JOURNAL OF DATA SCIENCE | Vol.2025, Issue 1, No.6
eISSN:2805-5160

Herbal Plant Identification Using Deep Learning

Md. Fouziya[1*], Ch. Divija[2], K. Lakshmi Priyanka[3], G. Swathi[4], N. Revathi[5]

[1,2,3,4,5] Vignan's Institute of Management and Technology for Women (VMTW), India

Email: fouziya@gmail.com[1*], divija3698@gmail.com[2], Khandepriyanka2004@gmail.com[3], 8500461078@gmail.com[4], narsingrevathi@gmail.com[5]

Abstract

From traditional medicine to today's research in pharmacology, herbal plants are seen as very important. Yet, correctly identifying herbal species is challenging since many species share the same features and must be classified by experienced taxonomists. Technological advances such as deep learning have provided a way to automate this work with improved accuracy. The proposed system identifies herbal plants by analyzing their images using Convolution Neural Networks (CNNs), which are known for being effective in computer vision. To ensure the dataset is strong, I used thousands of clear leaf pictures from various herbal plant species that were taken in many environmental settings. Before training, the images were processed in stages by normalizing them, creating variations, and separating important objects. To find the most suitable CNN, VGG16, ResNet50, and MobileNetV2 were assessed based on their accuracy, how efficient they are, and whether they could be used on mobile phones. By using transfer learning, the model could take advantage of previously trained models on huge image collections.
.

Keywords

Herbal Plant Identification, Deep Learning, Convolution Neural Networks, Image Classification, Plant Recognition, Computer Vision, Medicinal Plants, Automated Identification, Leaf Image Analysis, Artificial Intelligence in Botany

Introduction

Herbal plants have been an essential component of traditional medicine systems such as Ayurveda, Traditional Chinese Medicine, and African ethno medicine for centuries. Despite their importance, accurate identification of herbal species remains a significant challenge, especially in regions with rich biodiversity and overlapping morphological traits among plants. Misidentification can lead to ineffective or even harmful applications in medicinal and agricultural contexts. Traditionally, herbal plant identification relies on botanical expertise, field surveys, and morphological comparisons, which are time-consuming, subjective, and not easily scalable. Among these, deep learning, particularly Convolution Neural Networks (CNNs), has emerged as a powerful tool for image-based classification tasks.

Submission: 11 March 2025; **Acceptance**: 14 June 2025; **Available Online:** June 2025

This study explores the potential of deep learning in automating the identification of herbal plants based on leaf images. By leveraging large, annotated datasets and advanced CNN architectures, we aim to build a system capable of recognizing subtle visual cues such as leaf shape, venation patterns, and surface texture. The objective is not only to achieve high accuracy but also to create a solution that can be adapted for practical use in mobile applications, field research, and educational platforms.

In addition to improving identification accuracy, this research seeks to contribute to the digitization and preservation of traditional plant knowledge. As biodiversity continues to face threats from habitat loss and climate change, technologies like deep learning can play a pivotal role in documenting and protecting valuable plant species. This project thus sits at the intersection of technology, ecology, and medicine—an effort to blend innovation with conservation. High interpretability, which allows medical personnel to understand the main risk factors that contribute to diagnosis (Linardatos & Papastefanopoulos,2021).

On the other hand, Naive Bayes offers an efficient probabilistic approach in processing small datasets, assuming independence between features (Chen et al., 2020). Meanwhile, Random Forest stands out in handling complex relationships between variables with a high degree of accuracy, making it strong.

Figure 1. Sample of Dataset

Herbal plants represent a vast and largely untapped reservoir of bioactive compounds that have been utilized for centuries in traditional medicine and continue to be of immense interest in modern pharmacological research. In many rural and indigenous communities, herbal remedies serve as the primary source of healthcare, while in industrialized nations, there is a growing demand for plant-based and organic therapeutic alternatives. Despite the critical role these plants play, the accurate identification and classification of herbal species remain a complex task due to their morphological similarities, regional variations, and changing physical characteristics across growth stages and seasons.

As a final step, the entire process of model training and evaluation will be rigorously documented to ensure reproducibility of results and fair comparison between models. Thus, this research not only contributes to selecting the most accurate ML model for detecting lung cancer but also provides deeper insights into how artificial intelligence can change the landscape of medical diagnosis in the future. If applied correctly, this technology could potentially save more lives by providing faster, cheaper, and more accurate early detection than current conventional methods.

Material and Method

In this research, Convolutional Neural Networks (CNNs) form the main theoretical part used for labeling different herbal plants. It explains the concepts and technologies that are most important for the methodology.

Deep learning is a section of machine learning using multi-layered artificial neural networks to learn from complicated data. Contrary to ordinary machine learning, deep learning learns key features from raw images independently, giving it a clear advantage for processing images.

- **Convolutional Neural Networks (CNNs):** CNNs are made for handling data arranged in a grid-like pattern such as images. Most CNN architecture designs have multiple layers inside them.
- Convolutional Layers: Apply filters on the input image to detect areas such as edges, different textures, and shapes across the image.
- In most cases, ReLU (Rectified Linear Unit) is used to make the network non-linear.
- Downsizing feature maps by pooling layers allows us to focus on important points (max pooling is an example).
- High-level reasons and makes predictions after the extraction of features in the first layers.
- Softmax Layer: Transforms the network's output into estimates of the probability that the input belongs to a particular class.
- Plant identification with CNNs is reliable since they are able to detect delicate aspects of leaf shapes like veins, borders, and other textures.
- Transfer Learning: By using a well-trained model, we can tackle a different job. For this study, ResNet, MobileNet and InceptionV3 models, trained on ImageNet, are fine-tuned using a dataset of herbal leaves. Taking this approach leads to faster training and helps models perform well, mainly when the data collection is not large.

Image Processing Techniques:

- Preprocessing of images is used to increase the accuracy of deep learning models.
  With normalization, the pixel values are always in a consistent range.
  By augmenting data, we can use simple transformations to improve the ability to generalize.
  By segmenting the leaf from the background, noise is reduced.

The performance is checked for the model is checked using recognized classification measurements. This is the number of predictions got right over the number predicted in total.

- Precision & Recall tell how many of the suggested plants are important and how many of the important ones are recognized.

- F1 Score: It measures average precision and recall.
  A Confusion Matrix shows the number of true positives, false positives, false negatives, and true negatives.
- Identifying herbal plants helps to save traditional medicine, prevent any risks and assist in botanical science. To identify plants, one would typically rely on their looks and the expertise of others which can be both time-taking and incorrect. In recent times, CNNs have become important as they can accurately and quickly classify different types of plants in images.

Deep learning models are created by studying many pictures of herbal plants, so they can recognize subtle changes in things like leaf color and shape, the arrangement of veins and plant texture. They perform better than standard image processing techniques since they extract the essential features without needing humans to select them. After being trained, a model using deep learning can identify different plant species very precisely, even when they look similar. Consequently, deep learning has become very helpful in applications and tools needed for rapid identification on the go.

Despite its advantages, deep learning-based herbal plant identification faces some challenges. The quality and diversity of training data significantly influence model performance—poor lighting, occlusion, or background noise can reduce accuracy. Additionally, the scarcity of labeled datasets for rare or regional medicinal plants can limit the model's generalization ability. Ongoing research is focusing on improving dataset quality, leveraging transfer learning, and integrating multi-modal data (such as text and chemical profiles) to enhance the reliability and scope of deep learning in herbal plant identification.

Further enhancing accuracy, researchers have started incorporating advanced techniques such as transfer learning and data augmentation. Transfer learning allows models pre-trained on large datasets like ImageNet to be fine-tuned for specific herbal plant datasets, reducing the need for massive amounts of training data. Data augmentation techniques—such as rotation, scaling, and flipping—help in increasing the robustness of the model by exposing it to varied image conditions. These techniques help reduce overfitting and improve performance, particularly when dealing with limited data for certain plant species.

Moreover, deep learning is being integrated with mobile and cloud computing platforms to make herbal plant identification more accessible in real-time scenarios. Mobile applications embedded with trained models enable users, such as farmers, herbalists, and researchers, to identify plants on the spot using just a smartphone camera. This not only democratizes access to plant knowledge but also contributes to the creation of large, crowd-sourced datasets that can further train and refine existing models. As technology evolves, the combination of artificial intelligence and traditional botany is expected to revolutionize herbal plant research and application.

In addition to visual features, recent developments have explored multimodal approaches that combine image data with textual descriptions, GPS metadata, and environmental information. For example, pairing leaf images with known habitat locations or medicinal uses can improve the context-awareness of the deep learning model. This fusion of data types helps the model make more informed predictions, especially when visual similarities between different species pose a challenge. Such hybrid systems enhance both the accuracy and reliability of identification, making them suitable for practical applications in agriculture, medicine, and biodiversity conservation.

Furthermore, deep learning plays a critical role in large-scale plant biodiversity monitoring and conservation efforts. Automated identification systems can process thousands of images from field surveys, enabling faster cataloging of medicinal plants in various regions. This is particularly valuable in regions with rich but under-documented plant diversity. By supporting large-scale documentation, deep learning aids in the discovery of new plant species, supports ethno-botanical research, and helps prevent the misuse of lookalike toxic plants. As datasets grow and models become more sophisticated, the future of herbal plant identification will likely rely heavily on artificial intelligence to bridge the gap between traditional knowledge and cutting-edge technology.

Result and Discussion

The performance of various machine learning and deep learning models was evaluated for the task of herbal plant identification. The models tested include CNN, ResNet50, MobileNet, SVM, and Random Forest. Evaluation metrics such as F1-score for Class 0 and Class 1, accuracy, macro average, and weighted average were used to compare model performance.

Among the models, ResNet50 achieved the highest overall performance with a Class 0 F1-score of 0.92, Class 1 F1-score of 0.90, and an accuracy of 91%. CNN also performed well with an accuracy of 88%, followed by Mobile Net with 86%. Classical machine learning models like SVM and Random Forest showed comparatively lower accuracy (79% and 82%, respectively), indicating the superiority of deep learning techniques in handling complex visual patterns in herbal plant images.

The CNN model also performed well, with an accuracy of 88%, a Class 0 F1-score of 0.89, and a Class 1 F1-score of 0.86. Its macro and weighted averages (0.875 and 0.877, respectively) reflect strong generalization, though slightly lower than ResNet50. Mobile Net, a lightweight deep learning model, achieved an accuracy of 86%, which is slightly lower than CNN but still notable given its efficiency and lower computational requirements.

On the other hand, traditional machine learning models such as SVM and Random Forest had lower accuracy rates of 79% and 82%, respectively. Their F1-scores for both classes were also comparatively lower, indicating challenges in distinguishing between visually similar herbal species using manually engineered features. However, they still performed decently and may be suitable for applications with limited computational resources or smaller datasets.

In summary, the results clearly show that deep learning models, particularly those utilizing transfer learning, outperform classical models in herbal plant identification tasks, especially when high accuracy and generalization are required features, until the desired number of features is achieved. Both PCA and RFE are applied to the dataset to evaluate their impact on model performance.

The results from the study highlight the significant advantages of deep learning models, particularly ResNet50 and CNN, over traditional machine learning algorithms like SVM and Random Forest. ResNet50, a pre-trained model based on transfer learning, outperformed all other models, achieving the highest accuracy (91%) and F1- scores (Class 0: 0.92, Class 1: 0.90). This high performance can be attributed to the depth and complexity of ResNet50's architecture, which leverages residual connections to mitigate the vanishing gradient problem and allows the model to capture intricate features in herbal plant images. The success of transfer learning is particularly evident here, as ResNet50, trained on a large and diverse dataset (ImageNet), could be fine-tuned with relatively smaller amounts of plant-specific data, making it an ideal choice for scenarios where acquiring large labeled datasets may be challenging.

In comparison, CNN also showed robust performance, with an accuracy of 88% and F1-scores of 0.89 for Class 0 and 0.86 for Class 1. Although CNN did not perform as well as ResNet50, it still provided a reliable classification of herbal plant species. This suggests that CNNs, with appropriate architectures and sufficient training, can effectively handle the complexity of plant images, even without the pre-trained weights that ResNet50 benefits from. The macro average (0.875) and weighted average (0.877) of CNN further confirm that it maintained good performance across both classes, though with slightly lower precision and recall for Class 1.

The performance of Mobile Net, while slightly lower than CNN, emphasizes its suitability for environments where computational efficiency is critical. With an accuracy of 86%, Mobile Net balances performance with lower resource requirements, making it an excellent option for mobile applications or edge devices that need to perform plant identification in real-time. Despite this efficiency, its lower performance compared to the deeper models suggests that, while lightweight models are effective in certain scenarios, they may still struggle with more complex plant recognition tasks that require deeper feature extraction capabilities.

The SVM and Random Forest models, both of which rely on manually extracted features, showed the weakest performance with accuracy scores of 79% and 82%, respectively. These models likely struggled due to the complex and subtle visual features that define different herbal species, which traditional feature extraction methods may not capture as effectively as deep learning models. SVM, in particular, is sensitive to the choice of kernel and hyperparameters, and in this case, it was likely unable to sufficiently distinguish between the classes based on the limited features provided. Random Forest, while a more robust ensemble method, similarly had difficulty with complex image features and benefited less from the lack of feature engineering compared to deep learning models.

Despite their lower performance, SVM and Random Forest models still have value in specific contexts, such as when computational resources are limited or when working with small datasets. They could also serve as useful baseline models, providing a benchmark for evaluating more complex techniques. However, the results indicate that deep learning models, particularly those utilizing transfer learning and convolutional networks, are far superior when it comes to handling image-based herbal plant identification tasks.

Despite their lower performance, SVM and Random Forest models still have value in specific contexts, such as when computational resources are limited or when working with small datasets. They could also serve as useful baseline models, providing a benchmark for evaluating

more complex techniques. However, the results indicate that deep learning models, particularly those utilizing transfer learning and convolutional networks, are far superior when it comes to handling image-based herbal plant identification tasks.

In real-world applications, deep learning models like ResNet50 and CNN can revolutionize herbal plant identification by enabling fast, accurate, and scalable identification of medicinal plants. These models are not only capable of distinguishing between species but can also be integrated into mobile apps, aiding practitioners, farmers, and researchers in identifying plants on the spot, thus contributing to more efficient and informed decision-making in agriculture, herbal medicine, and biodiversity conservation.

Moreover, the macro and weighted average F1-scores from the deep learning models suggest that they are robust against class imbalances. This is important because many plant species may not be equally represented in datasets, yet these models were able to maintain balanced performance across classes. This characteristic could be crucial for applications where both classes (e.g., medicinal vs. non-medicinal plants) are equally important.

As the research progresses, expanding the dataset to include more diverse plant species, improving data augmentation techniques, and refining the models with more advanced architectures can further enhance performance.

Incorporating multimodal data, such as textual descriptions, environmental factors, and even chemical analysis, could also enrich the models' ability to identify plants under various conditions, further bridging the gap between traditional botanical knowledge and modern AI technologies.

This image is another screenshot from the "Herbal Plant-AI" web application, which uses artificial intelligence to identify herbal plants and explain their medicinal benefits. The latest includes a green leafy background, a purple navigation bar with a "Home" link, and a section labeled "Result" where the plant's details are displayed.

In the result, the plant identified is Clove. The description states that clove is used to treat digestive disorders, respiratory disorders, and toothache. This reflects the traditional medicinal uses of clove, which is known for its anti-inflammatory, antibacterial, and analgesic properties. The interface is simple and user-friendly, presenting the result in a clear and informative manner for users interested in natural remedies.

Conclusion

The integration of deep learning techniques, specifically Convolutional Neural Networks (CNNs) and transfer learning models like ResNet50, marks a significant advancement in the field of herbal plant identification. The results of this study demonstrate that deep learning models outperform traditional machine learning algorithms such as SVM and Random Forest in terms of accuracy, F1-scores, and overall reliability. With ResNet50 achieving the highest accuracy of 91%, followed closely by CNN models, these approaches have proven to be highly effective for classifying herbal plant species based on complex visual features. The ability of these models to capture subtle, intricate patterns, such as leaf morphology and color variations, has established them as the most reliable tools for plant identification.

The deep learning approach offers several key advantages, including automatic feature extraction, which eliminates the need for manual intervention in feature selection, and robustness to variations in lighting, background, and plant orientation. This is particularly valuable in real-world applications where plant images may not be ideal. Transfer learning, leveraging pre-trained models such as ResNet50, also demonstrated its effectiveness by allowing for accurate identification with relatively smaller and region-specific datasets. This ability to perform well even with limited data makes deep learning models particularly valuable in resource-constrained environments, where acquiring large-scale, labeled datasets might be challenging.

Despite all Deep Learning models, the study also acknowledges the importance of classical machine learning algorithms like SVM and Random Forest in certain contexts. While these models did not perform as well as the deep learning models, they still serve as valuable baseline tools, especially in settings with limited computational power or smaller datasets. Their efficiency and interpretability could be advantageous in applications where resources are constrained and where simplicity and speed are more critical than high accuracy.

Looking ahead, the potential for deep learning in herbal plant identification is vast. Models like ResNet50 and CNN have shown promise in improving the accuracy of plant identification systems for practical applications in fields such as agriculture, herbal medicine, and biodiversity conservation. The next steps for further enhancement could involve the integration of multimodal data, such as chemical profiles, environmental data, and textual descriptions, which would provide richer contextual information and allow for more accurate identification under diverse conditions. Additionally, expanding the dataset to include a broader variety of plant species, especially those with similar features, will continue to improve the model's robustness and generalizability.

Ultimately, real-time, mobile-based plant identification tools powered by deep learning have the potential to revolutionize the way farmers, researchers, medicinal practitioners, and even biodiversity conservationists interact with plant species. The combination of cutting-edge AI and traditional knowledge is bound to create more accessible, reliable, and effective methods for identifying and utilizing herbal plants. As technology continues to evolve, these models will not only become more accurate but also more efficient and accessible, enabling widespread use and contributing to the preservation of herbal plant knowledge globally.

Acknowledgement

We would like to express our heartfelt gratitude to everyone who supported and guided us throughout the completion of our project titled “Herbal Plant Identification Using Deep Learning”. We are especially thankful to Ms. Md.Fouziya, Assistant Professor, Department of Information Technology, for her expert guidance, continuous encouragement, and valuable feedback, which were instrumental in shaping this work. We also extend our sincere thanks to our co-authors Ms. Ch. Divija, Ms. K. Lakshmi Priyanka, Ms. G. Swathi, and Ms. N. Sahithi for their collaboration, support, and dedication throughout the project.

References

Arava, K., Paritala, C., Shariff, V., Praveen, S. P., & Madhuri, A. (2022). A generalized model for identifying fake digital images through the application of deep learning. *2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC)*, Coimbatore, India, pp. 1144-1147. https://doi.org/10.1109/ICESC54411.2022.9885341

Bulla, S., Basaveswararao, B., Rao, K. G., Chandan, K., & Swamy, S. R. (2022). A secure new HRF mechanism for mitigate EDoS attacks. *International Journal of Ad Hoc and Ubiquitous Computing, 40*(1-3), 20-29. http://dx.doi.org/10.1504/IJAHUC.2022.123524

Chamundeeswari, V. V., Sundar, V. S. D., Mangamma, D., Azhar, M., Kumar, B. S. S. P., & Shariff, V. (2024). Brain MRI analysis using CNN-based feature extraction and machine learning techniques to diagnose Alzheimer's disease. *2024 First International Conference on Data, Computation and Communication (ICDCC)*, Sehore, India, pp. 526-532. https://doi.org/10.1109/ICDCC62744.2024.10961923

Chitti, S., et al. (2019). Design, synthesis and biological evaluation of 2-(3, 4-dimethoxyphenyl)-6 (1, 2, 3, 6-tetrahydropyridin-4-yl) imidazo [1, 2-a] pyridine analogues as antiproliferative agents. *Bioorganic & Medicinal Chemistry Letters, 29*(18), 2551-2558.

Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture, 145*, 311–318. https://doi.org/10.1016/j.compag.2018.01.009

Grinblat, G. L., Uzal, L. C., Larese, M. G., & Granitto, P. M. (2016). Deep learning for plant identification using vein morphological patterns. *Computers and Electronics in Agriculture, 127*, 418–424. https://ri.conicet.gov.ar/bitstream/11336/52668/11/CONICET_Digital_Nro.bfec2c01-4dbd-4407-be27-54e24e26bb62_L.pdf

Jabassum, A., Ramesh, J. V. N., Sundar, V. S. D., Shiva, B., Rudraraju, A., & Shariff, V. (2024). Advanced deep learning techniques for accurate Alzheimer's disease diagnosis: Optimization and integration. *2024 4th International Conference on Sustainable Expert Systems (ICSES)*, Kaski, Nepal, pp. 1291-1298. https://doi.org/10.1109/ICSES63445.2024.10763340

Kodete, C. S., Pasupuleti, V., Thuraka, B., Gayathri, V. V., Sundar, V. S. D., & Shariff, V. (2024). Machine learning for enabling strategic insights to future-proof E-Commerce. *2024 5th International Conference on Smart Electronics and Communication (ICOSEC)*, Trichy, India, pp. 931-936. https://doi.org/10.1109/ICOSEC61587.2024.10722255

Kodete, C. S., Pasupuleti, V., Thuraka, B., Sangaraju, V. V., Tirumanadham, N. S. K. M. K., & Shariff, V. (2024). Robust heart disease prediction: A hybrid approach to feature selection and model building. *2024 4th International Conference on Ubiquitous Computing and Intelligent Information Systems (ICUIS)*, Gobichettipalayam, India, pp. 243-250. https://doi.org/10.1109/ICUIS64676.2024.10866501

Kodete, C. S., Saradhi, D. V., Suri, V. K., Varma, P. B. S., Tirumanadham, N. S. K. M. K., & Shariff, V. (2024). Boosting lung cancer prediction accuracy through advanced data processing and machine learning models. *2024 4th International Conference on Sustainable Expert Systems (ICSES)*, Kaski, Nepal, pp. 1107-1114. https://doi.org/10.1109/ICSES63445.2024.10763338

Kumar, C. S., et al. (2021). An adaptive deep learning model to forecast crimes. In *Proceedings of Integrated Intelligence Enable Networks and Computing: IIENC 2020*. Springer Singapore. https://arxiv.org/html/2407.19324v1

Lee, S. H., Chan, C. S., Mayo, S. J., & Remagnino, P. (2017). How deep learning extracts and learns leaf features for plant classification. *Pattern Recognition*, *71*, 1–13. https://doi.org/10.1016/j.patcog.2017.05.015

Mohan, V. M., et al. (2010). Mass transfer correlation development for the presence of entry region coil as swirl promoter in tube. *International Journal of Thermal Sciences*, *49*(2), 356-364. https://www.ijert.org/mass-transfer-studies-at-the-inner-wall-of-an-annular-conduit-in-the-presence-of-fluidizing-solids-with-coaxially-placed-spiral-coil-as-turbulence-promoter

Nagasri, D., Swamy, S. R., Amareswari, P., Bhushan, P. V., & Raza, M. A. (2024). Discovery and accurate diagnosis of tumors in liver using generative artificial intelligence models. *Journal of Next Generation Technology (ISSN: 2583-021X)*, *4*(2). https://www.researchgate.net/publication/381613787_Discovery_and_Accurate_Diagnosis_of_Tumors_in_Liver_using_Generative_Artificial_Intelligence_Models

N. S. Koti Mani Kumar Tirumanadham et al. (2025). Boosting student performance prediction in e-learning: A hybrid feature selection and multi-tier ensemble modelling framework with federated learning. *Journal of Theoretical and Applied Information Technology*, *103*(5). https://www.mecs-press.org/ijmecs/ijmecs-v17-n2/v17n2-3.html

Pasupuleti, V., Thuraka, B., Kodete, C. S., Priyadarshini, V., Tirumanadham, K. M. K., & Shariff, V. (2024). Enhancing predictive accuracy in cardiovascular disease diagnosis: A hybrid approach using RFAP feature selection and random forest modeling. *2024 4th International Conference on Soft Computing for Security Applications (ICSCSA)*, Salem, India, pp. 42-49. https://doi.org/10.1109/ICSCSA64454.2024.00014

Praveen, S. P., Jyothi, V. E., Anuradha, C., Venugopal, K., Shariff, V., & Sindhura, S. (2022). Chronic kidney disease prediction using ML-Based Neuro-Fuzzy model. *International Journal of Image and Graphics*. https://doi.org/10.1142/s0219467823400132

Praveen, S. P., et al. (2025). AI-powered diagnosis: Revolutionizing healthcare with neural networks. *Journal of Theoretical and Applied Information Technology*, *101*(3). https://www.jatit.org/volumes/Vol103No3/16Vol103No3.pdf

Rajkumar, K. V., Nithya, K. S., Narasimha, C. T. S., Shariff, V., Manasa, V. J., & Tirumanadham, N. S. K. M. K. (2024). Scalable web data extraction for Xtree analysis: Algorithms and performance evaluation. *2024 Second International Conference on Inventive Computing and Informatics (ICICI)*, Bangalore, India, pp. 447-455. https://doi.org/10.1109/ICICI62254.2024.00079

S., S., Kodete, C. S., Velidi, S., Bhyrapuneni, S., Satukumati, S. B., & Shariff, V. (2024). Revolutionizing healthcare: A comprehensive framework for personalized IoT and cloud computing-driven healthcare services with smart biometric identity management. *Journal of Intelligent Systems and Internet of Things*, *13*(1), 31–45. https://doi.org/10.54216/jisiot.130103

S., S., Raju, K. B., Praveen, S. P., Ramesh, J. V. N., Shariff, V., & Tirumanadham, N. S. K. M. K. (2025b). Optimizing diabetes diagnosis: HFM with tree-structured Parzen estimator for enhanced predictive performance and interpretability. *Fusion Practice and Applications*, *19*(1), 57–74. https://doi.org/10.54216/fpa.190106

Shariff, V., Aluri, Y. K., & Reddy, C. V. R. (2019b). New distributed routing algorithm in wireless network models. *Journal of Physics: Conference Series*, *1228*(1), 012027. https://doi.org/10.1088/1742-6596/1228/1/012027

Shariff, V., Paritala, C., & Ankala, K. M. (2025). Optimizing non small cell lung cancer detection with convolutional neural networks and differential augmentation. *Scientific Reports*, *15*(1). https://doi.org/10.1038/s41598-025-98731-4

Sirisati, R. S., Kumar, C. S., & Latha, A. G. (2021). An efficient skin cancer prognosis strategy using deep learning techniques. *Indian Journal of Computer Science and Engineering (IJCSE)*, *12*(1). https://www.researchgate.net/publication/349469992_An_efficient_skin_cancer_prognosis_strategy_using_deep_learning_techniques

Sirisati, R. S., Kumar, C. S., Latha, A. G., Kumar, B. N., & Rao, K. S. (2021). An enhanced multi layer neural network to detect early cardiac arrests. *2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, pp. 1514-1518. IEEE. https://ieeexplore.ieee.org/document/9531126/

Sirisati, R. S., Kumar, C. S., Latha, A. G., Kumar, B. N., & Rao, K. S. (2021). Identification of Mucormycosis in post Covid-19 case using Deep CNN. *Turkish Journal of Computer and Mathematics Education*, *12*(9), 3441-3450. https://turcomat.org/index.php/turkbilmat/article/download/11302/8362/20087

Sirisati, R. S., Prasanthi, K. G., & Latha, A. G. (2021). An aviation delay prediction and recommendation system using machine learning techniques. In *Proceedings of Integrated Intelligence Enable Networks and Computing: IIENC 2020* (pp. 239-253). Springer Singapore. https://www.researchgate.net/publication/370995631_Flight_Delay_Prediction_System_in_Machine_Learning_using_Support_Vector_Machine_Algorithm/fulltext/646e430a37d6625c002e31c1/Flight-Delay-Prediction-System-in-Machine-Learning-using-Support-Vector-Machine-Algorithm.pdf

Sirisati, R. S., Venuthurumilli, P., Ranjith, J., & Rao, K. S. (2023). Cancer Sight: Illuminating the Hidden-Advancing Breast Cancer Detection with Machine Learning-Based Image Processing Techniques. *2023 International Conference on Sustainable Communication Networks and Application (ICSCNA)*, pp. 1618-1625. IEEE. https://doi.org/10.1109/ICSCNA58489.2023.10370462

Sirisati, R. S., et al. (2024). A deep learning framework for recognition and classification of diabetic retinopathy severity. *Telematique*, *23*(01), 228-238. https://www.frontiersin.org/journals/medicine/articles/10.3389/fmed.2025.1551315/abstract

Sirisati, R. S., et al. (2024). Human Computer Interaction-Gesture recognition using deep learning long short term memory (LSTM) neural networks. *Journal of Next Generation Technology (ISSN: 2583-021X)*, *4*(2). https://www.aasmr.org/jsms/Vol14/No.1/Vol.14%20No.1.32.pdf

Sirisati, R. S., Kalyani, A., Rupa, V., Venuthurumilli, P., & Raza, M. A. (2024). Recognition of counterfeit profiles on communal media using machine learning artificial neural networks & Support Vector Machine algorithms. *Journal of Next Generation Technology (ISSN: 2583-021X)*, *4*(2). https://www.researchgate.net/publication/381613824_Recognition_of_Counterfeit_Profiles _on_Communal_Media_using_Machine_Learning_Artificial_Neural_Networks_Support_Vector_Machine_Algorithms

Swaroop, C. R., et al. (2024). Optimizing diabetes prediction through intelligent feature selection: A comparative analysis of Grey Wolf Optimization with AdaBoost and Ant Colony Optimization with XGBoost. *Algorithms in Advanced Artificial Intelligence: ICAAAI-2023*, *8*(311).

Swamy, S. R., & Mandapati, S. (2018). A rule selected fuzzy energy & security aware scheduling in cloud. *Journal of Theoretical & Applied Information Technology*, *96*(10). https://informatica.si/index.php/informatica/article/download/5741/3358

Swamy, S. R., Rao, P. S., Raju, J. V. N., & Nagavamsi, M. (2019). Dimensionality reduction using machine learning and big data technologies. *Int. J. Innov. Technol. Explor. Eng.(IJITEE)*, *9*(2), 1740-1745. https://www.researchgate.net/publication/364081288_Dimensionality_Reduction_using_Machine_Learning_and_Big_Data_Technologies

Swamy, S. R., et al. (2023). Multi-Features Disease Analysis Based Smart Diagnosis for COVID-19. *Computers, Systems & Science and Engineering*, *45*(1), 869-886.

Thatha, V. N., Chalichalamala, S., Pamula, U., Krishna, D. P., Chinthakunta, M., Mantena, S. V., Vahiduddin, S., & Vatambeti, R. (2025b). Optimized machine learning mechanism for big data healthcare system to predict disease risk factor. *Scientific Reports*, *15*(1). https://doi.org/10.1038/s41598-025-98721-6

Thuraka, B., Pasupuleti, V., Kodete, C. S., Chigurupati, R. S., Tirumanadham, N. S. K. M. K., & Shariff, V. (2024). Enhancing diabetes prediction using hybrid feature selection and ensemble learning with AdaBoost. *2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, Kirtipur, Nepal, pp. 1132-1139. https://doi.org/10.1109/I-SMAC61858.2024.10714776.

Thuraka, B., Pasupuleti, V., Kodete, C. S., Naidu, U. G., Tirumanadham, N. S. K. M. K., & Shariff, V. (2024). Enhancing predictive model performance through comprehensive pre-processing and hybrid feature selection: A study using SVM. *2024 2nd International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS)*, Erode, India, pp. 163-170. https://doi.org/10.1109/ICSSAS64001.2024.10760982

Tirumanadham, N. S. K. M. K., Priyadarshini, V., Praveen, S. P., Thati, B., Srinivasu, P. N., & Shariff, V. (2025d). Optimizing lung cancer prediction models: A hybrid methodology using GWO and Random Forest. In *Studies in computational intelligence* (pp. 59–77). https://doi.org/10.1007/978-3-031-82516-3_3

Uğuz, H., & Yildirim, A. (2021). Classification of medicinal plants using deep learning algorithms. *Engineering Science and Technology, an International Journal*, *24*(6), 1319–1327. https://doi.org/10.1016/j.jestch.2021.05.006

Vahiduddin, S., Chiranjeevi, P., & Mohan, A. K. (2023). An analysis on advances in lung cancer diagnosis with medical imaging and deep learning techniques: Challenges and opportunities. *Journal of Theoretical and Applied Information Technology*, *101*(17). https://www.google.com/search?q=http://www.jatit.org/volumes/Vol101No17/Vol101No17.pdf

Venkata Naga Ramesh, J., & Dedeepya, P., Chiranjeevi, P., Narasimha, V., Shariff, V., Ranjith, J. (2023). Image recognition and similarity retrieval with Convolutional Neural Networks.

2023 2nd International Conference on Automation, Computing and Renewable Systems (ICACRS), Pudukkottai, India, pp. 709-716. https://doi.org/10.1109/ICACRS58579.2023.10404664.

Venkata Narasimha, V., R. R. T., Kadiyala, R., Paritala, C., Shariff, V., & Rakesh, V. (2024). Assessing the resilience of machine learning models in predicting long-term breast cancer recurrence results. *2024 8th International Conference on Inventive Systems and Control (ICISC)*, Coimbatore, India, pp. 416-422. https://doi.org/10.1109/ICISC62624.2024.00077

Veerapaneni, E. J., Babu, M. G., Sravanthi, P., Geetha, P. S., Shariff, V., & Donepudi, S. (2024). Harnessing Fusion's potential: a State-of-the-Art information security architecture to create a big data analytics model. In *Lecture notes in networks and systems* (pp. 545–554). https://doi.org/10.1007/978-981-97-6106-7_34

Wäldchen, J., & Mäder, P. (2018). Plant species identification using computer vision techniques: A systematic literature review. *Archives of Computational Methods in Engineering*, *25*(2), 507–543. https://link.springer.com/article/10.1007/s11831-016-9206-z

Yarra, K., Vijetha, S. L., Rudra, V., Balunaik, B., Ramesh, J. V. N., & Shariff, V. (2024). A dual-dataset study on deep learning-based tropical fruit classification. *2024 8th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, Coimbatore, India, pp. 667-673. https://doi.org/10.1109/ICECA63461.2024.10800915