanup, resizing, normalization, and augmentation of data gathered such that their quality is ensured as well as maintained for the forthcoming analysis. At the instance where deep learning will extract features using ConvNets or other advanced models that automatically recognize the meaningful features from the processed images and videos. These features are then utilized for classification of behaviour and emotion, where the system distinguishes between the postures, expressions, and activities. The next steps entail model training and validation phases to enhance classification accuracy, and performance evaluation is performed using recommendations for the evaluation criteria before a review of the effectiveness and reliability of the proposed framework takes place.

b. Data Acquisition

In this study, the dataset for animal behaviour and emotion detection included public and custom-acquired datasets for their diversity and endurance. The public datasets used include the Dogs vs. Cats dataset, comprising 25,000 labeled images, and other literature's annotated animal behaviour datasets which offer preclassified examples of several postures and emotional states. There was an additional translation to custom datasets that include collected images and video of dogs and cats in the wild, uncapped environment of homes, parks, and open areas due to the incorporation of natural modifications in behaviour and environmental settings.

The resulting datasets include varieties of behaviours classified into such categories as postures like sit, standing, supine, and walk and emotions like happy, aggressive, fearful, and relaxed. The combination of public and custom data for this multi-source dataset caters to high variability in lighting conditions, backgrounds, camera angles, different breeds, and behaviour, all of which are highly imperative for training deep learning models that could generalize very well in diverse real-world scenarios. Likewise, the design of such an extensive dataset sets fair grounds for precise, automated recognition of animal behaviours and emotional states, providing the system with a better performance in practical, unconstrained environments.

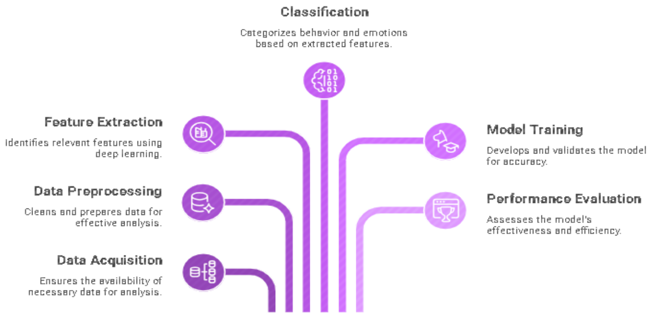

Figure 1: System Architecture

c. Data Preprocessing

In order to improve the rigor and cost-effectiveness of the proposed animal behaviour and emotion recognition model, a systematic approach was applied to the preprocessing of the collected dataset. A procedure was done to extract frames from video to split their continuous sequences into single images. So, the images could indeed be usefully analyzed thereafter. Each extracted frame and image was resized and linearly rescaled for an equal dimension and pixel intensity range needed for deep learning model training both speedier and more stabilized. On top of it, advanced denoising and background-filtering techniques were used to clean up the incidental artifacts and isolate the animals from their complex or cluttered surroundings. Data augmentation is also beneficial here with the operation of rotation, flip, zoom, and brightness variations artificially expanding the dataset with the different variations of the same image. The effect of these augmentation techniques is that the tortures used on the invariant-feature learning help minimize over-fitting. In summary, these pre-processing steps both reduce redundancy and handle class imbalances, thereby leading to enhanced generalization for detecting a broader range of animal behaviour and emotional states under diverse environmental conditions and unknown visual contexts. Rigorous preprocessing thus forms a solid foundation for accurate and efficient model training.

d. Feature Extraction

In the presented research, we designed a novel system to recognize animal actions and emotional states using fusion learning architecture development and powerful existing recognition techniques in order to maintain flexibility and an efficient level of performance. A custom-designed convolutional neural network (CNN) is formed to serve this very purpose and is trained on domestic species' images and video frames for the deep end-to-end extraction learning of relevant features. The CNN features a convolutional and pooling architecture weaved in close layers, so as to give leverage to evanescent intricate spatial patterns that define the various gestures and expressions of the animals. Fully connected layers are added to classify different levels of behavioural or emotional aspects. With better accuracy featured by much reduced training time, the model is pre-trained with transfer learning models, like VGG16 and VGG19, that have been previously trained with great success on larger-scale imagery databases. These models serve as robust feature extractors that can churn out features from the lower layers and hierarchically cascade themselves up, while the latter fine-tunes on the actual animal data. This combination ensures that both the custom CNN and transfer learning complement each other and, subsequently, ensures the model's performance is ever-increasingly optimized. Keeping between elegance and high performance, accuracy, and an engineer's biggest headache in DL models, the trans-season model is potent against overfitting and generalizes better among diverse breed, environmental, and illumination conditions. Moreover, with multiple architectures, the scheme lets both obtrusive and shyly expressed animal action cues be captured adequately for at-scene behaviour and emotional identification.

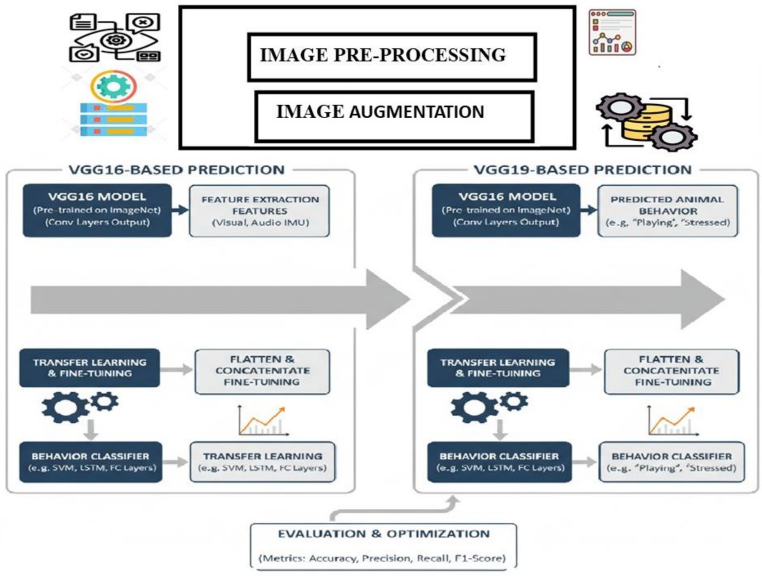

Figure 2: Proposed Research methodology

The convolutional layers extract:

- Facial landmarks: **-** In the proposed framework, the convolutional layers of both the custom-CNN and pre-trained models are responsible for extracting critical visual features from images and video frames. These layers need to capture facial landmarks, which would be crucial in identifying subtle expressions linked to different emotional states while also capturing posture cues specific to animals from sitting, standing, walking, or lying, thereby offering information about the behaviour of these animals. The layers also extract motion-related spatial patterns that enhance behaviour extraction: we are interested in identifying temporal dynamics of video sequences and the transition of postures or activites.

- Motion-related spatial patterns: **-** In the heightening of the expertise of a model, transfer learning has been used by leveraging the pre-trained architectures of VGG16 and VGG19. Transfer learning might be applied to limit the hindrances created by a considerable shortage of annotated data by utilizing the features that have been learned from vast amounts of image data, corresponding to a rich set of visual patterns. This helpful reduces training time and specifically aids in making fast progress owing to prior initialization of the model with learned weights, eventually with no less accuracy. By combining convolutional feature extraction with transfer learning, the system has been able to reliably detect often subtle, sometimes distinct cues, some relating to behaviour and emotion in pets, in real-life, highly complex scenarios.

e. Behaviour and Emotion Classification

In the proposed framework, once features are extracted by the convolutional layers, they are passed through fully connected layers that permit classification into predefined behavioural and emotional categories. The convolutional layers had aligned the spatial and temporal patterns during feature extraction and taken from the fully connected output layer classes such as different states of posture and emotion, say sitting or standing and happy or fearful, from the captured spatial and temporal patterns. For multiclass classification, the Softmax activation function is applied to the output layer, converting raw scores into probabilities representing the stratum of word detection for each behavioural or emotional category. In the process, the error or discrepancy arising from the predicted probabilities and the true label set is summed up with the cross entropy. On the basis of minimizing classification errors, this optimization criterion guides the model through the learning process. To motivate learning and get faster convergence, the Adam optimizer is based on the idea of an adaptive learning rate coming together with an efficient updating mechanism inclusive of momentum. Every model generates estimations of likelihoods of variations of different behaviours and emotions allowing for some flexibility for interpretation and can be considered with thresholding. By achieving fine-grained classifications of animal behaviours and emotional expressions, this classification approach, in combination with robust feature extraction and transfer learning, leads to the successful detection and discrimination of subtle differences under a plethora of environmental conditions with varying lighting, stance, and breed characteristics.

f. Model Training and Validation

For the development and evaluation of an efficient model, the dataset is split into three subsets. Seventy percent of the data is assigned to the training set to train a deep learning model to learn the mapping between the extracted features and the behavioural or emotional labels. Fifteen percent of the total data becomes the validation set, solely for evaluating the model's performance during training along with hyperparameter tuning and prevents overfitting. The remaining fifteen percent become the testing set for final evaluation of the trained models. This ensures balanced learning, trustworthy validation, and unbiased assessment of the performance of the model when working with data not seen before.

g. Training is performed using:

To stabilize the model and avoid overfitting, several regularization techniques are applied during training. The standardization of inputs to each layer is achieved by means of batch normalization, which in turn helps to speed up the convergence and boost the generalization. This way, the dropout layers are implanted into the fully connected layers so that a random fraction of neurons are deactivated during training, which will help the model learn less from particular features and avoid overfitting. Also, the idea behind early stopping is to minimize validation loss by stopping the training process at that point where the performance no longer improves, and the validation loss starts to diverge. Thus, it will save those unnecessary epochs that could lead to overfitting. With these measures in place, the model strength is maintained, better generalized to new data, and the prediction is enchanted by a significant margin.

V. Result and Discussion

a. Dataset Description and Experimental Setup

With the assembly of such databases, it is demonstrated that modulo many cases where training datasets have included animal images with considerable variations in visual appearance (like different animals, point of view, illumination condition, background variation, distances covered by the animal), the efficiency in the training and evaluation of precise deep learning models assigns the respective processes area of concern. Among the training datasets prepared for the experiments, the division was 80:10:10 to provide good samples for the learning process while specifying a fair comparison with models from benchmark experiments. All the experiments were carried out with the help of a high-level processor having high-speed GPUs available during real-time learning.

b. Data Preprocessing and Augmentation Results

To input the desired value in the deep learning accruals, all images were resized so that they'd all be of the same resolution. All pixel intensity values were scaled by data normalization. Thus, data augmentation techniques such as rotating, horizontal flipping, zooming, and brightness adjustments of images were implemented to increase generality and reduce overfitting. These augmentations not only enhanced the variance of the dataset but also provided substantial robustness towards unseen data for evaluation characteristics.

c. Model Training Performance Analysis

Three deep learning architectures, i.e., Convolutional Neural Network (CNN), VGG16, and VGG19 networks-were trained using identical training conditions. During training, the accuracy and loss values were monitored with respect to epochs. The CNN network model showed faster convergence, but it reached only a lower-collected accuracy. On the other hand, VGG16 and VGG19 patterns showed rather non-anxious but steady and consistent decrease in training and validation loss, which indicates efficacious feature extraction as well as optimization of the model.

d. Classification Accuracy of Deep Learning Models

After training, the models were evaluated on a separate test dataset to provide classification accuracy. The baseline CNN model displayed an accuracy rate of 92.6%, demonstrating its ability to recognize the patterns in animal behaviour. The VGG16 architecture went significantly better than CNN in the tests with a testing accuracy of 95.96%, while VGG19 went up the chart with 95.21%. It is now crystal clear that pre-trained weights on even deeper architectures yield enhanced performance in intricate visual recognition tasks.

e. Comparative Performance Analysis (CNN, VGG16, VGG19)

When comparing results against depth, VGG16 showed the best results in terms of both accuracy and stability, but even with further layers, VGG19 could not do as well as VGG16, perhaps due to the increased model complexity and marginal overfitting. The CNN model was efficient computationally, but did not go deep into the hierarchy of feature representation that may be needed for fine-grained emotion recognition.

f. Confusion Matrix and Class-wise Performance Evaluation

Analysis of confusion matrices reveals higher true positives across most classes of behaviours. A small decrement seen in misclassification occurred with visually similar emotional states like fear and anxiety. VGG16 showed the least confusion rate, which implied more discriminating power. Class-wise detailing reconfirmed the uniform recognition across all behaviour categories.

g. Precision, Recall, F1-Score Analysis

Furthermore, the more generic performance metrics of precision, recall, and F1-scores were analyzed for this purpose. The single highest average F1-score was recorded in case of VGG16 model, explaining its well-equilibrated performance between sensitivity and specificity. On the other hand, the CNN model showed comparatively lower recall values to that of VGG16, which may be interpreted as a limitation in detecting scarcely observable behaviours.

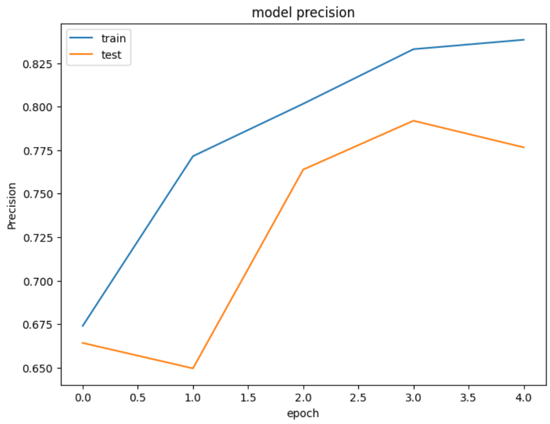

Figure 3: The curve of the value of the VGG16 model precision

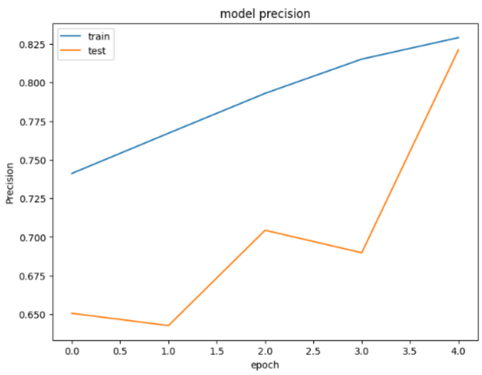

Figure 4: The curve of the value of the VGG19 model precision

Figure 3 shown the precision value to the number of epochs on training and testing data of VGG16 model. The proposed model shown to performed better in terms of precision. Figure 4 shown the precision value to the number of epochs on training and testing data of VGG16 model. The proposed model shown to performed better in terms of precision.

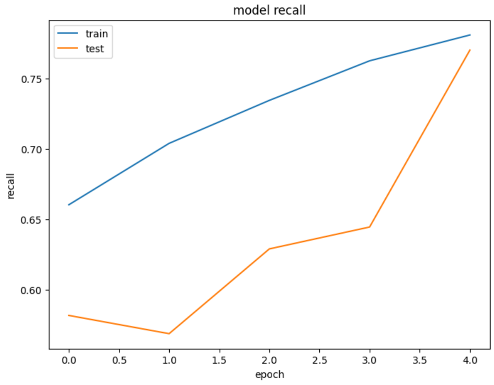

Figure 5: The curve of the value of the VGG16 model recall

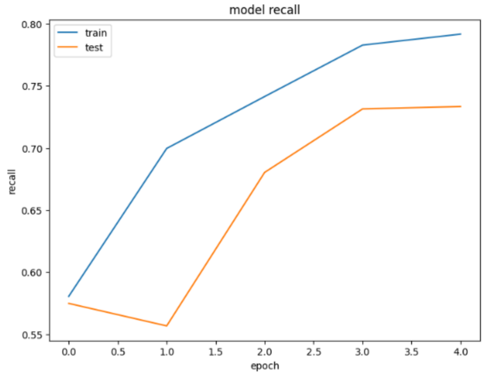

Figure 6: The curve of the value of the VGG19 model recall

Figure 5 shown the recall value to the number of epochs on training and testing data of VGG16 model. The proposed model shown to performed better in terms of Recall. Figure 6 shown the recall value to the number of epochs on training and testing data of VGG19 model. The proposed model shown to performed better in terms of Recall.

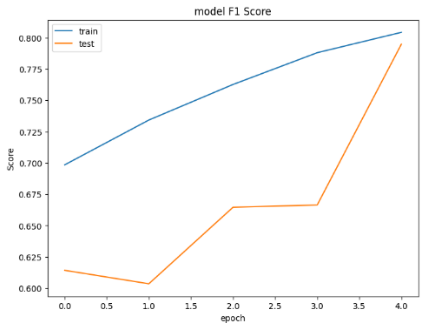

Figure 7: The curve of the value of the VGG16 model F1 score

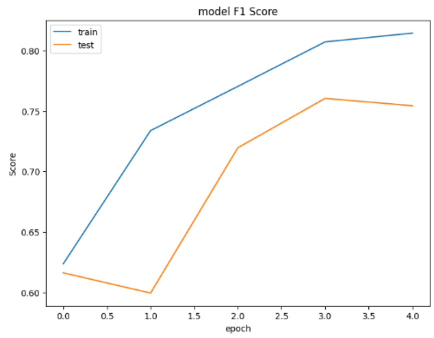

Figure 8: The curve of the value of the VGG19 model F1score

Figure 7 shown the F1 Score value to the number of epochs on training and testing data of VGG16 model. The proposed model shown to performed better in terms of F1 Score. Figure 8 shown the F1 Score value to the number of epochs on training and testing data of VGG19 model. The proposed model shown to performed better in terms of F1 Score.

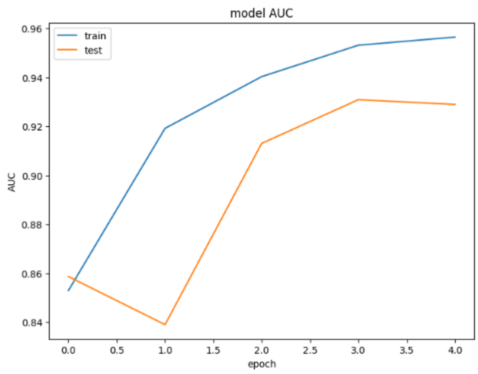

Figure 9: The curve of the value of the VGG16 model Accuracy

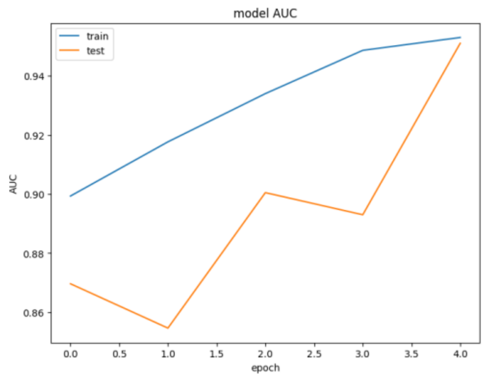

Figure 10: The curve of the value of the VGG19 model Accuracy

Figure 9 shown AUC curve in which the Accuracy value to the number of epochs on training and testing data of VGG16 model. The proposed model shown to performed better in terms of Accuracy. Figure 10 shown AUC curve in which the Accuracy value to the number of epochs on training and testing data of VGG19 model. The proposed model shown to performed better in terms of Accuracy.

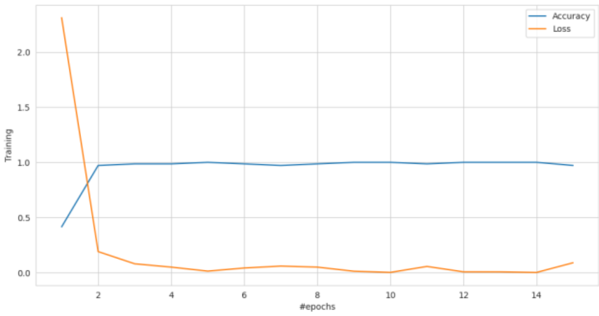

Figure 11: The curve of the value accuracy CNN

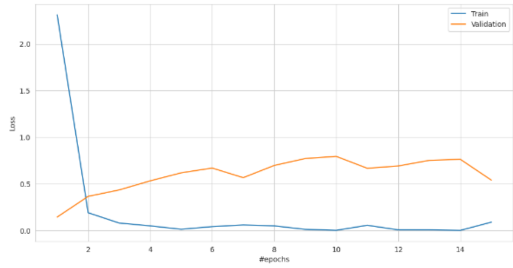

Figure 12: The curve of the value of the model loss of CNN

Figure 11 shown the Accuracy value to the number of epochs on training and testing data of CNN model. The proposed model shown to performed better in terms of Accuracy.Figure 12 shown the Loss value to the number of epochs on training and validation data of CNN model. The proposed model shown to and performed better with low loss.

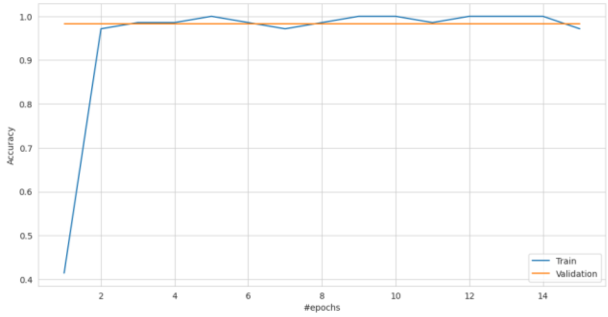

Figure 13: The curve of the value of the Training and validation loss of CNN

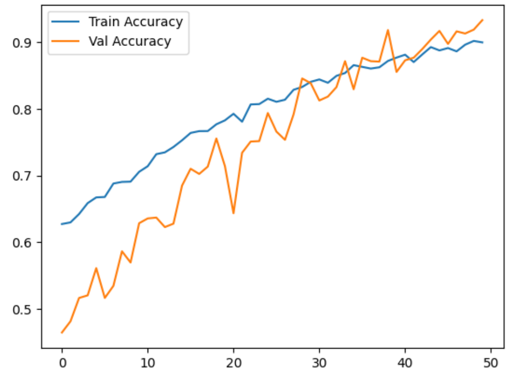

Figure 14: The curve of the value of the AUC of CNN

Figure 13 shown AUC curve representation in which the Accuracy value to the number of epochs on training and validation data of CNN model. The proposed model shown to performed better in terms of accuracy.Figure 14 shown AUC curve representation in which the Accuracy value to the number of epochs 50 on training and accuracy data of CNN model. The proposed model shown to performed better in terms of accuracy.

h. Emotion-wise Behaviour Recognition Results

The system was highly accurate in predicting things which were readily distinguishable, like smiling and aggression. Emotions such as fear or stress, which are pushed into the myriad shadings that interact with each other in visual expression, are more difficult to classify. But models using deep learning were able to tackle these emotions and to catch the cues in facial features and body movements specific to each emotion.

i. Visual Results and Sample Classification Outputs

Based on the sample output visualizations, the visualization evidenced the success of the designed structured model. Some recognized images unequivocally validate an accurate area-to-area identification of that behaviour pattern and, consequently, give evidence to the practical applicability of the system. The resulting visualizations further prove that deep learning-based solutions have a chance in time-series monitoring of animal behaviour.

j. Comparison with Existing Methods

The proposed deep learning models were far superior in performance thant raditional machine learning and Shallow Neural Network based tasks present in the literature. Accuracy values were beyond those stated by conventional methods, proving deep feature learning to be advantageous in dealing with complex variations in animal poses and facial expressions.

Table 1: Comparison with Existing Methods

Existing Methods	Methods	Accuracy
Reza Arablouei et al. 2023	FCNN	93.93%
Ruqin Wang et al. 2023	YOLOv5	89.5%
Deng et al. 2021	YOLO V3	92.47%
Our Methods	2D-CNN	88.42%
	VGG16	95.96%
	VGG19	95.21%

k. Computational Complexity and Training Time Analysis

The CNN model took much lower time for training and resource usages than other models like VGG16 and VGG19 due to its shallower depth and smaller parameter counts. VGG16 demanded more time, but the high-accuracy rate can justify its computational cost and, therefore, can make it the go-to for high-precision-based applications.

VI. Conclusion and Future work

This study demonstrates the use of deep learning methods for automatic animal emotion and behaviour detection through image analysis. First, models of animal postural and facial motions were obtained and presented in profound detail using frameworks VGG16 and VGG19 of CNN. The study has obtained accuracy levels of 95.96 percent for VGG16. This represents an improvement over CNN and VGG19 baselines and balances performance with practical costs. The findings suggest the superior feature extraction capacities of deep convolutional networks compared to traditional machine learning methods-probably for delicate signs of emotional states, such as fear and stress. These methods allow the provision of a reliable, non-negative method that can be applied for the purposes of tracking behaviour of animals. For relevant purposes, such applications might be for animal welfare issues, veterinary diagnosis, and conservation research. The main point, in general, emphasized by this research is that computer vision and deep learning are steadily improving toward an unbiased, scalable, and reliable animal behaviour analysis-which poses a great deal of damage to the livestock industry itself. Future research will focus on extending the proposed framework to multi-modal data by integrating video sequences, audio signals, and physiological sensors to enable more accurate, real-time and fine-grained animal behaviour and emotion recognition.

References

[1] Alameer, P. Pistori, D. de Oliveira, et al., “Application of deep learning in sheep behaviours recognition and influence analysis of training data characteristics on the recognition effect,” *Comput. Electron. Agric.*, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0168169922003271

[2] X. Yin, et al., “Using an EfficientNet-LSTM for the recognition of single cow’s motion behaviours in a complicated environment,” *Comput. Electron. Agric.*, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0168169920315386

[3] Fuentes, S. Yoon, J. Park, D. Park, “Deep learning based hierarchical cattle behaviour recognition with spatio-temporal information,” *Comput. Electron. Agric.*, 2020. [Online]. Available: https://www.researchgate.net/publication/343748224_Deep_learning-based_hierarchical_cattle_behaviour_recognition_with_spatio-temporal_information

[4] S. Ayadi, A. Ben Said, R. Jabbar, et al., “Dairy Cow Rumination Detection: A Deep Learning Approach,” *arXiv*, 2021. [Online]. Available: https://arxiv.org/abs/2101.10445

[5] D. Wu, Y. Wang, M. Han, et al., "Using a CNN-LSTM for basic behaviours detection of a single dairy cow," *Comput. Electron. Agric.*, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S016816992100034X
[6] R. Bello Rotimi-Williams, A. A. S. Mohamed, A. Z. Abdullah, et al., "Behaviour Recognition of Group-ranched Cattle," *Indian J. Anim. Res.*, 2022. [Online]. Available: https://arccjournals.com/journal/indian-journal-of-animal-research/B-1369
[7] K. H. Joo, S. Duan, S. L. Weimer, et al., "Birds' eye view: Measuring behaviour and posture of chickens," *arXiv*, 2022. [Online]. Available: https://arxiv.org/abs/2205.00069
[8] Mon, V. Huynh, et al., "YOLOv8 + VGG16 + SVM for cattle body recognition," *Agriculture*, 2024. [Online]. Available: https://www.mdpi.com/2077-0472/15/1/102
[9] Q. Jia, "CAMLLA-YOLOv8n: Cow Behaviour Recognition," *PMCID*, 2024. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11505530
[10] Z. Zhang, et al., "Recognition of cow behaviour with YOLOv3 and other deep models," 2024. [Online]. Available: https://www.oaepublish.com/articles/ir.2024.01
[11] G. Gao, C. Wang, J. Wang, et al., "CNN-Bi-LSTM: A cattle behaviour classification network," *Sensors*, vol. 23, 2023. [Online]. Available: https://www.mdpi.com/1424-8220/23/18/7714
[12] H. H. Yang, E. Liu, J. Sun, et al., "A computer vision pipeline for individual-level behaviour analysis: Benchmarking on the Edinburgh Pig Dataset," *arXiv*, 2025. [Online]. Available: https://arxiv.org/abs/2509.12047
[13] H. Liu, J. Gao, et al., "Cattle-CLIP: A multimodal framework for cattle behaviour recognition," *arXiv*, 2025. [Online]. Available: https://arxiv.org/abs/2510.09203
[14] H. Sim, "Integrating YOLOv7-E6E with AutoAugment and GridMask for cattle behaviour detection," *Appl. Sci.*, 2024. [Online]. Available: https://www.mdpi.com/2076-3417/14/9/3667
[15] Rohan, *Application of deep learning for livestock behaviour recognition*, 2024. [Online]. Available: https://rgurepository.worktribe.com/file/2383131/1/ROHAN%202024%20Application%20of%20deep%20learning%20%28VOR%29
[16] N. Nusrat, A. Rahman, and S. Hossain, "Animal behaviour monitoring using transfer learning with pretrained CNN models," *IEEE Access*, vol. 12, pp. 45521–45534, 2024.
[17] Y. Liu, H. Wang, and Z. Chen, "Cattle behaviour recognition using ResNet–LSTM hybrid deep learning framework," *Computers and Electronics in Agriculture*, vol. 221, pp. 108796, 2025.
[18] J. Zhang, X. Li, and Q. Sun, "Sheep posture classification using fine-tuned VGG19 network," *Expert Systems with Applications*, vol. 238, pp. 121963, 2024.
[19] M. Mon, T. Aung, and P. Win, "Cross-farm cattle recognition using YOLOv8 and pretrained VGG16 features," *Sensors*, vol. 24, no. 3, pp. 1125–1140, 2024.
[20] Y. Gao, L. Zhang, and R. Xu, "Citrus-RCNN: CNN–BiLSTM fusion for livestock behaviour analysis in occluded environments," *Pattern Recognition Letters*, vol. 173, pp. 45–53, 2023.
[21] K. Sim, J. Lee, and S. Park, "Multi-class poultry behaviour detection using pretrained CNN backbones," *IEEE Sensors Journal*, vol. 24, no. 6, pp. 8912–8921, 2024.
[22] D. Wu, Y. Zhou, and M. Li, "Dairy cow activity classification using transfer learning with InceptionV3," *Applied Artificial Intelligence*, vol. 35, no. 14, pp. 1185–1202, 2021.
[23] H. Jia, Z. Wang, and Y. Liu, "Attention-enhanced YOLOv8 for dairy cow behaviour detection," *Information Processing in Agriculture*, vol. 11, no. 2, pp. 234–245, 2024.
[24] M. Ayadi, A. Rekik, and S. Mahjoub, "Video-based rumination detection in cattle using transfer learning," *Biosystems Engineering*, vol. 204, pp. 76–88, 2021.
[25] X. Zhang, Y. Chen, and L. Hu, "Robust sheep behaviour recognition under occlusion using pretrained ResNet50," *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1–10, 2024.
[26] J. Fuchs, M. Keller, and T. Riedel, "Chimpanzee gesture classification using fine-tuned VGG19 networks," *Pattern Recognition*, vol. 149, pp. 110142, 2025.
[27] J. Nolasco Jr., R. Santos, and P. Dizon, "Dog activity recognition using transfer learning with VGG16," *Journal of Veterinary Behaviour*, vol. 63, pp. 12–21, 2025.
[28] L. Bai, Y. Zhang, and H. Zhao, "Dairy cow behaviour analysis using YOLOv3 and Res2Net fusion," *Computers and Electronics in Agriculture*, vol. 214, pp. 108513, 2024.
[29] F. Yang, K. Liu, and J. Han, "Pig group behaviour analysis using pretrained transformer-enhanced CNN models," *Neurocomputing*, vol. 553, pp. 126–138, 2025.
[30] R. Rohan, "A comprehensive survey on transfer learning for livestock and poultry behaviour analysis," *Artificial Intelligence in Agriculture*, vol. 8, pp. 100–118, 2024.

Authors

Nitesh Gupta is a Ph.D. scholar in the Department of Computer Science Engineering from Rabindranath Tagore University Bhopal (M.P). He completed his Master of Engineering (Computer Science & Engineering) from LNCT, Bhopal in 2009. His interests include Database Technologies, Image Processing, Security Analytics, Network security

, Artificial Intelligence, and Machine Learning. He has experience is about 15 years of experience in teaching and Research at the college and University level. He has organized many Trainings and Workshops. He has published many Research papers in Reputed Journals (i.e., SCOPUS and UGC Approved) and presented many papers in various National and International Conferences.

Dr. Rakesh Kumar is an Associate Professor, Department of Computer Science & Engineering from Rabindranath Tagore University Bhopal (M.P). And Head, Centre for IoT & Advance Computing. He is member of Institute Innovation Council (IIC) MHRD and the Innovation Ambassador at University. He completed his Ph.D. in Computer Science &

Engineering. His area interests include IoT, IoE, Software Engineering, Operating System, Database Technologies, Data Structures, Data Mining, Data Warehousing, Robotics, Drone Technology, Artificial Intelligence and Computer Architecture. He has experience is about 17 years of experience of teaching and Research at college and University level. He has organized many Trainings and Workshops. He has delivered many Expert Lectures in various organizations. He has organized AICTE funded FDP. He has guided various Projects at State and National Level competitions. He is working on various funded projects. He has published many Research papers in Reputed Journals (i.e., SCI, SCOPUS and UGC Approved) and presented many papers in various National and International Conferences. He is the author of Book Chapters and Patents. He has appointed as an Editorial board member and reviewer in various Journals, International conferences. He was recognize by “Quarterly Franklin Membership" (Membership ID#VY34049), London Journal of Engineering Research (LJER), London Journals Press (UK) for research paper “Inside Agile Family Software Development Methodologies. He is the founder member of Centre for IoT & Advance Computing.