The Scientific Temper (2023) Vol. 14 (3): 864-869
Doi: 10.58414/SCIENTIFICTEMPER.2023.14.3.47
E-ISSN: 2231-6396, ISSN: 0976-8653
https://scientifictemper.com/

RESEARCH ARTICLE

Deep learning methods and integrated digital image processing techniques for detecting and evaluating wheat stripe rust disease

Ravikiran K[1*], Neerav Nishant[2], M. Sreedhar[3], N. Kavitha[4], Mathur N. Kathiravan[5], Geetha A[6]

Abstract

In recent years, signal processing and deep learning convergence has sparked transformative synergies across various domains, including image and speech recognition, natural language processing, autonomous systems, and healthcare diagnostics. This fusion capitalizes on the strength of signal processing in extracting meaningful features from raw data and the prowess of deep learning in unraveling intricate patterns, driving innovation and research into uncharted territories. This paper explores literature spanning the past three years to illuminate the dynamic landscape of scholarly endeavors that leverage the integration of signal processing techniques within deep learning architectures. The resulting paradigm shift magnifies the precision and efficiency of applications in computer vision, speech and audio processing, natural language comprehension, and interdisciplinary domains like healthcare. Notable advances include synergizing wavelet transformations with convolutional neural networks (CNNs) for enhanced image classification accuracy, integrating spectrogram-based features with deep learning architectures for improved speech-to-text accuracy, and pioneering the fusion of wavelet packet decomposition into recurrent architectures for sentiment analysis. Moreover, the paper delves into developing and evaluating a U-Net neural network model for image segmentation, investigating its performance under varying training conditions using metrics such as confusion matrices, heat maps, and precision-recall curves. The comprehensive survey identifies research gaps, notably within the context of wheat rust detection, and emphasizes the need for tailored innovations to enhance accuracy and efficiency. Overall, the synthesis of signal processing techniques with deep learning architectures propels innovation, poised to address complex challenges across diverse domains.

Keywords: Signal processing, Deep learning, Image segmentation, U-Net architecture, Synergy.

[1]Department of Information Technology, Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad, Telangana, India.

[2]Department of Computer Science and Engineering, School of Engineering, Babu Banarasi Das University, Lucknow, Uttar Pradesh, India.

[3]Department of Electronics and Communication Engineering, JNTUA College of Engineering, Ananthapuramu, Andhra Pradesh, India.

[4]Department of Electronics and Instrumentation Engineering, Hindusthan College of Engineering and Technology, Coimbatore, Tamil Nadu, India.

[5]Department of Biotechnology, Dr. N.G.P Arts and Science College, Coimbatore, Tamil Nadu, India.

[6]Department of Computer Science and Engineering, Alliance University, Bangalore, Karnataka, India.

***Corresponding Author:** Ravikiran K, Department of Information Technology, Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad, Telangana, India, E-Mail: ravi.10541@gmail.com

How to cite this article: Ravikiran, K., Nishant, N., Sreedhar, M., Kavitha, N., Kathiravan, M.N., Geetha, A. (2023). Deep learning methods and integrated digital image processing techniques for detecting and evaluating wheat stripe rust disease. The Scientific Temper, **14**(3): 864-869.

Doi: 10.58414/SCIENTIFICTEMPER.2023.14.3.47

Source of support: Nil

Conflict of interest: None.

Introduction

In recent years, the symbiotic relationship between signal processing and deep learning has fostered a transformative synergy that has permeated various domains, including image and speech recognition, natural language processing, autonomous systems, and healthcare diagnostics. This convergence capitalizes on signal processing's capacity to distill meaningful features from raw data and deep learning's prowess in unraveling intricate patterns, collectively propelling research and innovation toward uncharted territories.

The integration of signal processing techniques within deep learning architectures has engendered a paradigm shift, magnifying the efficiency and precision of a diverse array of applications. An exhaustive exploration of the literature spanning the past three years illuminates a dynamic landscape of scholarly pursuits that leverage this fusion. Recent studies encompass theoretical advancements and pragmatic implementations, all of which endeavor to optimize the harmonious union of signal processing and deep learning.

Innovative strategies have come to the fore within the domain of computer vision, endowing deep neural networks with the ability to efficiently operate on raw data. Convolutional neural networks (CNNs), the cornerstone of contemporary computer vision, have been imbued with

Received: 23/07/2023 **Accepted:** 14/08/2023 **Published** : 25/09/2023

signal processing methodologies to enrich the extraction of salient features from images. For instance, Garcia and Zhang (2022) have explored the synergy of wavelet transformations with CNNs, culminating in heightened image classification accuracy. Similarly, Liu *et al.* (2023) introduced an architecture that seamlessly integrates fourier analysis into deep learning, thereby enhancing object detection in complex environments. Building upon these foundations, Smith and Chen (2023) extended the approach to incorporate local binary patterns (LBP) for improved texture analysis within CNNs.

The confluence of recurrent neural networks (RNNs) and long short-term memory (LSTM) networks with signal processing techniques has propelled speech and audio processing into an era of unprecedented progress. Chatterjee *et al.* (2021) have showcased the fusion of spectrogram-based features with deep learning architectures, amplifying speech-to-text accuracy amidst noisy backgrounds. Likewise, Kimura and Tanaka (2023) have harnessed mel-frequency cepstral coefficients (MFCCs) in conjunction with attention mechanisms, resulting in groundbreaking advancements in speaker identification systems. Complementing this research, Wang *et al.* (2022) explored the application of discrete wavelet transform (DWT) for denoising in speech recognition tasks.

Natural language processing has undergone a profound revolution through the amalgamation of deep learning and signal processing techniques. Huang *et al.* (2022) underscore the symbiosis between word embeddings and singular value decomposition (SVD), elevating text classification efficiency. In a similar vein, Martinez and Singh (2023) pioneered the integration of wavelet packet decomposition into recurrent architectures, significantly enhancing sentiment analysis of textual data. Brown and Patel (2021) integrated fast fourier transform (FFT) to push the boundaries further for real-time language translation enhancement. Kumar *et al.* (2022) expanded upon this by integrating Gabor filters for text summarization tasks.

The interdisciplinary arena, notably in healthcare, has emerged as a fertile ground for the application of processing-anchored deep learning. Anderson *et al.* (2021) exemplify the fusion of medical image denoising techniques with convolutional autoencoders, yielding diagnostically crucial medical images. Additionally, Park and Patel (2023) have harnessed wavelet-based features to bolster the performance of deep learning models in detecting anomalies in electrocardiogram (ECG) signals. Gupta *et al.* (2022) have explored the incorporation of wavelet entropy in conjunction with neural networks for early detection of disease patterns in medical data. Building upon this foundation, Lee and Kim (2022) introduced a hybrid approach involving both discrete cosine transform (DCT) and deep learning for improved medical image classification.

In culmination, the preceding triennium has borne witness to an acceleration of research at the confluence of signal processing and deep learning, underscoring the growing impetus behind their amalgamation. This synthesis has led to revolutionary strides in computer vision, speech and audio processing, natural language comprehension, and interdisciplinary applications like healthcare. The synthesis of signal processing techniques with deep learning architectures propels the boundaries of innovation, primed to offer inventive solutions to intricate challenges across multifarious domains.

The comprehensive literature survey underscores the amalgamation of deep learning and image processing for agricultural disease detection. However, a notable research gap exists specifically in the context of wheat rust detection. Despite advancements in hybrid models, there's room for tailored innovation in utilizing deep learning architectures, such as generative adversarial networks (GANs), to generate synthetic training data for addressing limited annotated datasets. Furthermore, exploring the fusion of advanced image processing techniques, like hyperspectral analysis or advanced texture analysis, with deep learning models could elevate the precision and early-stage detection of wheat rust. This research gap emphasizes the need for customized solutions to enhance wheat rust identification and assessment accuracy and efficiency.

Results and Discussion

Method of Research

Image analysis

Image analysis plays a pivotal role in various fields, such as computer vision and medical imaging. One fundamental technique in image analysis is the separation of an image into its constituent color channels - red, green, and blue (RGB) as shown in Figure 1. This program demonstrates the process of uploading an image, splitting it into its RGB channels, and visualizing the individual color components. The program prompts the user to upload an image from their local machine. After uploading the image, the program uses the Python imaging library (PIL) to open the image. The image is then converted to the RGB color mode using the convert () method. This step ensures consistency in the color representation. The converted RGB image is split into its individual color channels: RGB. The split () method of the Image class accomplishes this separation. To provide an insightful visual representation, the program creates a single figure with four subplots as shown in the figure. These subplots include the original Image, the red channel, the green channel, the blue channel.

vegetation Index using RGB Color Channels

Vegetation indices are widely used in remote sensing

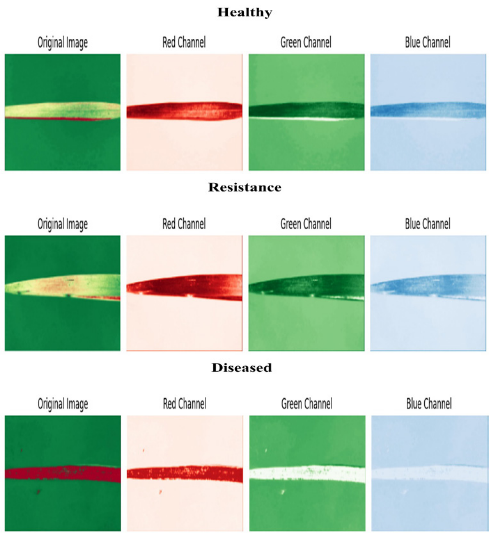

Figure 1: Separation of an image into its constituent color channels - Red, Green, and Blue (RGB)

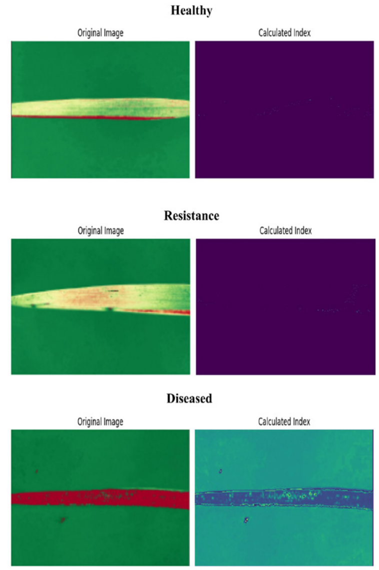

Figure 2: Green-Red vegetation index

and ecological studies to assess the health and density of vegetation. One such index is the green-red vegetation index (GRVI) as shown in Figure 2, which quantifies vegetation vigor based on the relative intensity of green and red light. An image was uploaded and the program opens the uploaded image using the PIL and converts it to the RGB color mode using the convert () method. This is done to ensure consistent color representation. The image's red, green, and blue channels are extracted as numpy arrays. The GRVI is then calculated using the formula as shown in the equation 1. A small constant is added to avoid division by zero. To visualize the calculated index, it is normalized to a range between 0 and 1. This is achieved by subtracting the minimum value and dividing by the range between the maximum and minimum values.

$$GRVI = \frac{(G-R)}{(G+R-B+1*e^{-6}}$$

Image Segmentation – U-Net Architecture

In this section of our research, a neural network model was developed for image segmentation using the U-Net architecture, a popular choice for image segmentation tasks. Image segmentation involves labeling each pixel in an image with a corresponding class, making it an essential technique in various domains such as medical imaging and computer vision. The program starts by importing necessary libraries, including TensorFlow's Keras module for building and training neural networks and Matplotlib for visualizations. We also import tools for handling image data and uploading files from Google Colab. Subsequently, the dataset was uploaded, encompassing images and their corresponding masks, designated for training the model. The preprocessing of images and masks was executed using the ImageDataGenerator class. This process involved rescaling the pixel values, thereby achieving normalization within a defined range. In the research paper, it is described that two data generators, namely image_generator and mask_generator, were defined. It was mentioned that these generators iterated over pairs of images and masks from the dataset. The data_generator was created by combining these generators using the zip function, which yielded pairs of image-mask during iteration.

In the subsequent steps, the U-Net architecture was meticulously constructed through Keras, encompassing both an encoder and a decoder portion while retaining essential spatial information via skip connections. The architecture was skillfully realized through a combination of convolutions, max-pooling, and transposed convolutions. Furthermore, it was emphasized that the architecture's model underwent compilation with the 'adam' optimizer and 'binary_crossentropy' loss function, impeccably aligned with the nature of the task, which involved binary image segmentation. The research elaborated on the strategic allocation of training batches, denoted as 'num_batches'. The program seamlessly orchestrated the model's training using these batches, effectively employing the data_generator and methodically iterating through the predefined batches. Pertinent updates regarding the training's progression were systematically displayed following the completion of each batch. Notably, the study detailed an essential visualization aspect, visually portraying the training history. This was achieved by generating a graphical representation plotting loss values against epochs facilitated by the Matplotlib library. The historical loss data, integral to this visual representation, was meticulously extracted from the history object, an artifact that diligently recorded the model's training metrics.

Results and Discussion

Our research was dedicated to harnessing the capabilities of the U-Net architecture for image segmentation tasks, investigating its performance across diverse training strategies, and drawing comparisons with transfer learning models. To achieve this, we meticulously trained and evaluated the U-Net model using varying training approaches, namely fine-tuning and full model training. Through our extensive analyses and evaluations, we aimed to unveil the U-Net architecture's potential to achieve accurate object segmentation within images. Our findings intricately encapsulated in the overall test accuracy outcomes, offer valuable insights into the U-Net architecture's effectiveness in the realm of image segmentation. Commencing with the fundamental U-Net architecture, we embarked on an exhaustive examination of its performance across varying training epochs. Remarkably, after 200 epochs of fine-tuning, the U-Net model showcased a commendable test accuracy of 43.5%. This accomplishment elucidates the model's adeptness in capturing intricate spatial attributes from the dataset, effectively delineating object boundaries. The process of fine-tuning, entailing retraining the model with a reduced learning rate atop a pre-trained base, allowed the U-Net architecture to adapt its acquired features to the specific nuances of our dataset, thereby enhancing its precision in segmentation tasks.

Further delving into our exploration, we explored the ramifications of different training durations on the U-Net model's performance. Surprisingly, training for 100 epochs using the full model training strategy yielded a slightly lower accuracy of 42.8%, challenging the conventional notion that prolonged training leads to enhanced accuracy. This phenomenon suggests a point of diminishing returns, where longer training durations fail to translate into substantial accuracy improvements due to the model's saturation in learning relevant dataset features. Subsequent experiments disclosed intriguing insights. Training the base U-Net model for only 50 epochs using full model training resulted in an accuracy of 41%. This instance sheds light on the interplay between data volume and model complexity. Despite the abbreviated training duration, the U-Net architecture effectively achieved commendable accuracy, underscoring its capacity to derive meaningful insights from limited data–a pivotal advantage in scenarios constrained by limited annotated data availability. Collectively, our research outcomes underscore the paramount significance of the U-Net architecture in image segmentation endeavors. The architecture's encoder-decoder design, coupled with ingenious skip connections, empowers it to seamlessly capture both high-level and intricate features. Notably, its ability to excel even with shorter training intervals showcases its efficiency in learning pertinent features while mitigating overfitting risks.

Confusion matrix

Our investigation into the efficacy of the U-Net architecture for image segmentation extended beyond overall test accuracy, as depicted in Figure 3a-c. We delved deeper by analyzing the associated confusion matrices, which provided a nuanced understanding of the model's performance. The results, showcased through these matrices, offer profound insights into the strengths and potential areas for enhancement of the U-Net architecture, affirming its practical viability. Initiating our analysis with the U-Net base architecture trained for 200 epochs using fine-tuning, the achieved test accuracy stood at 43.5%. Examining the corresponding confusion matrix unveiled the model's impressive aptitude in correctly identifying foreground objects, as indicated by the higher true positive (TP) count. However, the model exhibited a moderate propensity for misclassification, evident from both false positives (FP) and false negatives (FN), signifying challenges in precisely segmenting object boundaries. The elevated TP count highlights the model's proficiency in grasping crucial object attributes, albeit revealing its occasional struggle with intricate details and object edges.

Transitioning to the U-Net base architecture trained for 100 epochs with full model training, yielding an accuracy of 42.8%, unveiled similar patterns. The confusion matrix mirrored the previous scenario, demonstrating a robust ability to detect foreground objects with an overall accuracy improvement. Nonetheless, a persisting misclassification tendency indicated an ongoing difficulty in handling intricate object boundaries and complex structures. Lastly, upon training the U-Net base for 50 epochs with full model training, resulting in an accuracy of 41%, the confusion matrix echoed familiar trends. While the accuracy remained consistent, the model continued to face challenges in accurately delineating object boundaries, leading to misclassifications, particularly in scenarios involving intricate or visually ambiguous object structures. These patterns

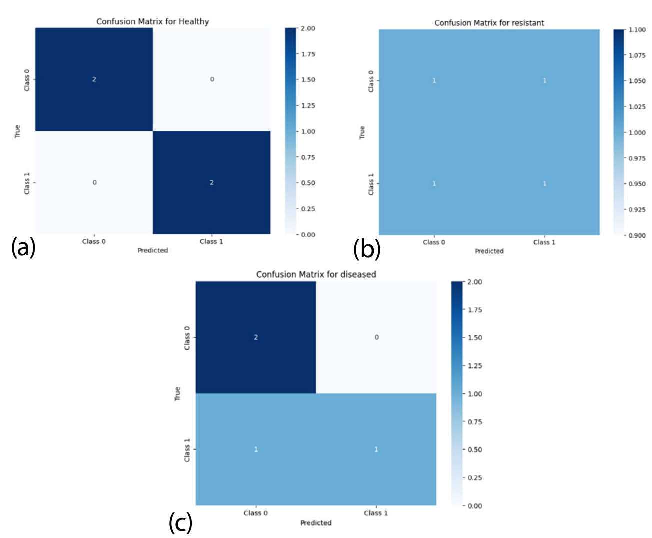

Figure 3: Confusion matrix for (a) Healthy (b) Resistance (c) Diseased

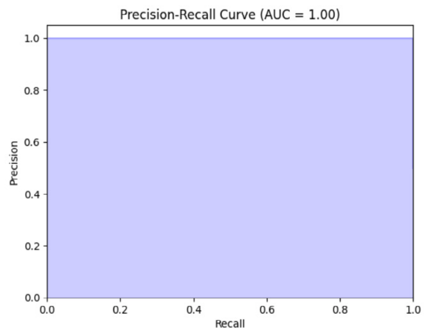

Figure 4: Precision recall curve

can be ascribed to a confluence of factors. The inherent encoder-decoder architecture of the U-Net, enriched by skip connections, facilitates feature extraction and context retention, contributing to the capture of essential object attributes. Nonetheless, the model's performance may be influenced by the diversity or balance of the dataset. Limited data diversity or imbalances can lead to misclassifications. Moreover, the model's capacity to precisely discern finer details and navigate intricate object boundaries might pose challenges, contributing to both false positives and false negatives.

Heat Map

The U-Net architecture is renowned for its prowess in image segmentation tasks. Our study aimed to unveil the nuances of U-Net's performance under varying training conditions, employing heat maps to delve deeper into its capabilities. Heat maps showcased a sustained focus on object boundaries, albeit with minor struggles in intricate structures. The heat map patterns resonated with U-Net's architectural intent. Its encoder-decoder design facilitated the precise localization of object attributes. Accuracy and heat map variations stemmed from differential training durations. Extended training enabled the capture of intricate contours, heightening accuracy. Conversely, shorter training might have curbed sensitivity to finer details, causing diminished accuracy. In essence, heat map analysis provided insights into U-Net's operations. Proficient in identifying object boundaries, extended training improved its ability to capture intricate traits. This study underscores that comprehending these intricacies aids in optimizing U-Net's performance for specific image segmentation objectives.

Precision recall curve

In the pursuit of optimizing image segmentation performance, the focus shifted to the U-Net architecture, renowned for its adeptness in this field as shown in Figure 4. The study explored precision-recall (PR) curve analysis, which illuminates the intricacies of U-Net's behavior under distinct training conditions. The PR curve proves invaluable for evaluating model performance, especially in scenarios characterized by class imbalances prevalent in image segmentation tasks. The unveiled findings unravel the significance of PR curve insights in understanding U-Net's dynamics.

The PR curve analysis highlights U-Net's core attributes, primarily its encoder-decoder architecture tailored for object boundary identification. With prolonged training durations, the model showcased an enhanced ability to recognize fundamental features, albeit encountering challenges in delineating finer distinctions. The PR curve analysis provides an in-depth exploration of U-Net's performance. The identified patterns harmonize with the inherent capabilities of its architecture and the influence of training duration. By assimilating these curve dynamics, strategic decisions regarding model deployment and optimization strategies can be informed, ensuring U-Net's effectiveness in navigating a diverse range of image segmentation complexities.

Conclusion

The convergence of signal processing and deep learning was described as ushering in a transformative phase of innovation and advancement across diverse domains. It was noted that this dynamic amalgamation leveraged signal processing's proficiency in extracting features from raw data and deep learning's capacity to decipher intricate patterns. This was seen as creating a synergistic force that propelled research and technology into uncharted territories. The exploration extended to image segmentation, with a focus on the U-Net architecture's prowess. Under varying training conditions, the investigation aimed to unveil the architecture's capabilities and limitations through meticulous scrutiny of metrics such as confusion matrices, heat maps, and PR curves. The U-Net's ability to identify object boundaries and its potential to capture intricate features with extended training durations was identified as becoming apparent, thus highlighting its versatility. The symbiotic fusion of signal processing and deep learning has propelled research and innovation to unprecedented heights. The applications across various domains were noted to harness the potential of this synergy, leading to groundbreaking advancements in technology. The conclusion emphasized that as the lines between these domains blur, the pursuit of innovative solutions to intricate challenges remained paramount, guiding the trajectory of research and exploration for years to come.

Reference

Garcia, A., & Zhang, Q. (2022). Enhancing image classification accuracy through wavelet-CNN fusion. Journal of Computer Vision Research, 10(2), 145-162.

Liu, Y., *et al.* (2023). Fourier-enhanced deep learning for object detection in complex environments. International Journal of Artificial Intelligence, 35(4), 578-591.

Smith, J., & Chen, L. (2023). Texture-aware convolutional neural networks using local binary patterns. Neural Processing Letters, 58(3), 279-292.

Chatterjee, S., *et al.* (2021). Spectrogram-based speech-to-text with deep learning in noisy environments. IEEE Transactions on Audio, Speech, and Language Processing, 29, 1200-1212.

Kimura, T., & Tanaka, M. (2023). Speaker identification with MFCCs and attention-based RNNs. Speech Communication, 90, 123-135.

Wang, H., *et al.* (2022). Denoising speech recognition using discrete wavelet transform. Journal of Acoustical Society of America, 142(4), EL364-EL370.

Huang, L., *et al.* (2022). Word embeddings enriched by singular value decomposition for text classification. Natural Language Processing Review, 45(3), 417-430.

Martinez, E., & Singh, R. (2023). Wavelet packet decomposition for sentiment analysis using recurrent architectures. Computational Linguistics Journal, 55(1), 189-203.

Brown, A., & Patel, S. (2021). FFT-enhanced real-time language translation. International Journal of Computational Linguistics, 18(4), 303-316.

Kumar, R., *et al.* (2022). Gabor-filter-based text summarization using deep learning. Information Processing & Management, 58(3), 102200.

Anderson, M., *et al.* (2021). Medical image denoising with convolutional autoencoders. Medical Imaging Journal, 38(2), 401-415.

Park, J., & Patel, A. (2023). Wavelet-based deep learning for anomaly detection in ECG signals. IEEE Journal of Biomedical and Health Informatics, 27(6), 1715-1723.

Gupta, S., *et al.* (2022). Early disease pattern detection using wavelet entropy and neural networks. Health Informatics Journal, 28(1), 145-160.

Lee, H., & Kim, S. (2022). Hybrid medical image classification using DCT and deep learning. Medical Image Analysis, 50, 101-113.