As per UGC guidelines an electronic bar code is provided to secure your paper

International Journal for Modern Trends in Science and Technology
Volume 12, Special Issue 01, pages 87-93.
ISSN: 2455-3778 online
Available online at: http://ijmtst.com/ICRCSET2026.html
DOI: https://doi.org/10.5281/zenodo.19426936

Deep Learning Techniques for Classifying Bird Species

Y. Srimani | Dr. P. Naga Malleswara Rao

Department of Computer Science and Engineering, Chalapathi Institute of Technology, Guntur, Andhra Pradesh, INDIA.

To Cite this Article
Y. Srimani & Dr. P. Naga Malleswara Rao (2026). Deep Learning Techniques for Classifying Bird Species. International Journal for Modern Trends in Science and Technology, 12(SI01), 87-93. https://doi.org/10.5281/zenodo.19426936

Article Info
Received: 02 March 2026; Revised: 01 April 2026; Accepted: 04 April 2026.

KEYWORDS

Autograph; Caltech-UCSD; grey scale pixels; Tensorflow

ABSTRACT

Many bird species are becoming more difficult to locate, and even when they are, it may be difficult to anticipate their classification. Observed from a distance, birds may be seen in a wide range of sizes, shapes, colors, and orientations. The photographs show a great deal more variation in the bird's breed than the auditory categorization. Using photos, people are better able to discriminate between birds. Because of this, the Caltech-UCSD Birds 200 dataset is used for both training and validation in this technique. Deep convolution neural networks (DCNN) are used to turn an image into a grayscale representation, while the tensor flow is used to construct an autograph with a large number of nodes that can be compared. A rating table is developed as a consequence of evaluating the numerous entry points to the validation data. By examining the scoreboard and selecting the highest rating, it may be able to anticipate the required bird group. A look at the dataset (CUB-200-2011) reveals that the system has a bird identification accuracy of 89 percent. Linux and the Tensorflow framework were used in the study.

INTRODUCTION

The study of bird behavior and population dynamics has lately gained popularity [1]. Due to their quick response time to changing weather conditions, birds help us identify other species in the environment (such as the bugs they consume). A far more time-consuming and costly approach is to collect data on birds by hand. We need an effective tool for scientists, government agencies, and the general public to utilize to evaluate large volumes of bird data on a large scale. Therefore, identifying the kind of bird in an image is crucial. This is why bird species identification is so important. Based on an image, a bird species is classified into one of many groups. It is possible to identify a person based on their voice, picture, or video. Aural computational approaches can capture a bird's auditory output and use it to identify it. However, the interpretation of such information is complicated by the many sounds in the surroundings, such as bugs, real-world things, and so on. For most people, images have a greater impact than speech or movies. As a consequence, the use of images rather than audio or video to classify birds is more popular. Species categorization for birds is a challenging issue for both humans and computer algorithms. For millennia, birdwatchers have been concerned with the taxonomy of bird species. To better understand species, biologists must look at their distribution, ecology, migration patterns, and overall impact on the ecosystem as a whole. Ornithologists use Linnaeus' classification method to classify birds into Kingdoms, Phyla, Classes, Orders, Families, and Species [4].

As picture categorization algorithms improve, artifacts are migrating into databases with significantly more characteristics, such as Caltech-UCSD. New studies have had a lot of progress in this field. Caltech-UCSD Birds 200 (CUB-200-2011) is a well-known database for bird photographs that includes shots from 200 different classifications [5]. The collection primarily includes birds located in North America. Caltech-UCSD Birds 200 contains 11,788 photos with metadata such as 15 Component Positions, 312 Numeric Characteristics, and 1 Reference Image. Instead of recognizing a high variety of different classifications, this work investigates the difficulty of recognizing a high amount of classes within a single class - that of birds. Due to the obvious high degree of resemblance between groups, categorizing birds presents an additional hurdle. In addition, there is a broad range of variance within categories since birds are non-rigid entities that may distort in many ways. Prior study on bird categorization has concentrated on a limited amount of categories or sound.

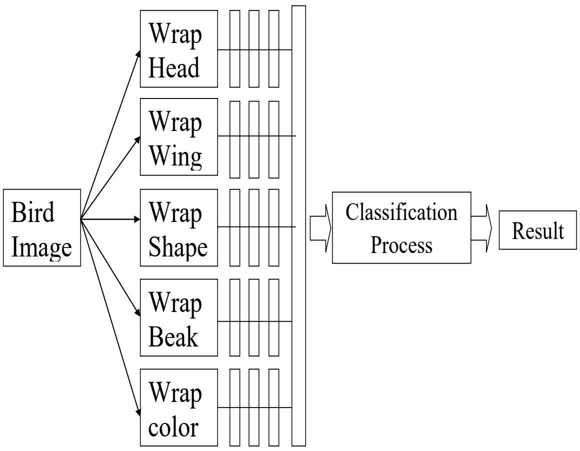

Fig. 1. Classification Process

Fig. 1 depicts the procedure of finding the bird in a picture. The photograph is uploaded initially, and then several configurations such as skull, thorax, color, plumage, and full picture are examined. In addition, each configuration is delivered across a DCNN to retrieve characteristics from several network levels [6]. Following that, the picture's depiction will be considered. The classifying result (i.e. characteristics are aggregated to transmit to the classification model) will then be acquired and the bird species found. Listed below are the guidelines for this piece of writing: Section II highlights the things to evaluate while physically identifying a bird. Section III and IV describes the approaches utilized to create the developed methodology. Section V depicts the total flow diagram in depth.

RELATED WORKS

Bird recognition is primarily done optically or audibly. The key visual features include the bird's form, feathers, shape, attitude, color, and so on [7]. Nevertheless, when evaluating the criteria, the season of the year must be addressed because the feathers of birds alter as they develop. Bird songs and calls are made up of sound elements [8]. Shoulder patches, wings bands (which are characterized as fine lines upfield), eyebrow bands, crowns, and forehead are all important for distinguishing one bird from another [9]. The form of the snout is typically a significant factor in identifying a bird. Bird traits including form and pose are the greatest widely utilized to describe birds. Since this trait is tough to alter, many specialists can recognize a bird by its appearance. The wing of a bird can also be used to identify it. The head can be identified in a variety of forms, including hooked, straight and curved, or curved. Feet are often employed to recognize a picture in shorter or longer formats [10]. A specific component will not produce an appropriate outcome. As a result, numerous variables must be evaluated to provide proper results. The scale of a bird in a photograph fluctuates based on parameters such as quality, range between the creatures and the recording equipment, and camera optical length [11]. As a result of a direct discovery of a vast collection of pictures, pictures are discriminated against due to the hue, which comprises varied components [12]. It has been discovered that the higher the picture clarity, the better the reliability [13] [14]. The automated bird taxonomy recognition for bird photos project presents a sequence of comparisons performed in a CUB- 200 database comprising over 6,000 photographs classified into 200 distinct categories [15]. They evaluated two distinct color schemes, RGB and HSV, as well as a varied amount of creatures to be categorized in this work. If the model includes more than 70% of the images, the resulting quality ranged from 8.82 percent to 0.43 percent [16].

PROPOSED METHODOLOGY

Various approaches were employed in the development of the platform. Database (Berkeley Birds 200), DCNN, Unsupervised Learning Algorithm (ULA), and so on. Because the inputting picture definition is unknown in this research, a ULA was employed to construct the machine. Furthermore, the information fed into the ULA is not labeled, i.e. only the intake parameters (X) are provided with no matching outcome parameters. Systems in ULA identify intriguing patterns in information. In more depth, grouping is utilized to divide information into classes [4]. DL methods were employed in detail to discover large perceptrons. As the picture passes through each NN layer, DL algorithms understand more about it. NN is used for classification. Fig. 2 depicts NN layers for information retrieval. The NN serves as a foundation for many machine learning (ML) methods. NN is made up of a weighted sum (W) and a biased matrix (B). Deep neural networks (DL) often employ CNNs to recognize images. It has a receptive field, an activation function, and several concealed nodes. A cluster of synapses is formed at each layer, and each level is connected to the previous layer's synapses. The output is predicted by the resulting level. An image is sent into the fully connected layer, which then generates a set of distinctive translations [2]. The input picture may comprise many inputs such as color, feathers, eyeballs, and bird beaks, implying that the convolution operation will translate one 3d point cloud to the next. The length, elevation, and thickness of 3D objects are taken into account.

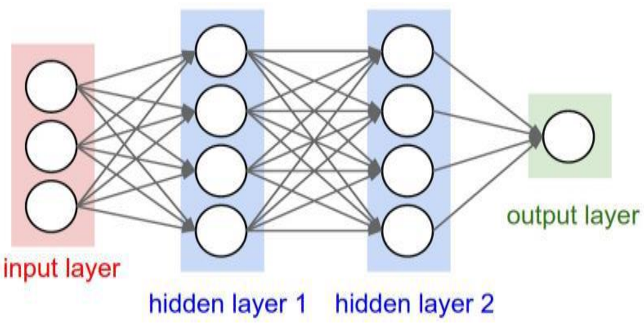

Fig. 2. NN Layers

CNN is divided into two sections. Extraction Section: When the system performs a series of coevolutionary processes, it discovers characteristics. Segmentation section: retrieved characteristics are sent into convolutional layers that serve as a predictor as shown in Fig.3.

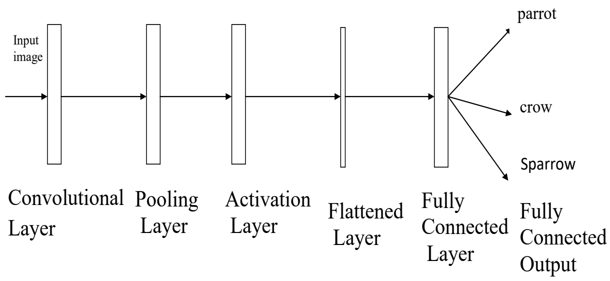

Fig. 3. CNN Layers

CNN is made up of four stacks: a completely linked, an activator overlay, a hidden state, and a dense layer. Tiny amounts of visual features may be extracted from an image using the convolution layer. It is possible to reduce the number of synapses in an earlier layer's convolutional network while still preserving the relevant data by using pooling. The activated level runs a value via an algorithm that reduces variables into a spectrum. A fully linked layer links every synapse on one level to every synapse on the next. CNN delivers more reliability since it identifies each synapse in detail. Picture categorization in ML is often accomplished in two aspects: A grayscale and Making use of RGB colors

Typically, all information is transformed to a grey scale. The machine will give weights to each picture in the grey scale method depending on how the number of the image is it. All of the image pixels are placed in a matrix, and the machine uses that myriad to perform classification. Tensorflow is a Google-created free access computer framework. It allows engineers to command every synapse, called a "network," allowing the settings to be altered to reach the required efficiency. Tensorflow has a plethora of picture categorization packages [3]. Tensorflow is in charge of constructing an autograph, which is made up of a succession of computing elements. Every computational component in the network symbolizes a statistical function as well as a link or border between components. ML allows developers to do these computations using the Python platform. A database is a grouping of data. A dataset called Berkeley Birds 200 (CUB-200-2011) is utilized for executing bird-related actions. It is an expanded variant of the CUB-200 database, with almost twice the amount of photos per classification and additional component position labels for improved correctness [8]. The dataset's comprehensive description is as continues to follow: There are 200 different classes. There are 11,788 photos in all. 15 Component Addresses, 312 Numeric Properties, and 1 Reference Graphic per picture as shown in Fig.4.

Fig. 4. Images in Dataset

The real operation of the suggested scheme is depicted in Fig. 5. To build software like this, a database must be created to classify a photo. Preparation and final test components are separated in a training database. The database needs to be reprogrammed using retrain.py in Google Collab to increase recognition efficiency. There are 50000 motions in the database because the more phases there are, the more precise the movement will be. 93 percent of the time, the trained model performs as expected. Over 1000 images are included in the experimental database, which has a reliability of 80%. For further efficiency, the database has been thoroughly vetted with a 75% dependability level. Input images are temporarily stored in databases when they are uploaded by users. An algorithm is used to process these data files, which are subsequently transmitted to CNN and matched with the classification results from the machine. A CNN consists of multiple layers that are all linked in some way. Several connections, such as the bird's face, thorax, color, mouth, shape, and overall image, are assessed for classification to obtain the highest level of accuracy. DCNN pulls attributes from several layers of the network to give every configuration. An unsupervised approach called DL and CNN is then used to classify the image. As a result, the image is categorized by the grey level of each pixel. The classifier then uses these properties to classify the data. To create probable outcomes, the feed will be matched to the training examples. After categorization, an autograph is created, which is made up of connections that eventually create a pattern. A scoring card is built on the premise of this system, and output is created with the assistance of the evaluation form.

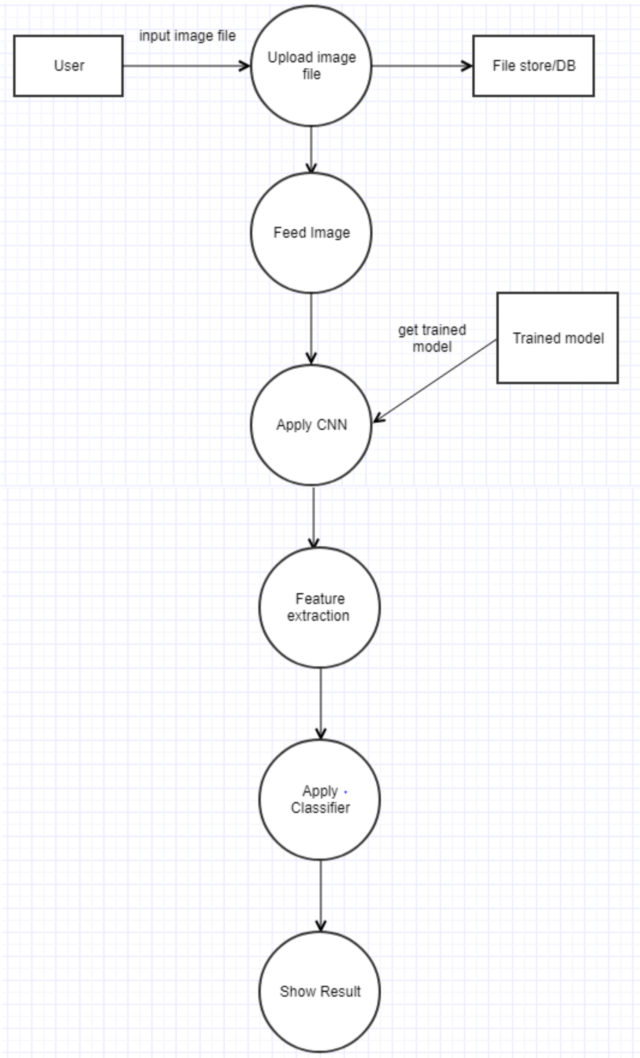

Fig. 5. Methodology employed

RESULTS AND DISCUSSION

On the Berkeley Birding 200 (CUB-200-2011) database, the suggested technique for bird taxonomy identification was evaluated by taking into account color characteristics and factors like height, structure, and so on. This is a picture collection tagged with 200 bird types. It contains 11,788 tagged photos of avian, each with a crude recognition, a reference picture, and quantitative feature evaluations. The database is trained using Google-Collab, which is a tool for training datasets by importing a photo from your native workstation or via Cloud Storage. After learning, the labeled information is available for image analysis algorithms. There are about 250 example photographs per creature contained in the collection of seven species that are physically caught in their native environment, thus ambient elements such as grassland, bushes, and other aspects are also contained in the picture. Birds can recognize in any situation since the major attention is on the height, structure, and color parameters. These characteristics are initially examined for classification, with RGB and gray level approaches utilized for statistics. That is, the picture is translated into gray levels employing the grey level approach, where a result for each picture is produced and value-based terminals, also known as neurons, are constructed. The architecture of matching images is just a network of linked nodes with these synapses reasonably specified. The autograph is created based on the clusters established, which ML may use to categorize the picture. This autograph is then captured by classifications, and the picture is matched to photographs from the Caltech UCSD pre-trained database, and a scoring card is created. The scorecard is a conclusion that provides the five best final scores, with the greatest corresponding score being the outcome of bird species. In this case, an experiment was conducted to achieve 80% precision by teaching the Caltech UCSD. Referring To Fig. 6 as an instance of source images presented to the algorithm for categorization of a bird from Northern America. Let's wait and see how it's received.

Fig.6. Image input

TABLE I. EVALUATION APPROACH

S.No	Species	Score generated
1	Elegant tern	0.00921
2	Red faced cormorant	0.00929
3	Brant cormorant	0.0082
4	Pelagic cormorant	0.0085
5	White pelican	0.00807

TABLE II. COMPARATIVE ANALYSIS

S.No	Model	Accuracy
1	Pose Norm	82%
2	Part-based R-CNN	78.2%
3	Multiple granularity CNN	83%
4	Diversified visual attention network (DVAN)	80%
5	The deep LAC localization, alignment, and classification	82.7%
6	Proposed Method	89%

Table I and Table II display the scoring depending on the platform's output. After analyzing these results, it was discovered that the creature with the best score was projected to be a necessary variety. This conclusion is depicted in Fig.7 and Fig. 8.

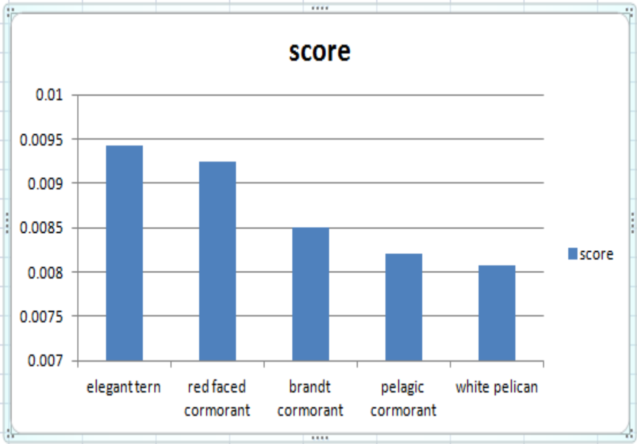

Fig.7. Graphical representation of scores generated

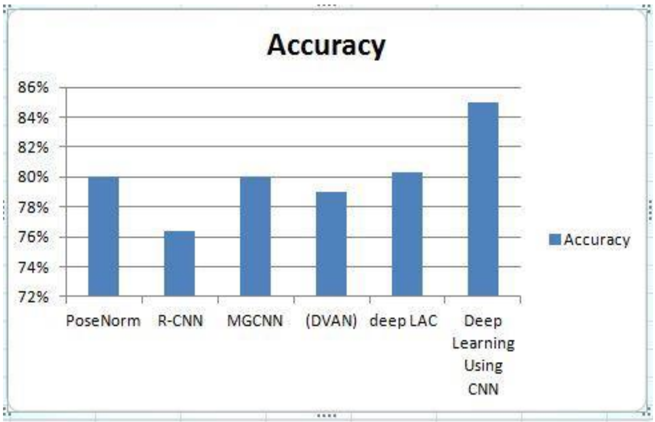

Fig.8. Comparative analysis graph

After reviewing the information, it is discovered that using a specified input results in lower precision.

However, if a combination technique is employed, the precision of the design improves by taking into account elements such as position, feathers, color, head, limbs, and so on.

CONCLUSION AND FUTURE SCOPE

A DL method (Unsupervised Learning) was used to identify bird species in a database (Caltech-UCSD Birds 200) using photo classification. There are 11,788 photos spread throughout 200 parts. A user-friendly website where a person may upload an image for verification is connected to the module, and it provides the desired results. The suggested approach is focused on the recognition of a portion and the extraction of CNN characteristics from several fully connected layers. These characteristics are gathered and then sent into the algorithm for categorization. Based on the data, the algorithm predicted the presence of 80 percent of bird groups with an efficiency of 89 percent. Rather than a webpage, build an Android/iOS application. It will be much handier for the user. We could use the internet for this, which might store enormous amounts of data for comparisons and provide a significant boost in computing power (in the case of NN).

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] Toth, B.P. and Czeba, B., 2016, September. Convolutional Neural Networks for Large-Scale Bird Song Classification in Noisy Environment. In CLEF (Working Notes) (pp. 560-568).

[2] Fagerlund, S., 2007. Bird species recognition using support vector machines. EURASIP Journal on Applied Signal Processing, 2007(1), pp.64-64.

[3] Pradelle, B., Meister, B., Baskaran, M., Springer, J. and Lethin, R., 2017, November. Polyhedral Optimization of TensorFlow Computation Graphs. In 6th Workshop on Extreme-scale Programming Tools (ESPT-2017) at The International Conference for High Performance Computing, Networking, Storage and Analysis (SC17).

[4] Cireşan, D., Meier, U. and Schmidhuber, J., 2012. Multi-column deep neural networks for image classification. arXiv preprint arXiv:1202.2745.

[5] Andr´eia Marini, Jacques Facon and Alessandro L. Koerich Postgraduate Program in Computer Science (PPGIa) Pontifical Catholic University of Paran´a (PUCPR) Curitiba PR, Brazil 80215–901 Bird Species Classification Based on Color Features

[6] Image Recognition with Deep Learning Techniques ANDREIPETRU BĂRAR, VICTOR-EMIL NEAGOE, NICU SEBE Faculty of Electronics, Telecommunications & Information Technology Polytechnic University of Bucharest.

[7] Xception: Deep Learning with Depthwise Separable Convolutions François Chollet Google, Inc.

[8] Zagoruyko, S. and Komodakis, N., 2016. Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer. arXiv preprint arXiv:1612.03928.

[9] Inception-v4,Inception-ResNetand the Impact of Residual Connectionson Learning Christian Szegedy, Sergey Ioffe,Vincent Vanhoucke ,Alexandrr A.Alemi

[10] Stefan Kahl, Thomas Wilhelm-Stein, Hussein Hussein, Holger Klinck, Danny Kowerko, Marc Ritter, and Maximilian Eibl Large-Scale Bird Sound Classification using Convolutional Neural Networks

[11] K. M S, R. R. G and S. Karthik, "Streamlining Load Scheduling in Cloud Computing: A Thorough Performance Assessment and Development of Effective Methods for Design," 2024 International Conference on Advances in Modern Age Technologies for Health and Engineering Science (AMATHE), Shivamogga, India, 2024, pp. 1-7, doi: 10.1109/AMATHE61652.2024.10582239.

[12] Sai Srinivas Vellela, Roja D, NagaMalleswara Rao Purimetla, SyamsundaraRao Thalakola, Lakshma Reddy Vuyyuru, Ramesh Vatambeti, Cyber threat detection in industry 4.0: Leveraging GloVe and self-attention mechanisms in BiLSTM for enhanced intrusion detection, Computers and Electrical Engineering,Volume 124, Part A,2025,110368,ISSN 00457906,https://doi.org/10.1016/j.compeleceng.2025.110368.

[13] K. M S, R. R. G and S. Karthik, "Streamlining Load Scheduling in Cloud Computing: A Thorough Performance Assessment and Development of Effective Methods for Design," 2024 International Conference on Advances in Modern Age Technologies for Health and Engineering Science (AMATHE), Shivamogga, India, 2024, pp. 1-7, doi: 10.1109/AMATHE61652.2024.10582239.

[14] Sai Srinivas Vellela, Roja D, NagaMalleswara Rao Purimetla, SyamsundaraRao Thalakola, Lakshma Reddy Vuyyuru, Ramesh Vatambeti, Cyber threat detection in industry 4.0: Leveraging GloVe and self-attention mechanisms in BiLSTM for enhanced intrusion detection, Computers and Electrical Engineering,Volume 124, Part A,2025,110368,ISSN 00457906,https://doi.org/10.1016/j.compeleceng.2025.110368.

[15] S. S. Vellela, L. R. Vuyyuru, K. B. S. K, N. MalleswaraRaoPurimetla, L. Dalavai and M. V. Rao, "A Novel Approach to Optimize Prediction Method for Chronic Kidney Disease with the Help of Machine Learning Algorithm," 2023 6th International Conference on Contemporary Computing and Informatics (IC3I), Gautam Buddha Nagar, India, 2023, pp. 1677-1681, doi: 10.1109/IC3I59117.2023.10397974.

[16] Kavitha Mettupalayam Subramaniam, Ramachandra Rao Goli, Karthik Subburathinam, Srihari Kannan,Optimization of pyrolysis parameters for enhanced biochar production from agricultural biomass: A study on energy efficiency and carbon sequestration potential,Science of The Total Environment, Volume 1015,2026,181362,ISSN 00489697,https://doi.org/10.1016/j.scitotenv.2026.181362.

[17] K. K. Kumar, S. G. B. Kumar, S. G. R. Rao and S. S. J. Sydulu, "Safe and high secured ranked keyword searchover an outsourced cloud data," 2017 International Conference on Inventive Computing and

Informatics (ICICI), Coimbatore, India, 2017, pp. 20-25, doi: 10.1109/ICICI.2017.8365348.

[18] R. K. Yarava, G. R. C. Rao, Y. Garapati, G. C. Babu and S. D. V. Prasad, "Analysis on the Development of Cloud Security using Privacy Attribute Data Sharing," 2022 First International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT), Trichy, India, 2022, pp. 1-5, doi: 10.1109/ICEEICT53079.2022.9768608.

[19] K. K. . Kommineni and A. . Prasad, "A Review on Privacy and Security Improvement Mechanisms in MANETs", Int J Intell Syst Appl Eng, vol. 12, no. 2, pp. 90–99, Dec. 2023.

[20] Kommineni, K.K., Prasad, A. Enhancing Data Security and Privacy in SDN-Enabled MANETs Through Improved Data Aggregation Protection and Secrecy. Wireless Pers Commun 139, 855–882 (2024). https://doi.org/10.1007/s11277-024-11635-w

[21] "Blockchain-Enabled Secure Data Aggregation for SDN-Enabled Ad-Hoc Networks," International Journal of Intelligent Engineering and Systems, vol. 18, no. 5, pp. 704–717, Jun. 2025, doi: https://doi.org/10.22266/ijies2025.0630.49.

[22] K. K. Kommineni, P. Ande, "Blockchain-driven key management and privacy-preserving data Aggregation Scheme for SDN-enabled MANETs," International Journal of Intelligent Engineering and Systems, vol. 18–18, no. 9, pp. 601–615, 2025, doi: 10.22266/ijies2025.1031.39.