INTERNATIONAL JOURNAL OF RESEARCH CULTURE SOCIETY ISSN(O): 2456-6683
Monthly Peer-Reviewed, Refereed, Indexed Journal [Impact Factor: 9.241]
Volume - 8, Issue - 10, October - 2024

DOIs:10.2017/IJRCS/202410012 --:-- Research Paper / Article / Review

A Survey on Prediction Models for the Iris Dataset: Comparative Analysis of Machine Learning and CNN Approaches

[1]Arpan G Singh, [2] Prof. Manish Kumar Singhal
[1]M.tech Scholar, [2]Associate Professor & H.O.D
[1,2]Department of Information Technology (IT)
[1,2]NRI Institute Of Information Science And Technology, Bhopal (Mp), India,
Email :[1]arpansingh61997@gmail.com [2]manishsinghal.nirt@gmail.com

Abstract: *In this survey paper discuss the The Iris dataset, a classic benchmark in pattern recognition and machine learning, has been widely studied for predicting species of iris flowers based on sepal and petal dimensions. This survey presents a comparative analysis of traditional machine learning models and Convolutional Neural Networks (CNN) applied to the Iris dataset. While machine learning algorithms like Decision Trees, k-Nearest Neighbors (k-NN), Support Vector Machines (SVM), and Logistic Regression have shown high accuracy in classifying the species, recent advancements in deep learning, specifically CNNs, provide a fresh perspective on feature extraction and classification, even for structured datasets like Iris. We evaluate the performance, strengths, and limitations of each model in terms of accuracy, interpretability, and computational cost. This survey paper explores demonstrate that while CNNs may not significantly outperform traditional methods in this simple dataset, they offer advantages in handling more complex data, showcasing the potential of deep learning for future studies on similar tasks.*

Key Words: *Machine Learning, Iris Flower, Logistic Regression, K-Nearest Neighbors, Decision Tree, Random Forest etc.*

1. INTRODUCTION:

A most fascinating research topic is Machine Learning today, today many people work in the machine learning field, and some publishers or researchers discovering or updating new algorithms and methods for Machine Learning, Basically, machine learning is the process of causing the machine to make the same decisions as the human brain. machine learning which is the main part of artificial intelligence has two main categories supervised learning and unsupervised learning. Hence, the learning phase is classified as Supervised learning, unsupervised learning, and Reinforcement Learning. As part of the supervised learning process, an output target is presented, which assists or makes the system in learning also it contains instances of training data that consist of different input attributes and an output. A subpart of supervised learning is Classification, where the program learns from the input data given to it and uses this process to classify new observations. Classification techniques have various types like Decision Trees, Neural Networks, Bayes Classifiers, Support Vector Machine, K-Nearest neighbors, and many more. Here are some examples of Machine learning classification tasks with both discrete and continuous data: Classifying credit card transactions, Detection of diseases in the human body classifying protein as alpha-helix as secondary structures of beta-sheet or random coil, weather forecasts, and categorizing news stories as finance, sports, and entertainmen. The Iris dataset, one of the most well-known datasets in machine learning, consists of 150 samples of iris flowers, with measurements such as sepal length, sepal width, petal length, and petal width. These samples are categorized into three distinct species: *Iris setosa*, *Iris versicolor*, and *Iris virginica*. This dataset is widely used for classification tasks and serves as an excellent starting point for applying machine learning algorithms due to its simplicity and well-structured data.

The goal of this project is to build a prediction model capable of classifying the species of an iris flower based on its measurements. To accomplish this, we will explore both traditional machine learning models, such as k-Nearest Neighbors (k-NN), Decision Trees, and Support Vector Machines (SVM), as well as modern deep learning approaches, specifically Convolutional Neural Networks (CNNs). CNNs, though typically applied to image data, can

also be used in unique ways for tabular data classification through feature extraction layers The Iris dataset, introduced by Sir Ronald A. Fisher in 1936, is one of the most well-known datasets in the field of data science and machine learning. It contains 150 records of iris flowers, with four features: sepal length, sepal width, petal length, and petal width. These features are used to classify the flowers into three species: Iris-setosa, Iris-versicolor, and Iris-virginica. Due to its simplicity and well-defined structure, the Iris dataset has become a standard benchmark for testing various machine learning models.

Traditional machine learning techniques, such as decision trees, support vector machines (SVMs), and k-nearest neighbors (KNN), have been widely used to classify the Iris dataset with remarkable accuracy. These methods are effective for structured, tabular data where features are predefined and relationships between data points can be easily inferred. However, with the advent of deep learning, more advanced approaches like Convolutional Neural Networks (CNNs) have also been explored for classification tasks. While CNNs are primarily used for image processing, their ability to capture complex patterns can be leveraged for structured data as well.

2. LITERATURE REVIEW:

Mashael M. Khayyat, et.al (2024), Author are presented a new The iris functions as a resilient instrument for dependable human identification, showcasing substantial promise in recognizing individuals with a considerable level of assurance. The crucial step in iris recognition lies in extracting effective features. Traditionally, various handcrafted features, devised by biometrics specialists, have been employed for implementing iris recognition systems. However, given the remarkable success of Fuzzy-deep-learning in addressing computer vision challenges, local binary patterns (LBP) features learned by Convolutional Neural Networks (CNNs) have garnered considerable interest for application in iris recognition systems. This study evaluates the LBP features followed by the Fuzzy-CNN model for classification. The system's performance is compared with several machine and deep learning models. The proposed model obtained an accuracy of 99.55%, 98.85% precision, 99.47% recall, and 99.22% F1 Score. Rigorous testing is conducted on four public datasets, namely IITD, CASIA-Iris-V1, CASIA-Iris-thousand, and CASIA-Iris-V3 Interval. The proposed iris recognition system demonstrates outstanding results, achieving a notably high accuracy rate [01].

Karima Boukari,et.al,(2024), Author are analysis Deep learning networks enable direct learning of feature representation schemes from images. This work presents a fast and efficient iris recognition system without relying on a successful iris segmentation step, which may not always be possible under uncontrolled conditions. The system calculates CNN features and compares them for two iris image databases (Org-IrD and Prep-IrD). The accuracy for the Org-IrD subset is better without any processing, indicating the strength of the CNN features in characterizing iris images based on texture and morphology. The recognition rates for original images out-perform those of normalized images, regardless of the used model (EfficientNet, Xception, ResNet50, or VGG19) and classifier (SVM or Softmax), for all tested databases (CASIA- Lamp V4, IITD, and MMU2). In a second experiment, a multi-modal recognition system is proposed using different models (EfficientNetBO, Xception, ResNet50, VGG19, and MobileNetv2). The system employs two parallel CNN structures for the left and right iris eyes, which generate two characteristic vectors that are fused and fed to the SVM classifier. Three fusions of features (concatenation, sum, and maximization) were tested, and the proposed system was evaluated on several public datasets (CASIA-Lamp V4, CASIA-Interval V3, IITD, and MMU2). The system achieved satisfying results, especially with MobileNet. However, these results need to be generalized to larger databases [02].

Chya Fatah Aziz, et.al, (2023), Author are study — Supervised Machine Learning algorithm has an important approach to Classification. We are predicting the deal type of the Iris plant using various algorithms of machine learning. Iris plants are determined by numerous factors such as the size of the length and width of the property. A horticultural skill announces that some of the plants are different in some physical appearances like size, shape, and color. Hence it is difficult to recognize any species. Versicolor, Setosa, and Virginica have three identical subspecies of The Iris flower species. This paper uses machine learning algorithms to recognize all classes of the flower with an accuracy degree of %100 for KNN, %95 for RF, %97 for DT, and %98 for LR. The Iris dataset is frequently available, and it is implemented using Scikit tools. and build the prediction model for Plants. Here, algorithms of machine learning such as Logistic Regression (LR), Decision Tree (DT), K Nearest Neighbor (KNN), and Random Forest (RF) are employed to construct a predictive model [03].

Ferdi Özbilgin et.al, (2023), Researcher are Coronary Artery Disease (CAD) occurs when the coronary vessels become hardened and narrowed, limiting blood flow to the heart muscles. It is the most common type of heart disease and has the highest mortality rate. Early diagnosis of CAD can prevent the disease from progressing and can make treatment easier. Optimal treatment, in addition to the early detection of CAD, can improve the prognosis for these patients. This study proposes a new method for non-invasive diagnosis of CAD using iris images. In this study, iridology, a method of analyzing the iris to diagnose health conditions, was combined with image processing techniques to detect the disease in a total of 198 volunteers, 94 with CAD and 104 without. The iris was transformed

into a rectangular format using the integral differential operator and the rubber sheet methods, and the heart region was cropped according to the iris map. Features were extracted using wavelet transform, first-order statistical analysis, a Gray-Level Co-Occurrence Matrix (GLCM), and a Gray Level Run Length Matrix (GLRLM). The model's performance was evaluated based on accuracy, sensitivity, specificity, precision, score, mean, and Area Under the Curve (AUC) metrics. The proposed model has a 93% accuracy rate for predicting CAD using the Support Vector Machine (SVM) classifier [04].

Swathi Gowroju, et. al (2022), Authors presented Iris recognition is a secure and best-chosen biometric application in the digital world because of its unique characteristics. Day by day, the digital world plays a significant role in human life for various applications. The applications are vastly spread over secure applications of the nation such as border control applications, criminal investigations, postmortem studies, access the digital equipment, smart homes, smart appliances, smart cars etc. Due to the digitalization of the world, all the research communities, scientists, and industries are focusing on the biometric-based secured iris recognition system. The researcher has implemented various algorithms based on traditional and neural network architectures. In this scenario, this paper gives a brief on different techniques and algorithms used by researchers to predict the age of human people using the iris [05].

Smita Khade, et.al, (2021), Author are study Iris biometric identification allows for contactless authentication, which helps to avoid the transmission of diseases like COVID-19. Biometric systems become unstable and hazardous due to spoofing attacks involving contact lenses, replayed video, cadaver iris, synthetic Iris, and printed iris. This work demonstrates the iris presentation attacks detection (Iris- PAD) approach that uses fragmental coefficients of transform iris images as features obtained using Discrete Cosine Transform (DCT), Haar Transform, and hybrid Transform. In experimental validations of the proposed method, three main types of feature creation are investigated. The extracted features are utilized for training seven different machine learning classifiers alias Support Vector Machine (SVM), Naive Bayes (NB), Random Forest (RF), and decision tree(J48) with ensembles of SVM+RF+NB, SVM+RF+RT, and RF+SVM+MLP (multi-layer perceptron) for proposed iris liveness detection. The proposed iris liveness detection variants are evaluated using various statistical measures: accuracy, Attack Presentation Classification Error Rate (APCER), Normal Presentation Classification Error Rate (NPCER), Average Classification Error Rate (ACER). Six standard datasets are used in the investigations. Total nine iris spoofing attacks are getting identified in the proposed method. Among all investigated variations of proposed iris-PAD methods, the 4 × 4 of fragmental coefficients of a Hybrid transformed iris image with RF algorithm have shown superior iris liveness detection with 99.95% accuracy [06].

Ananya Zabin et.al, (2020), Researcher are —An efficient iris sensor identification algorithm can be used in certain forensic applications, i.e. detecting mislabeled iris data at large scale iris datasets, and verifying the validity of the data origin of collected iris datasets that are available to be shared. Such knowledge can potentially increase the overall iris recognition system accuracy by offering the operator the option to match same-sensor or cross-sensor iris images. In either case the knowledge of the origin of the sensor used to collect these data, when not available, or the correction of mislabeled data, is expected to result in higher iris matching accuracy. Another benefit of iris sensor identification is that it can assist in improving the detection of fake iris data, i.e. when knowing the iris sensor, we can apply more appropriate models for fake detection that are tuned for a specific iris sensor. In this paper we propose an efficient deep learning-based iris recognition algorithm that is sensor inter-operable. Our approach utilizes a moderate amount of data and is adaptable to learning rate variations as well as variations of the amount of data used for training per class. Our proposed approach uses a set of iris datasets that include iris images captured at different standoff distances. We are using the original captured, dual eye, or periocular images rather than the iris itself, after detecting, segmenting, and normalizing the iris [07].

Sue Chin Yow (2019) – Identity recognition through human iris organ is claimed as one of the famous biometric techniques due to its reliability promising higher accurate return as compared to other traits. Reviewing past literatures, poor imaging condition, low flexibility of model, and small size iris images dataset are the limitations needing solutions. In this paper, a proposed algorithm development flow and systematic analysis has been conducted to achieve high efficiency in the iris recognition task. A transfer learning method that does not involve iris segmentation phase is proposed to capitalise pre-trained Convolutional Neural Network (ConvNet) model introduced in the ImageNet Large Scale Visual Recognition Competition (ILSVRC) on iris recognition system. Both data augmentation and Bayesian optimisation are also involved in optimising the network and prevent it from overfitting. Simulation results showed the transferability of a pre-trained model on new target task is improved and meanwhile, the high recognition rate of the algorithm on small-size Institute of Automation, Chinese Academy of Sciences (CASIA) Iris-Interval V1 iris image dataset is achieved [08].

3. IRIS FLOWERS CLASSIFICATION:

The Iris flower classification problem is a popular dataset used in machine learning and statistics to classify iris flowers into one of three species: Iris setosa, Iris versicolor, and Iris virginica. The dataset contains 150 samples, each with four features: sepal length, sepal width, petal length, and petal width, which describe the physical characteristics of the flowers. The objective is to use these features to build a model that can predict the species of an iris flower. Due to its simplicity and balanced classes, it is often used for demonstrating classification algorithms such as k-nearest neighbors (KNN), decision trees, and support vector machines (SVM). The Iris dataset was first introduced by the British statistician and biologist Ronald Fisher in 1936, making it one of the earliest examples of multivariate analysis.

The dataset has become a benchmark for testing and comparing machine learning algorithms due to its relatively small size and clean structure, making it an ideal starting point for beginners in data science. The three classes of iris flowers are linearly separable in some dimensions, but not in others, which challenges models to learn non-linear decision boundaries. Various approaches can be applied to solve the classification problem, such as logistic regression, neural networks, or ensemble methods like random forests. Data preprocessing, feature scaling, and splitting the data into training and testing sets are common practices used when working with the Iris dataset to ensure that models are trained effectively and generalize well to unseen data. Furthermore, the dataset can also be visualized easily through scatter plots or pair plots, providing insights into the relationships between different features and how they contribute to distinguishing the iris species. Overall, the Iris flower classification remains an essential exercise for those exploring supervised learning techniques in machine learning.

Flowers Classification, we will be dealing with Logistic Regression Machine Learning Algorithm. First, we will see logistic Regression, and then we will understand the working of an algorithm with the Iris flowers dataset. We all know about Iris Dataset, and it contains features of different flower species. Independent features in this dataset are Sepal Length, Sepal Width, Petal Length, and Petal Width. All these lengths were in centimeters. And Dependent feature, which will be the output for the model, is Species. It contains the name of the species to which that particular flower with those measurements belongs.

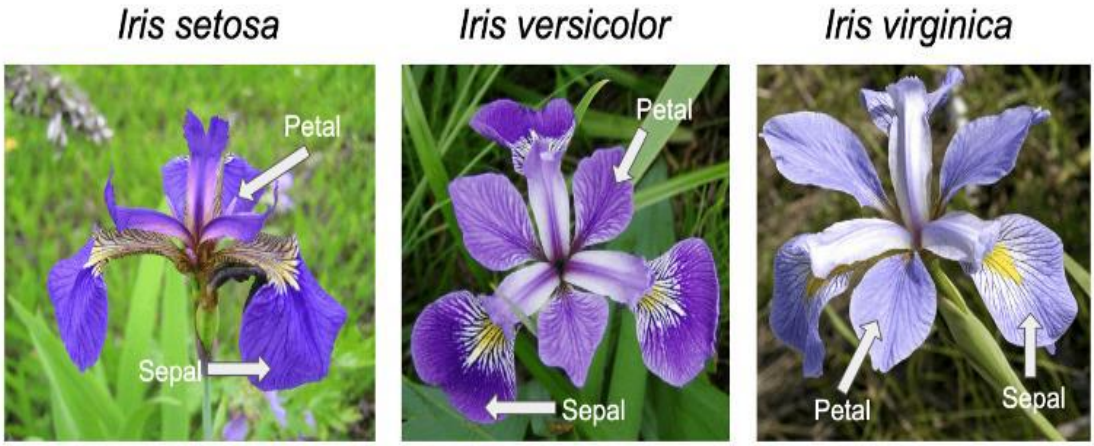

Figure 1: Iris Flowers Classification

4. CONCLUSION:

In this survey paper discuss on the prediction model based on the Iris dataset utilizing both traditional machine learning and convolutional neural network (CNN) approaches demonstrates significant promise in classifying iris species effectively. The comparative analysis reveals that while classical machine learning algorithms, such as decision trees, support vector machines, and k-nearest neighbors, offer reliable performance with relatively simple feature extraction, the CNN model showcases superior accuracy and robustness in handling the dataset's inherent complexities. The ability of CNNs to automatically extract relevant features from the data further enhances their effectiveness in classification tasks. the versatility of machine learning techniques in botanical classification and underscores the potential for applying more advanced models like CNNs to achieve even higher accuracy and efficiency in similar predictive tasks. Future work may explore additional datasets and model architectures to further improve classification performance and generalizability.

REFERENCES

1. Mashael M. Khayyat, Nuha Zamzami, Li Zhang, Michele Nappi , Muhammad Umer. "Fuzzy-CNN: Improving personal human identification based on IRIS recognition using LBP features." Volume 83, June 2024, 103761.
2. Karima Boukari. "Deep Learning-Based Iris Recognition System Using Unprocessed Images" J. Electrical Systems 20-3 (2024): 2119-2129.
3. Chya Fatah Aziz, and Banan Jamil Awrahman. "Prediction Model based on Iris Dataset Via Some Machine Learning AlgorithmsVol.10, No.2, Aug., 2023, pp 64-69.
4. Ferdi Özbilgin, Çetin Kurnaz and Ertan Aydın. "Prediction of Coronary Artery Disease Using Machine Learning Techniques with Iris Analysis." Volume 13 , Issue 6 , 13 March 2023.
5. Swathi Gowroju, Aarti & Sandeep Kumar "Review on secure traditional and machine learning algorithms for age prediction using IRIS imagePublished: 28 June 2022,Volume 81, pages 35503–35531, (2022)
6. Smita Khade, Shilpa Gite, Sudeep D. Thepade, Biswajeet Pradhan And Abdullah Alamri6. " Detection of Iris Presentation Attacks Using Hybridization of Discrete Cosine Transform and Haar Transform With Machine Learning Classifiers and Ensembles." Volume 9, 2021.
7. Ananya Zabin, Thirimachos Bourlai "A Deep Learning Based Approach to Iris Sensor Identification" IEEE/ACM ASONAM 2020, December 7-10, 2020 978-1-7281-1056-1/20/$31.00 © 2020 IEEE
8. Sue Chin Yow and Ahmad Nazri Ali. " Iris Recognition System (IRS) Using Deep Learning Technique." Journal of Engineering Science, Vol. 15(2), 125–144, 2019.
9. Sabhanayagam, T., Venkatesan, V. P. & Senthamaraikannan, K. (2018). A comprehensive survey on various biometric systems. Int. J. App. Eng. Res., 13, 2276–2297.
10. Al Rifaee, M. (2014). Unconstrained iris recognition. PhD diss., De Montfort University.
11. Masek, L. (2003). Recognition of human iris patterns for biometric identification. BEng diss., University of Western Australia.
12. Cruz, F. R. G., Hortinela, C. C., Redosendo, B. E., Asuncion, B. K. P., Leoncio, C. J. S., Linsangan, N. B. & Chung, W. -Y. (2016). Iris recognition using Daugman algorithm on Raspberry Pi. 2016 IEEE region 10 conference (TENCON). New Jersey, USA: Institute of Electrical and Electronics (IEEE), 2126–2129, https://doi.org/10.1109/ TENCON.2016.7848401.
13. Kunik, Z., Bykowski, A., Marciniak, T. & Dąbrowski, A. (2017). Raspberry Pi based complete embedded system for iris recognition. 2017 signal processing: algorithms, architectures, arrangements, and applications (SPA). New Jersey, USA: IEEE, 263–268, https://doi.org/10.23919/ SPA.2017.8166876.
14. Wang, Z., Li, C., Shao, H. & Sun, J. (2018). Eye recognition with mixed convolutional and residual network (MiCoRe-Net). IEEE Access, 6, 17905–17912. https://doi.org/10.1109/ACCESS.2018.2812208.
15. Klein, A., Falkner, S., Bartels, S., Hennig, P. & Hutter, F. (2016). Fast bayesian optimization of machine learning hyperparameters on large datasets. arXiv preprint arXiv:1605.07079.
16. Bazrafkan, S., Thavalengal, S. & Corcoran, P. (2018). An end to end Deep Neural Network for iris segmentation in unconstrained scenarios. Neural Networks, 106, 79–95, https://doi.org/10.1016/j.neunet.2018.06.011.
17. Arsalan, M., Naqvi, R. A., Kim, D. S., Nguyen, P.H., Owais, M. & Park, K. R. (2018). IrisDenseNet: Robust iris segmentation using densely connected fully convolutional networks in the images by visible light and near-infrared light camera sensors. Sensors, 18(5), 1501, https://doi. org/10.3390/s18051501.
18. LeCun, Y., Haffner, P., Bottou, L. & Bengio, Y. (1999). Object recognition with gradient-based learning. In Shape, contour and grouping in computer vision. Lecture notes in computer science, vol. 1681. Berlin: Springer, 319–345, https://doi.org/10.1007/3-540-46805-6_19.
19. Khan, S., Rahmani, H., Shah, S. A. A. & Bennamoun, M. (2018). A guide to convolutional neural networks for computer vision. Synthesis lectures on computer vision, vol. 8, no. 1. San Rafael, CA, USA: Morgan & Claypool Publishers, https://doi.org/10.2200/S00822ED1V01Y201712COV015.
20. Ciaburro, G. & Venkateswaran, B. (2017). Neural Networks with R Smart models using CNN, RNN, deep learning, and artificial intelligence principles. Birmingham, UK: Packt Publishing.
21. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., et al. (2014). ImageNet large scale visual recognition challenge. Int. J. Comput. Vis., 115(3), 211–252. https://doi.org/10.1007/s11263-015-0816-y.
22. Nguyen, K., Fookes, C., Ross, A. & Sridharan, S. (2017). Iris recognition with off-the-shelf CNN features: A deep learning perspective. IEEE Access, 6, 18848–18855, https://doi.org/10.1109/ACCESS.2017.2784352.
23. Kim, P. (2017). MATLAB deep learning with machine learning, neural networks and artificial intelligence. New York, USA: Apress.