Original Article
ISSN (Online): 2582-7472

ShodhKosh: Journal of Visual and Performing Arts
May 2024 5(5), 397–403

A SYSTEMATIC ANALYSIS OF CURRENT DEVELOPMENTS AND POTENTIAL CHALLENGES IN APPLIED DEEP LEARNING-BASED SEED YIELD PREDICTION

Pawan Kumar Gupta[1], Dr. Akhilesh A. Waoo[2]

[1]Department of Computer Science Engineering, AKS University, Satna

Corresponding Author
Akhilesh A. Waoo,
akhileshwaoo@gmail.com

DOI
10.29121/shodhkosh.v5.i5.2024.1890

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ABSTRACT

In recent years, numerous studies have explored applying DL-based approaches for seed yield prediction across various crops, including wheat, maize, soybean, and rice. These studies have demonstrated the efficacy of DL models, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and deep belief networks (DBNs), in capturing intricate relationships between environmental factors, agronomic practices, and seed yield outcomes (Waoo and Soni 2023). By leveraging large-scale datasets, high-resolution imagery, and advanced computational techniques, DL-based seed yield prediction models have achieved remarkable performance improvements compared to traditional methods.

Keywords: Deep Learning, Seed Yield Prediction, Agriculture, Crop Management, Precision Agriculture

1. INTRODUCTION

Seed yield prediction is crucial in modern agriculture, aiding farmers and stakeholders in making informed decisions regarding crop management, resource allocation, and harvest planning. Traditionally, seed yield prediction relied on empirical models and historical data, which often lacked accuracy and robustness, particularly in dynamic agricultural environments. However, with the advent of deep learning (DL) techniques, there has been a paradigm shift in seed yield prediction research, enabling the development of sophisticated models capable of analyzing complex spatial and temporal data with unprecedented accuracy.

In recent years, numerous studies have explored applying DL-based approaches for seed yield prediction across various crops, including wheat, maize, soybean, and rice. These studies have demonstrated the efficacy of DL models, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and deep belief networks (DBNs), in capturing intricate relationships between environmental factors, agronomic practices, and seed yield outcomes (Waoo and Soni 2023). By leveraging large-scale datasets, high-resolution imagery, and advanced computational techniques, DL-based seed yield prediction models have achieved remarkable performance improvements compared to traditional methods.

Despite the promising advancements, several challenges persist in the practical implementation of DL-based seed yield prediction systems. Data scarcity, particularly in regions with limited access to agricultural data, poses a significant obstacle to model development and validation. Additionally, the interpretability of DL models remains a concern, as complex neural networks often operate as "black boxes," hindering the understanding of underlying decision-making processes. Furthermore, scalability issues may arise when deploying DL models in resource-constrained environments, necessitating efficient computational infrastructures and optimization techniques.

Crop yield prediction plays a crucial role in modern agriculture, facilitating informed decision-making for farmers and stakeholders. Traditionally, agricultural yield forecasting relied on statistical models and empirical methods, often limited by their inability to capture complex relationships inherent in agricultural systems. However, with the advent of deep learning techniques, particularly in the domain of seed yield prediction, there has been a paradigm shift towards more accurate and efficient prediction models.

This paper presents a systematic analysis of current developments and potential challenges in applied deep learning-based seed yield prediction. With the rapid advancements in deep learning algorithms, there is an increasing interest in leveraging these techniques to enhance the accuracy and reliability of crop yield forecasts. By systematically reviewing existing literature, this study aims to provide insights into the methodologies, datasets, models, and outcomes of recent research endeavors in this domain.

Through a comprehensive examination of the state-of-the-art approaches, this paper seeks to identify trends, gaps, and opportunities for further advancement in applied deep learning-based seed yield prediction. By understanding the strengths and limitations of existing methodologies, stakeholders in the agricultural sector can make informed decisions regarding the adoption and implementation of deep learning techniques for crop yield forecasting.

Overall, this systematic analysis aims to contribute to the ongoing discourse on the application of deep learning in agriculture, with a specific focus on seed yield prediction. By highlighting current developments and addressing potential challenges, this paper strives to pave the way for future research and innovation in this important area of study.

2. LITERATURE REVIEW

In our research on seed yield prediction, we employed a systematic review methodology to ensure a thorough and unbiased compilation of relevant academic research in this area. The wide range of studies about seed yield prediction highlights the suitability of this approach for our research goals. This systematic methodology allows for a comprehensive examination, critique, and synthesis of studies that address our specific research questions. This includes analyzing the different types of seeds studied, the diverse deep learning frameworks utilized, the primary types of data employed, and the challenges and prerequisites crucial for accurate yield prediction.

Forecasting crop yields pose a significant challenge within precision agriculture, with numerous models having been proposed and validated to date. Addressing this challenge entails leveraging multiple datasets, as crop yield is influenced by a multitude of factors including climate, weather, soil conditions, fertilizer application, and seed type (Xu et al., 2019). This underscores the complexity inherent in crop yield prediction, comprising various intricate stages. Presently, crop yield prediction models can reasonably estimate actual yields, yet there remains a continual pursuit for enhanced performance in yield prediction (Filippi et al., 2019).

Chlingaryan and Sukkarieh conducted a comprehensive review on nitrogen status estimation utilizing machine learning techniques (Chlingaryan et al., 2018). The study concludes that rapid advancements in sensing technologies and machine learning methods are poised to yield cost-effective solutions in the agricultural domain. Elavarasan et al. conducted a survey of literature concerning machine learning models for predicting crop yields based on climatic parameters, advocating for a broader exploration to identify additional parameters influencing crop yield (Elavarasan et al., 2018). Liakos et al. (2018) presented a review paper on the utilization of machine learning in various agricultural domains, including crop management, livestock management, water management, and soil management. Li, Lecourt, and Bishop conducted a review study focusing on the determination of fruit ripeness for optimal harvest timing and yield prediction (Li et al., 2018). Beulah investigated diverse data mining approaches utilized in crop yield prediction, concluding that the application of such techniques holds promise for resolving challenges associated with crop yield prediction (Beulah, 2019).

Adopting a systematic approach is crucial for reducing bias and establishing a transparent and replicable framework for analysis. This level of precision is fundamental to our research objectives, which aim not only to summarize the current state of seed yield prediction but also to identify gaps and emerging trends that can inform future studies. By systematically reviewing the literature, we can provide valuable insights into the advancements, challenges, and opportunities in seed yield prediction using deep learning techniques. In pursuit of this objective, we followed the methodology outlined by the authors Sordello et. al. (2021), who emphasized the Systematic Literature Review (SLR) as a method originally devised for the medical domain in the United Kingdom. The primary aim of an SLR is to ensure transparency in the methodologies employed and to garner collective peer validation of the chosen approach, as stated by the next author. Key elements of an SLR, as highlighted by author Nambiema et. al. (2021), encompass clearly defined objectives, predetermined eligibility criteria, an explicit and replicable methodology, a comprehensive search strategy, rigorous evaluation of the selected studies, and a systematic synthesis and presentation of the characteristics and findings of these studies.

In the review focusing on seed, our objective was to conduct a comprehensive analysis of research endeavors concerning deep learning techniques and their utilization in predicting crop yields. To fulfill this objective, we scrutinized studies from diverse angles. In this segment, we will summarize and emphasize the fundamental aspects of the methodology utilized. A key feature of a systematic literature review is its typical initiation with extensive bibliographic exploration of the subject at hand, aiding in elucidating the research concept, as noted by Siddaway et al. (2019). This initial step aims to enhance comprehension regarding prior publications within the research domain.

Ferhat et. al (2020), the authors employed DL models to detect sunflower seeds and addressed overfitting by utilizing optimization algorithms. They reported that the optimized GoogleNet model achieved an accuracy of 95%. However, this model required human intervention to arrange individual seeds rather than maintaining them in bulk. Additionally, the authors only considered a single viewpoint of seeds during model training, suggesting potential enhancements in model robustness and reliability through training on multiple perspectives.

To tackle the challenges identified in the study by reference (Ferhat et. al., 2020), researchers in Zhao et. al. (2021) examined the entire surface of soybean seeds. They utilized a circumrotating mechanism for full surface detection, achieving an accuracy of 98.87%. Furthermore, they enhanced classification accuracy by utilizing the MobileNet model on a dataset comprising defective seeds.

Furthermore, Zhu et. al. (2020) proposed a method for identifying Soybean seeds, employing pre-trained CNN models such as AlexNet, ResNet18, Xception, Inception-v3, DenseNet201, and NASNetLarge to demonstrate the impact of transfer learning. They claimed that among all models, NASNetLarge reported the highest accuracy of 97.2%. They also suggested that integrating hyperspectral imaging with transfer learning yielded higher accuracy at lower computational costs.

To expand the application of ML models in weed identification, researchers Granitto et. al. (2005) utilized the naïve Bayes algorithm for identifying weed seeds based on their morphological and textural features. The model achieved an accuracy of 98% on grayscale and black-and-white images, although a notable decrease in accuracy was observed for colored images.

The authors Sinthupinyo et. al. (2011) developed an imaging device to address variations in light intensity, reflection, and shadow observed on seeds. Utilizing a Support Vector Machine (SVM) for seed categorization, they achieved an accuracy of 95.6% for intact seeds and 80.6% for defective ones. However, a notable drawback is the misclassification rate of approximately 19%, limiting its practical application. Similarly, researchers Zhang et. al. (2021), Tsuchikawa et. al. (2020), and Nie et. al. (2019) pursued a comparable research path, employing hyperspectral imaging for image capture. They proposed a Deep Convolutional Neural Network (DCNN) model for classifying four types of corn seeds using hyperspectral imaging. Additionally, SVM and K-means algorithms were utilized for seed classification. Comparative analysis revealed that the DCNN-based model achieved the highest training accuracy of 100%. Nevertheless, the system exhibited a considerable number of misclassifications on the testing dataset.

Similarly, researchers Ali et. al. (2020) utilized an MLP machine learning classifier, achieving a classification accuracy of 98.93% for six varieties of corn seeds. Furthermore, Veeramani et. al. (2018), DeepSort for classifying haploid maize seeds, employing SVM, Random Forest (RF), and logistic regression.

Furthermore, Chao et. al. (2019), ResNet, VGGNet, and AlexNet models were employed for the automatic inspection of maize kernels, with ResNet demonstrating superior performance and achieving an accuracy of 98.2%. Likewise, Huang et. al. (2019), AlexNet, VGG-19, GoogleNet, and SVM were applied for classifying maize seeds based on defects, with GoogleNet outperforming all models with an accuracy of 95%.

3. METHODOLOGY

In this paper, we conduct a systematic analysis of current developments and potential challenges in applied DL-based seed yield prediction. We begin by reviewing recent literature on DL-based approaches for seed yield prediction, focusing on model architectures, training methodologies, and performance evaluation metrics. Subsequently, we examine the key datasets utilized in seed yield prediction research, including publicly available repositories, remote sensing data, and field observations. Furthermore, we discuss the advantages and limitations of DL-based seed yield prediction models, highlighting their potential applications and practical considerations.

3.1 Data Collection: The first step in developing a deep learning-based seed yield prediction model is to gather relevant data sources. These may include historical yield data, meteorological records, soil characteristics, crop management practices, and remote sensing imagery. Data collection methods may vary depending on the availability of resources and access to field observations. Publicly available datasets, such as those provided by agricultural research institutions and government agencies, can supplement proprietary data sources and facilitate model training and validation.

3.2 Data Preprocessing: Once the data is collected, it undergoes preprocessing to ensure consistency, quality, and compatibility with the modeling framework. Preprocessing steps may include data cleaning to remove outliers and errors, feature engineering to extract relevant variables, and normalization to scale numerical features. Additionally, data augmentation techniques may be applied to increase the diversity of the dataset and improve model generalization. For remote sensing data, preprocessing techniques such as radiometric calibration, atmospheric correction, and image registration may be employed to enhance data quality and suitability for analysis.

3.3 Model Selection and Architecture Design: The choice of model architecture is a critical aspect of deep learning-based seed yield prediction. Convolutional neural networks (CNNs) are commonly used for analyzing spatial data, such as remote sensing imagery, due to their ability to capture hierarchical features and spatial dependencies. Recurrent neural networks (RNNs), on the other hand, are well-suited for modeling temporal sequences, such as weather patterns and crop phenology. Hybrid architectures, such as convolutional recurrent neural networks (CRNNs), may be employed to integrate spatial and temporal information for improved prediction accuracy.

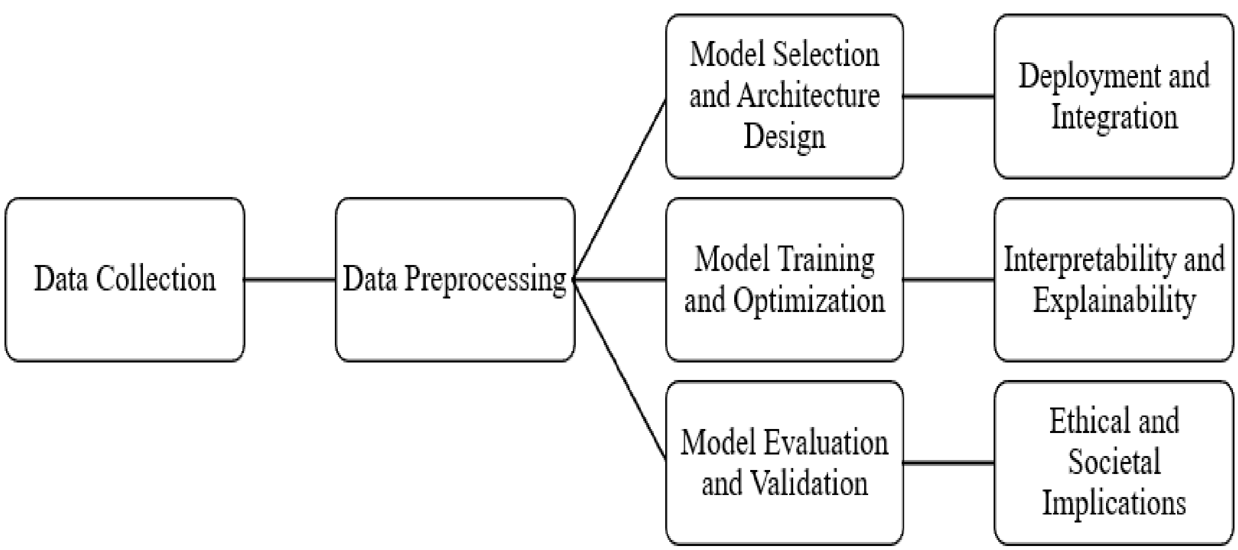

Fig 1: Block diagram illustrating the research methodology for developing a deep learning-based seed yield prediction model

3.4 Model Training and Optimization: Once the model architecture is defined, it is trained using the preprocessed dataset. The dataset is typically split into training, validation, and testing sets to assess model performance and prevent overfitting. During training, optimization algorithms such as stochastic gradient descent (SGD), Adam, or RMSprop are employed to minimize the loss function and update the model parameters iteratively. Hyperparameter tuning techniques, such as grid search or random search, may be used to optimize model performance and generalization capabilities.

3.5 Model Evaluation and Validation: After training, the model is evaluated using the validation set to assess its performance metrics, such as accuracy, precision, recall, and F1 score. Additionally, model uncertainty and confidence intervals may be estimated to quantify prediction uncertainty and provide insights into model reliability. Cross-validation techniques, such as k-fold cross-validation, may be employed to validate model robustness and assess its generalization capabilities across different subsets of the data.

3.6 Deployment and Integration: Once the model is trained and validated, it can be deployed for real-world applications in seed yield prediction. This may involve integrating the model into decision support systems, agricultural management platforms, or mobile applications to provide timely and actionable insights to farmers and stakeholders. Model deployment considerations include scalability, reliability, and interoperability with existing agricultural infrastructure.

3.7 Interpretability and Explainability: Model interpretability and explainability are crucial for gaining insights into the underlying factors driving seed yield predictions and building trust among end-users. Techniques like feature importance analysis, attention mechanisms, and saliency maps may be employed to visualize and interpret model predictions. Additionally, post-hoc explainability methods, such as LIME (Local Interpretable Model-agnostic Explanations) or SHAP (Shapley Additive explanations), can provide explanations for individual predictions and highlight the contribution of input features to the model output.

3.8 Ethical and Societal Implications: Finally, it is essential to consider the ethical and societal implications of deep learning-based seed yield prediction. This includes addressing concerns related to data privacy, algorithmic bias, and equitable access to technology. Stakeholder engagement and participatory approaches may be employed to ensure that the benefits of seed yield prediction technologies are equitably distributed and contribute to sustainable agricultural development.

4. RESULTS AND DISCUSSION

Our analysis reveals a growing body of research dedicated to DL-based seed yield prediction, with an increasing emphasis on model refinement, interpretability, and scalability. CNNs have emerged as popular for analyzing remote sensing imagery and extracting spatial features related to crop health, canopy cover, and vegetation indices. RNNs and long short-term memory (LSTM) networks have been employed to model temporal dependencies in weather patterns, soil moisture levels, and crop phenology, enabling dynamic yield forecasting over multiple growing seasons. Moreover, attention mechanisms and transfer learning techniques have been proposed to enhance model interpretability and generalization capabilities, facilitating knowledge transfer across different crop types and geographical regions.

Despite the progress made in DL-based seed yield prediction, several challenges remain to be addressed. Data quality and representativeness are critical factors influencing model performance, necessitating the collection of diverse and geographically distributed datasets for robust model training and validation. Furthermore, model interpretability is essential for gaining insights into the underlying factors driving seed yield variations, enabling stakeholders to make informed decisions regarding agronomic practices, irrigation scheduling, and pest management strategies. Additionally, the scalability of DL models must be considered when deploying seed yield prediction systems in real-world agricultural settings, ensuring efficient utilization of computational resources and timely delivery of actionable insights to end-users.

5. POTENTIAL CHALLENGES AND OPPORTUNITIES

Despite the promising advancements in deep learning-based seed yield prediction, several challenges remain to be addressed. One major challenge is the availability of high-quality, labeled datasets for model training and validation. The scarcity of annotated data, especially for rare crops or specific geographic regions, hinders the development of robust and generalizable models. Moreover, the interpretability of deep learning models poses a significant challenge, as complex neural network architectures often lack transparency in their decision-making process. Ensuring the reliability and interpretability of model predictions is essential for gaining the trust of end-users and stakeholders in agricultural decision-making. Additionally, the scalability of deep learning models to large-scale agricultural systems remains a pressing concern, as training and deploying complex neural networks require substantial computational resources and infrastructure. Addressing these challenges requires interdisciplinary collaboration between researchers, agronomists, data scientists, and policymakers to develop innovative solutions and foster the adoption of deep learning in precision agriculture.

6. CONCLUSION AND FUTURE DIRECTIONS

In conclusion, DL-based seed yield prediction holds immense potential for revolutionizing agricultural decision-making processes and enhancing crop productivity. By leveraging advanced DL techniques, researchers and practitioners can gain deeper insights into the complex interactions between environmental factors and seed yield outcomes, enabling proactive management strategies and risk mitigation measures. However, addressing the challenges of data scarcity, model interpretability, and scalability is essential for the widespread adoption of DL-based seed yield prediction systems. Future research directions may include the development of hybrid models combining DL with traditional statistical methods, the integration of multi-source data for comprehensive yield forecasting, and the implementation of explainable AI techniques for enhancing model transparency and trustworthiness.

CONFLICT OF INTERESTS

None

ACKNOWLEDGMENTS

None

REFERENCES

Sordello, R.; Villemey, A.; Jeusset, A.; Vargac, M.; Bertheau, Y.; Coulon, A.; Deniaud, N.; de Lachapelle, F.F.; Guinard, E.; Jactel, H. Conseils Méthodologiques pour la Réalisation D'une Revue Systématique à Travers L'expérience de COHNECS-IT. (2021).

Guoyang Zhao, Longzhe Quan, Hailong Li, Huaiqu Feng, Songwei Li, Shuhan Zhang, and Ruiqi Liu. 2021. The real-time recognition system of soybean seed full-surface defects based on deep learning. Comput. Electron. Agric. 187, 106230, May (2021).

Jun Zhang, Limin Dai, and Fang Cheng. 2021. Corn seed variety classification based on hyperspectral reflectance imaging and deep convolutional neural network. J. Food Meas. Charact. 15, 1, 484–494 (2021).

Nambiema, A.; Fouquet, J.; Guilloteau, J.; Descatha, A. La revue systématique et autres types de revue de la littérature: Qu'est-ce que c'est, quand, comment, pourquoi? Arch. Des Mal. Prof. L'Environ. 2021, 82, 539–552. (2021)

Ferhat Kurtulmuş. 2020. Identification of sunflower seeds with deep convolutional neural networks. J. Food Meas. Charact. 0123456789 (2020).

Shaolong Zhu, Jinyu Zhang, Maoni Chao, Xinjuan Xu, Puwen Song, Jinlong Zhang, and Zhongwen Huang. 2020. A rapid and highly efficient method for the identification of soybean seed varieties: Hyperspectral images combined with transfer learning. Molecules 25, 1 (2020).

Te Ma, Satoru Tsuchikawa, and Tetsuya Inagaki. 2020. Rapid and non-destructive seed viability prediction using near-infrared hyperspectral imaging coupled with a deep learning approach. Comput. Electron. Agric.177, April (2020).

Aqib Ali, Salman Qadri, Wali Khan Mashwani, Samir Brahim Belhaouari, Samreen Naeem, Sidra Rafique, Farrukh Jamal, Christophe Chesneau, and Sania Anam. 2020. Machine learning approach for the classification of corn seed using hybrid features. Int. J. Food Prop. 23, 1, 1097–1111 (2020).

R. Beulah; "A survey on different data mining techniques for crop yield prediction", Int. J. Comput. Sci. Eng., 7 (1) pp. 738-744 (2019).

Chao Ni, Dongyi Wang, Robert Vinson, Maxwell Holmes, and Yang Tao. 2019. Automatic inspection machine for maize kernels based on deep convolutional neural networks. Biosyst. Eng. 178, 131–144 (2019).

Sheng Huang, Xiaofei Fan, Lei Sun, Yanlu Shen, and Xuesong Suo. 2019. Research on Classification Method of Maize Seed Defect Based on Machine Vision. J. Sensors 2019, 1 (2019).

Pengcheng Nie, Jinnuo Zhang, Xuping Feng, Chenliang Yu, and Yong He. 2019. Classification of hybrid seeds using near-infrared hyperspectral imaging technology combined with deep learning. Sensors Actuators, B Chem. 296, May, 126630 (2019).

Siddaway, A.P.; Wood, A.M.; Hedges, L.V. How to do a systematic review: A best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses. Annu. Rev. Psychol. 2019, 70, 747–770. (2019).

X. Xu, P. Gao, X. Zhu, W. Guo, J. Ding, C. Li, X. Wu; "Design of an integrated climatic assessment indicator (ICAI) for wheat production: a case study in Jiangsu Province" Ecol. Ind., 101, pp. 943-953, (2019).

P. Filippi, E.J. Jones, N.S. Wimalathunge, P.D.S.N. Somarathna, L.E. Pozza, S.U. Ugbaje, T.F.A. Bishop; "An approach to forecast grain crop yield using multi-layered, multi-farm data sets and machine learning", Precis. Agric., 1–15 (2019).

Chlingaryan, S. Sukkarieh, B. Whelan; "Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: a review", Comput. Electron. Agric., 151, pp. 61-69, (2018).

D. Elavarasan, D.R. Vincent, V. Sharma, A.Y. Zomaya, K. Srinivasan; "Forecasting yield by integrating agrarian factors and machine learning models: a survey" Comput. Electron. Agric., 155, pp. 257-282, (2018).

B. Li, J. Lecourt, G. Bishop; "Advances in non-destructive early assessment of fruit ripeness towards defining optimal time of harvest and yield prediction—a review" Plants, 7 (1) (2018).

Balaji Veeramani, John W. Raymond, and Pritam Chanda. 2018. DeepSort: Deep convolutional networks for sorting haploid maize seeds. BMC Bioinformatics 19, Suppl 9 (2018).

Douglas F. Pereira, Priscila T.M. Saito, and Pedro H. Bugatti. 2016. An image analysis framework for effective classification of seed damages. Proc. ACM Symp. Appl. Comput. 04-08-Apri, 61–66 (2016).

Kantip Kiratiratanapruk and Wasin Sinthupinyo. 2011. Color and texture for corn seed classification by machine vision. 2011 Int. Symp. Intell. Signal Process. Commun. Syst. "The Decad. Intell. Green Signal Process. Commun. ISPACS 2011, 7–11 (2011).

Pablo M. Granitto, Pablo F. Verdes, and H. Alejandro Ceccatto. 2005. Large-scale investigation of weed seed identification by machine vision. Comput. Electron. Agric. 47, 1, 15–24 (2005).

Waoo, AA, Soni, BK. (2023) Chapter 5 - Recurrent neural network model for identifying neurological auditory disorder, Editor(s): Ajith Abraham, Sujata Dash, Subhendu Kumar Pani, Laura García-Hernández, Artificial Intelligence for Neurological Disorders, Academic Press, 2023, Pages 77-89, ISBN 9780323902779, .