Journal of Soft Computing and Decision Analytics

Journal homepage: www.jscda-journal.org
ISSN: 3009-3481

# Data-Efficient Vision Transformer Models for Robust Classification of Sugarcane

İshak Paçal[1], İsmail Kunduracıoğlu[1,*]

[1] Department of Computer Engineering, Faculty of Engineering, Igdir University, Iğdır, Turkey

**ARTICLE INFO**

***Article history:***
Received 6 April 2024
Received in revised form 30 May 2024
Accepted 16 June 2024
Available online 23 June 2024

***Keywords:***
Artificial intelligence; Deep-Learning; ViT; Plant diseases; Image processing.

**ABSTRACT**

Sugar cane is an important agricultural product that provides 75% of the world's sugar production. As with all plant species, any disease affecting sugarcane can significantly impact yields and planning. Diagnosing diseases in sugarcane leaves using traditional methods is slow, inefficient and often lacking in accuracy. This study presents a deep learning-based approach for accurate diagnosis of diseases in sugarcane leaves. Specifically, training and evaluation were conducted on the publicly available Sugarcane Leaf Dataset using leading ViT (Vision Transformer) architectures such as DeiT3-Small and DeiT-Tiny. This dataset includes 11 different disease classes and a total of 6748 images. Additionally, these models were compared with popular CNN models. The findings of the study show that there is no direct relationship between model complexity, depth and accuracy for the 11-class sugarcane dataset. Among the 12 models tested, the DeiT3-Small model showed the highest performance with 93.79% accuracy, 91.27% precision, and 90.96% F1-score. These results highlight that rapid, accurate and automatic disease diagnosis systems developed using deep learning techniques can significantly improve sugarcane disease management and contribute to increased yields.

## 1. Introduction

Agriculture serves as a crucial income source for rural populations in developing nations. Nevertheless, agricultural productivity must be enhanced to satisfy the food demands of a growing population [1]. The agricultural sector, however, faces numerous challenges, including plant diseases, pests, and varying weather conditions. These changing weather patterns hasten the spread of diseases, raising concerns about food safety. Plant diseases significantly threaten agricultural output. Factors like temperature, precipitation, wind speed, and extreme weather events such as droughts, heavy rainfall, hail, and hurricanes influence agriculture's sensitivity to climate. These events can reduce yields and damage soil [2]. Early detection can mitigate the damage caused by plant diseases, but manual methods make early diagnosis difficult. Diseases often begin on lower leaves and spread throughout the crop, making visual monitoring, rapid detection, and prevention critical. Artificial

* *Corresponding author.*
*E-mail address: ismail.kunduracioglu@igdir.edu.tr*

https://doi.org/10.31181/jscda21202446

intelligence (AI) and classification methods can automate this process. AI, particularly machine learning and convolutional neural networks (CNNs), can enhance precision agriculture by efficiently detecting and classifying pests with minimal labor.

Recently, AI techniques have been utilized to create expert systems for problem-solving and decision-making. Image processing techniques analyze pixel regions to identify patterns and develop algorithms for detecting behavioral trends. As a subset of AI, deep learning is a powerful feature extraction and classification method with significant potential in agriculture. Traditional diagnostic methods based on visual inspections are labor-intensive, costly, and relatively less sensitive, leading to substantial yield losses, especially for rural farmers. The interest in non-invasive methods has grown, providing automatic, quick, and accurate solutions [3]. Among these solutions, image processing techniques are prominent, yielding promising disease detection and management results using advanced cameras with sensitive sensors. Technological advancements have enhanced the use of technology across various fields [4]. This integration of technology promotes sustainability and reduces environmental impact by minimizing resource waste. These trends position autonomous agriculture as a key player in meeting global food production needs and addressing challenges like climate change and resource scarcity [5].

Deep learning algorithms are increasingly applied to diagnose and identify diseases in sectors such as healthcare and agriculture. The adoption of these technologies in agriculture marks a significant step towards boosting yields, ensuring food security, and promoting sustainability. Advances in deep learning and plant disease diagnosis are critical for the sustainability of agriculture. Using advanced technologies and methods can enhance productivity and foster economic growth by enabling early disease detection [6,7]. The significant progress in agricultural practices through deep learning models underscores the growing importance of research in this area.

Sugarcane, a member of the Poaceae family, is used to make molasses, white sugar, and jaggery (palm sugar), among other byproducts. It has a high sucrose content. Sugar cane is used to produce 75% of the sugar produced worldwide. Sugar cane juice's alkaline composition lowers blood pressure, promotes kidney and liver function, and lowers the risk of breast and prostate cancer. On the other hand, disease outbreaks can severely lower sugarcane production [8]. Effective planting requires regular monitoring of plant health. Diseased leaves, stems, fruits, and other afflicted parts can be identified using deep learning and image processing approaches. To discern between healthy and unhealthy plants, several deep learning algorithms are employed.

In order to determine whether sugarcane illnesses were present, Bashir & Sharma [9] described a discrete transform technique that used a particular wavelength. They also used the tree of decision to categorize pictures. When compared to conventional ANN approaches, the Elementary Learning Machine (ELM) performed better in predicting the development of sugarcane in different regions [10].

Hamuda *et al.* [11] developed an algorithm to automatically detect products, specifically broccoli, in video streams under various weather conditions and natural lighting. Their algorithm achieved a remarkable accuracy of 99.04% and a precision of 98.91% when compared to manually labeled ground truth data. Akbarzadeh *et al.* [12] proposed a plant classification method using support vector machines, which demonstrated a high accuracy rate of 97% in their experimental findings. Trong *et al.* [13] introduced a novel approach for weed classification using multimodal deep learning models, such as Inception-ResNet, MobileNet, NASNet, ResNet, and VGG. This method achieved an impressive accuracy of over 98.7%, allowing for real-time weed classification. Bhosle and Ahirwadkar [14] conducted experiments using structured data from hyperspectral images to identify cotton, sugarcane, and mulberry crops. They found that a deep learning CNN achieved an accuracy of

99.33%, while a deep FFNN achieved 96.6% accuracy. Significant progress has been achieved in the identification of plant diseases and deep learning in recent years. Natural language processing has gained interest because of the success of CNN and ViT models [15]. Using image processing techniques to extract plant traits and identify the presence of illness, many researchers have developed methods for classifying sugarcane diseases [8,16]. Texture in plant leaves and illnesses has been analyzed using the structure of color transition. After segmenting and applying Gabor's filter on the leaves, Arivazhagan *et al.* [17] trained a network of artificial neurons (ANN) to distinguish between classes.

Hashemi-Beni *et al.* [18] examined the use of aerial imagery for classifying weeds and crops using deep learning architectures such as U-Net, SegNet, FCNs, and DepLabV3+. DepLabV3+ achieved the highest accuracy at 84.3%. Veziroglu *et al.* [19] evaluated various models, including VGG, ResNet, DenseNet, EfficientNet, Inception, and Xception, on the Paddy Doctor dataset. The EfficientNetv2_Small model outperformed the others with a test accuracy of 98.01% and an F1-score of 97.99%. In another study, Kılıçarslan and Pacal [20] used DenseNet, ResNet50, and MobileNet architectures to detect diseases in tomato leaves. DenseNet provided the best performance with an accuracy of 99%. Pacal [21] tested 28 CNN models and 36 ViT models on a new dataset combining PlantVillage, PlantDoc, and CD&S datasets, achieving an accuracy rate of 99.24%.

These methods have also been applied to visualize lesions on products like guava [22], tea [23], and apple [24]. Additionally, Goluguri *et al.* [25] developed a neural network to predict rice blast disease using meteorological parameters like wind speed, temperature, rainfall, and relative humidity. Militante and Gerardo [26] trained models using 14,725 images of healthy and diseased sugarcane leaves, achieving a maximum training accuracy of 95.40% with VGGNet, followed by LeNet at 93.65% and StridedNet at 90.10%. Chen *et al.* [27] studied the impact of data augmentation and varying lighting conditions on detecting sugarcane stem nodes, identifying YOLO v4 as the top performer with an average precision of 95.17%, compared to Faster R-CNN (78.87%), SSD300 (88.98%), RetinaNet (90.88%), and YOLO v3 (92.69%). Wang *et al.* [28] improved an algorithm, resulting in a mean average precision (MAP) of 99.11% and a detection accuracy of 97.07%, surpassing the Faster-RCNN and YOLOv4 algorithms.

The Vision Transformer (ViT) structure, based on how individuals classify images of specific elements, was recently introduced to help segmentation applications. When a person looks at a photograph, they focus in a certain part of the image to discover the object of interest, according to Borhani *et al.* [29]. This methodology is applied by the ViT structure for picture categorization. Vision transformer (ViT) with hard patch embedding as input is suggested by Dosovitskiy *et al.* [30]. In order to encode the spatial location of each patch within the image, ViT also uses positional embeddings.

## 2. Methodology

### *2.1 Deep-Learning*

Machine learning methods have achieved significant successes for the advancement and modernization of society. These methods are widely used in various applications ranging from finding web search queries to filtering social media content and providing recommendations on e-commerce sites. Furthermore, with advanced technology, machine learning has become an integral part of our daily lives through smart devices. Machine learning also includes numerous applications such as object recognition in images, converting speech to written text, and matching specific news or social media posts to users' interests [31].

The primary goal of image processing is to create autonomous systems that can perform tasks beyond the capabilities of human visual systems [32]. In the early stages of image processing research, algorithms were developed to detect edges, curves, corners, and other basic shapes from input images. Prior to deep learning, image processing relied on gray-level segmentation, which was not robust enough to represent complex classes. Modern computer vision algorithms, based on artificial neural networks, have dramatically improved performance and accuracy compared to traditional image processing approaches [33].

These applications encompass deep learning techniques, a branch of machine learning methods that utilize iterative processes to run and analyze data until they can discern differences and identify or describe features in images. Deep learning enables data to be learned using computation models and algorithms. It can detect complex structures in large datasets and learn using the backpropagation algorithm. The growing interest in deep learning is due to its ability to process large amounts of heterogeneous data and integrate solutions into various hardware. Deep learning enables automatic feature extraction and is effective in many image-processing tasks, including image classification, object detection, and semantic segmentation [31].

Deep learning models play a crucial role in tasks such as image, video, speech, and audio processing, while recurrent networks enable the exploration of sequential data such as text and speech. These methods have elevated technologies developed in various fields such as speech recognition, image recognition, and object detection to the highest levels [31]. In agriculture, these methods form the backbone of modeling and automating agricultural activities such as disease identification, weed detection, and yield estimation [34].

### *2.2 Vision Transformer (ViT)*

Recently, the computer vision community has begun to apply this approach to the field of image processing, considering the success and flexibility of transformer models in the field of natural language processing (NLP). Transformer models have become the de facto standard in the field of text processing, and this success has also attracted great interest in processing visual data [16,30].

Vision Transformer (ViT) processes visual data based on transformer blocks instead of traditional Convolutional Neural Networks (CNN). ViT divides the input image into patches and provides a sequence of linear embeds of these patches as input to a Transformer. This approach is designed to process image patches as a string of words, like tokens in NLP. Its general architecture consists of three main components: patch embedding, feature extraction with stacked transformer encoders, and the classification head [30,35]. ViT embeds the input image (in the form of height, width, channels) into a feature vector using a set of transformations. This process splits the input image into a group of image patches and then feeds these image patch groups into the transformer encoder network by embedding them into encoded vectors. The Transformer encoder consists of a two-layer MLP with multi-head attention (MHA), layer normalization, and residual connections to learn features from embedded patches. The last MLP block is used as the output of the transformer and produces classification outputs with a softmax function in the case of image classification (Figure 1).

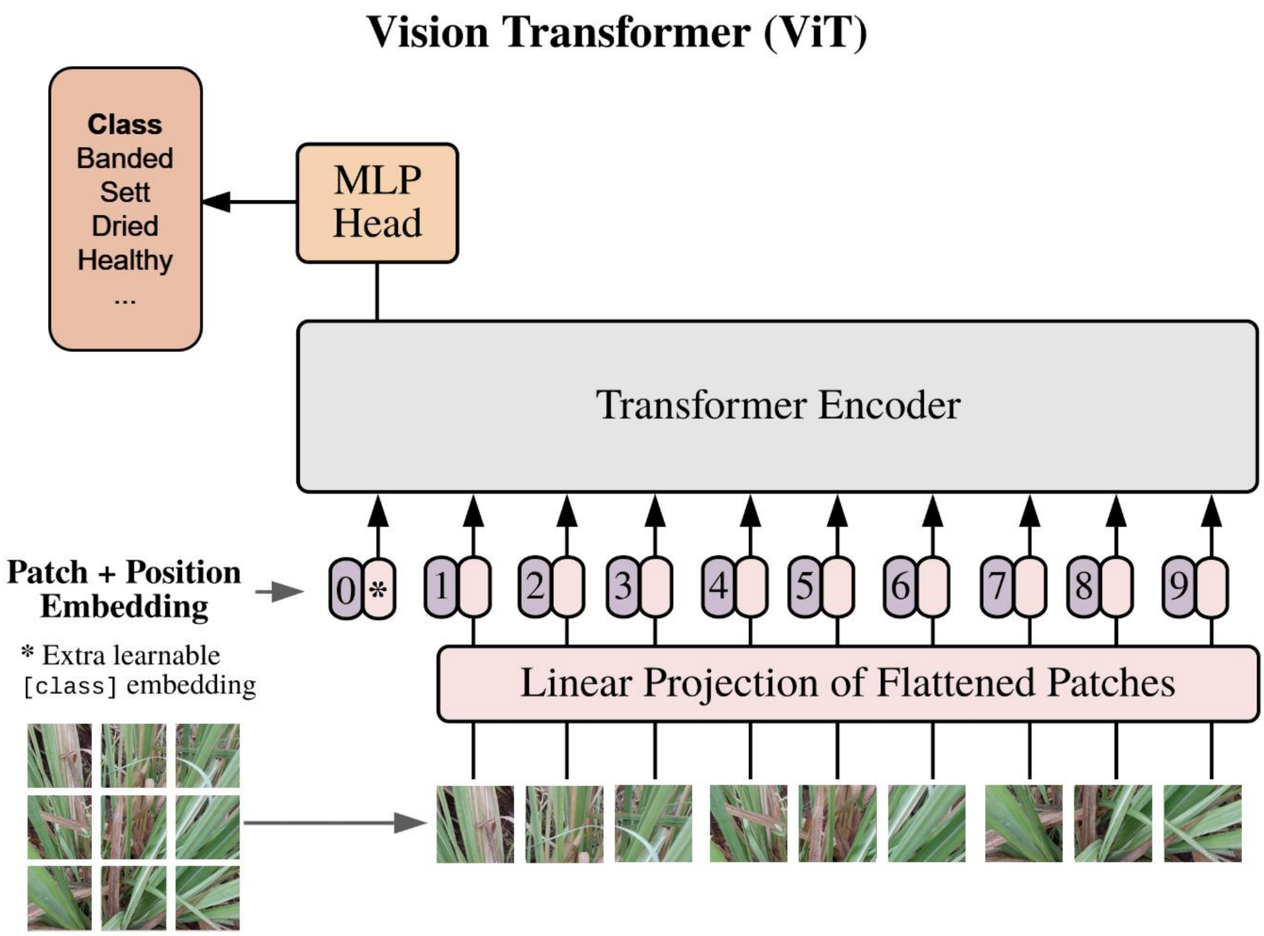


**Fig. 1.** Operating principle of ViT

Compared to traditional CNN architectures, Vision Transformers (ViT) offer differences such as image size processing, computational complexity, transfer learning and performance. Especially when trained using pre-training and transfer learning with large datasets, ViT models can achieve high success in various computer vision tasks. Additionally, the absence of convolution operations makes them more efficient in terms of computational cost and allows processing of large-sized images. The salient features of ViT are that it is an object detection model inspired by transformer models in NLP, published by Dosovitskiy *et al.* [30]. ViT is the first model to apply transformers directly to images without traditionally combining CNN and transformers. This model divides the image into patches and processes them by providing a sequence of linear embeddings of these patches as input to a Transformer. ViT can be used as a building block in various computer vision tasks, such as image classification. Besides Vision Transformers, different architectures such as patch-based, hybrid, token-based, scale-specific and mobile Vision Transformers have also been developed to increase performance and efficiency. For example, the Data-Efficient Image Transformer (DeiT) can perform well with less training data, while the SWIN architecture is designed to provide high performance with large-scale and complex datasets.

DeiT is a type of Image Transformer known as Data-Efficient Image Transformer and is specifically designed for image processing tasks. Compared to other Image Converter architectures, DeiT can perform well with less training data. DeiT uses techniques such as boosting, interpolation, and distillation to improve performance and efficiency.

SWIN is a new transducer architecture that works effectively on large-scale image data. Unlike traditional converter architectures, SWIN consists of a set of hierarchical blocks that combine the advantages of scalability and parallelism. These blocks can effectively handle small and large-scale objects and increase the generalization ability of the model. Using the attention mechanism, SWIN can model relationships at different scales and obtain a more comprehensive understanding by combining various features of images. As a result, SWIN can provide high performance and accuracy when working with large and complex data sets, making it an effective option for a variety of applications.

Today, Vision Transformers are used in many visual applications such as image classification, image-to-text, text-to-image generation, image segmentation, object detection, and provide significant progress in the field.

### *2.3 Dataset*

Datasets play a crucial role in both machine learning and deep learning, acting as essential resources that provide rich visual information. These datasets are instrumental for researchers, developers, and professionals to effectively train and validate their models, algorithms, and theories. Image datasets focusing on specific agricultural plants are particularly valuable, offering researchers and farmers invaluable tools to identify, classify, and study various diseases affecting their crops. By analyzing these images, experts can develop more accurate disease detection algorithms and early warning systems, which help speed up disease management and prevent extensive crop damage and yield loss.

**Table 1**
Categories and number of images of sugarcane dataset

| Categories | Train(70%) | Validation(15%) | Test(15%) | The Number of Images |
|---|---|---|---|---|
| Banded Chlorosis | 330 | 71 | 70 | 471 |
| Brown Rust | 220 | 47 | 47 | 314 |
| Brown Spot | 1205 | 258 | 259 | 1722 |
| Grassy Shoot | 242 | 52 | 52 | 346 |
| Pokkah boeng | 208 | 45 | 44 | 297 |
| Sett Rot | 456 | 98 | 98 | 652 |
| Smut | 221 | 47 | 48 | 316 |
| Viral Disease | 464 | 99 | 100 | 663 |
| Yellow Leaf | 836 | 179 | 179 | 1194 |
| Dried Leaves | 240 | 51 | 52 | 343 |
| Healthy Leaves | 301 | 64 | 65 | 430 |
| Total Number of Images | 4723 | 1011 | 1014 | 6748 |

Table 1 shows the number of images and split states (train, validation, test) of the sugarcane diseases, dry leaf, and healthy leaves in the sugarcane leaf dataset. This dataset includes Banded Chlorosis, Brown Rust, Brown Spot, Grassy Shoot, Pokkah boeng, Sett Rot, Smut, Viral Disease, Yellow Leaf diseases, and Dried Leaves, Healthy Leaves plant leaf images.

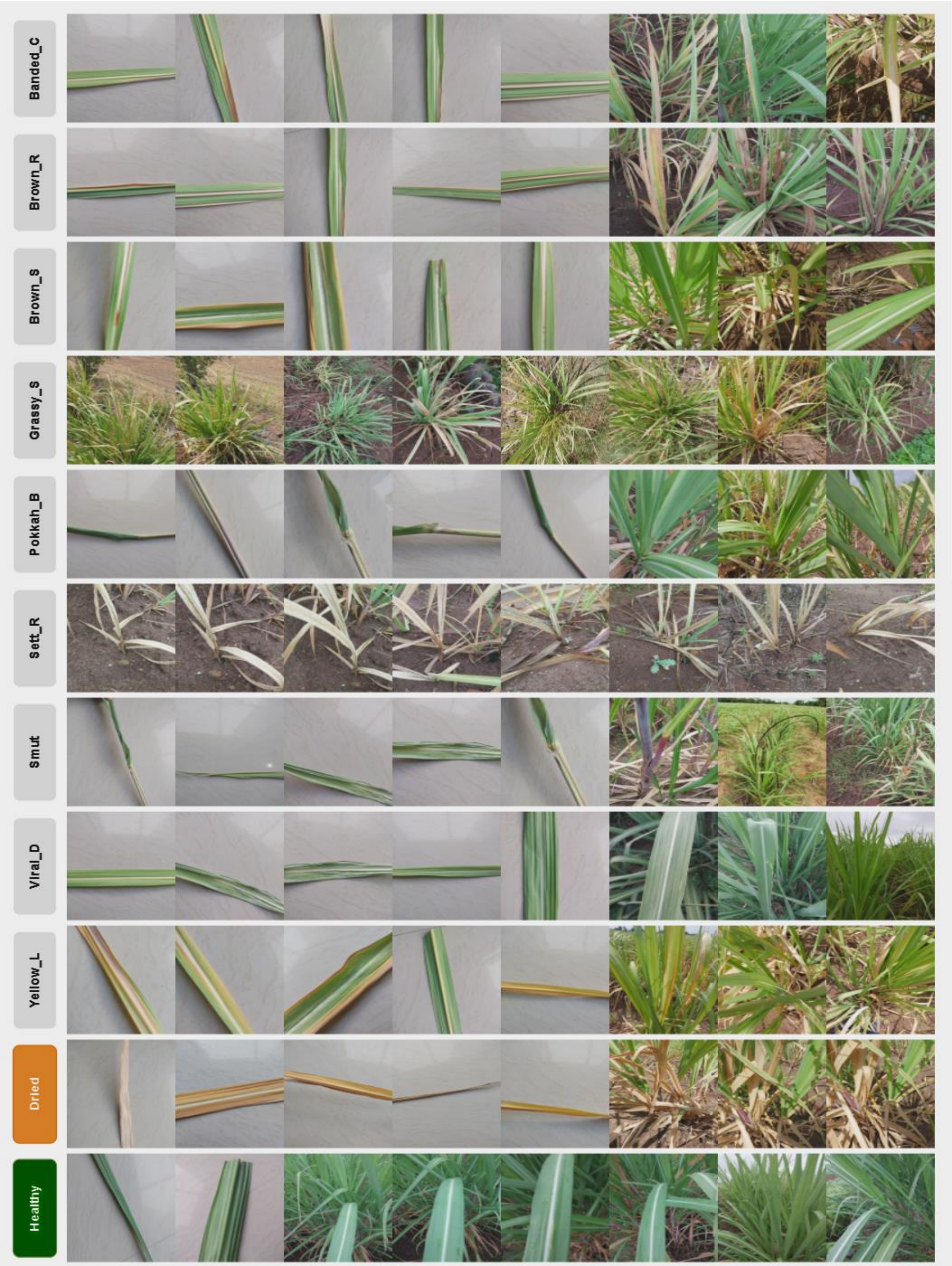


**Fig. 2.** Examples of leaf images in the sugarcane leaf dataset

The Sugarcane Leaf Dataset comprises 6748 high-resolution images of sugarcane leaves, categorized into 11 disease classes including dried leaves and healthy leaves [36]. It covers a range of common foliar diseases, making it easy to access and identify specific examples of diseases. This dataset enables the detection of diseases caused by sugarcane leaves. The aim of this study was to classify diseases using these images. Figure 2 illustrates examples of leaf classes in the sugarcane leaf dataset, which includes images of Banded Chlorosis, Brown Rust, Brown Spot, Grassy Shoot, Pokkah boeng, Sett Rot, Smut, Viral Disease, Yellow Leaf diseases, and images of Dried Leaves and Healthy Leaves plant leaves.

## 4. Results

Table 2 provides performance metrics for Deit, Swin ViT models and CNN models like ResNet50, Xception and EfficientNetv2-Small. The DeiT3-Small model shows the highest performance when the given metrics are examined. However, it requires a few parameters. The DeiT-Tiny model is notable because it provides high accuracy and good balance in least parameters.

**Table 2**
The performance measures of classification by DeiT and other CNN/ViT models

| Model | Model Architecture | Params (M) | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|---|---|
| ResNet50 | CNN-Based | 23.52 | 0.9260 | 0.9887 | 0.8907 | 0.8916 |
| Xception | CNN-Based | 20.83 | 0.9290 | 0.9039 | 0.8944 | 0.8920 |
| EfficientNetv2-Small | CNN-Based | 20.19 | 0.9300 | 0.9068 | 0.9081 | 0.9036 |
| ViT-Base | ViT-Based | 85.81 | 0.9300 | 0.9078 | 0.8904 | 0.8928 |
| DeiT-Tiny | ViT-Based | 5.53 | 0.9310 | 0.9069 | 0.9017 | 0.8980 |
| DeiT-Small | ViT-Based | 21.67 | 0.9221 | 0.8951 | 0.8834 | 0.8850 |
| DeiT-Base | ViT-Based | 85.81 | 0.9280 | 0.9093 | 0.9007 | 0.8984 |
| DeiT3-Small | ViT-Based | 21.68 | 0.9379 | 0.9127 | 0.9099 | 0.9096 |
| DeiT3-Medium | ViT-Based | 38.34 | 0.9260 | 0.8967 | 0.8917 | 0.8937 |
| DeiT3-Base | ViT-Based | 85.82 | 0.9221 | 0.8906 | 0.8939 | 0.8893 |
| DeiT3-Large | ViT-Based | 303.36 | 0.9250 | 0.9155 | 0.8886 | 0.8905 |
| Swin-Base | ViT-Based | 86.75 | 0.9250 | 0.8934 | 0.8893 | 0.8874 |

If we look at what the given metrics mean; ResNet50 is not the model with the lowest number of parameters, but it has a very high accuracy (92.60%) and F1-score (89.16). This model can achieve higher accuracy with fewer parameters than all other ViT-based models. Although the number of parameters is slightly lower than ResNet50 (20.83 M), Xception has almost the same accuracy (92.90%) and a slightly lower F1-score value (89.20). Xception is generally known as an effective model in visual recognition tasks. EfficientNetv2-Small, whose number of parameters is 20.19 M, has a very high accuracy (93.00%) and F1-score (90.36). Although the number of parameters is low, it performs effectively.

ViT-Base, which has the highest number of parameters (85.81 M) among Vision Transformer-based models, performs very well with 93.00% accuracy and 89.28 F1-score. However, compared to other ViT-based models, such as DeiT-Tiny, DeiT-Small, and Swin-Base, they have close accuracy values with fewer parameters (5.53 M, 21.67 M, and 86.75 M). With only 5.53 M parameters, DeiT-Tiny performs very effectively with 93.10% accuracy and 89.80 F1-score. It has the lowest number of parameters among other DeiT models. With 21.68 M parameters, DeiT3-Small shows a high performance with 93.79% accuracy and 90.96 F1-score. It is one of the models with the lowest number of parameters in the DeiT3 family. With 303.36 M parameters, DeiT3-Large shows effective performance despite a large number of parameters, with an accuracy of 92.50% and an F1-score of 89.05. However, since the number of parameters is high, it has a lower accuracy and F1-score value compared to other models. With 86.75 M parameters, Swin-Base performs effectively among ViT-based models with an accuracy of 92.50% and an F1-score of 88.74. It has similar performance to other ViT-based models.

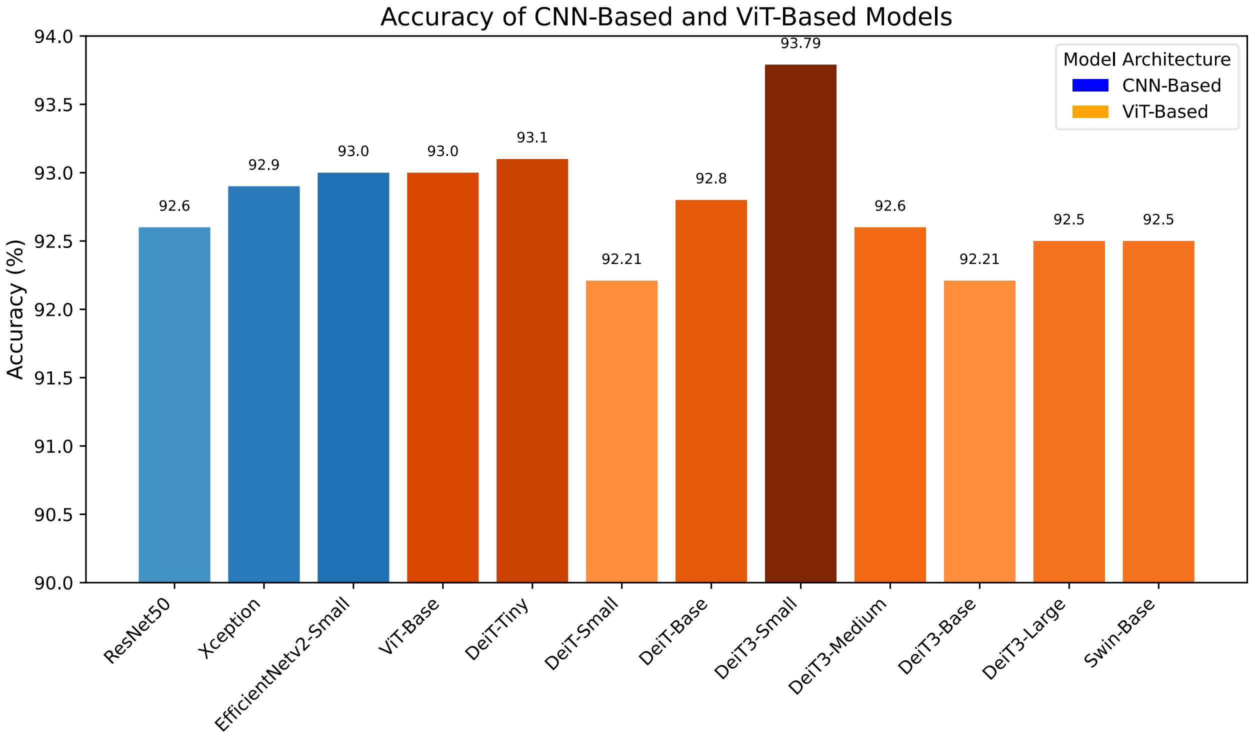


**Fig. 3.** Accuracy of CNN-Based and ViT-Based models

Accordingly to Figure 3, it seems that models with both CNN-Based and ViT-Based architectures have very high accuracy values. While the accuracy values of CNN-Based models vary between 92.60% and 93.00%, the accuracy values of ViT-Based models vary between 92.21% and 93.79%. These results show that both architectures are successful and accomplish the task successfully. In particular, the DeiT3-Small model stands out as the highest-performing model with 93.79% accuracy. Other ViT-Based models also have very close accuracy values between 92.50% and 92.80%. However, CNN-Based models also achieved very competitive results. These results show that both CNN-Based and ViT-Based models can be used effectively in a variety of visual tasks.

F1-score is a metric that measures the performance of a classification model. In addition to accuracy, it provides a combination of a model's precision and recall metrics. F1-score helps us understand how accurate and comprehensive predictions a model makes.

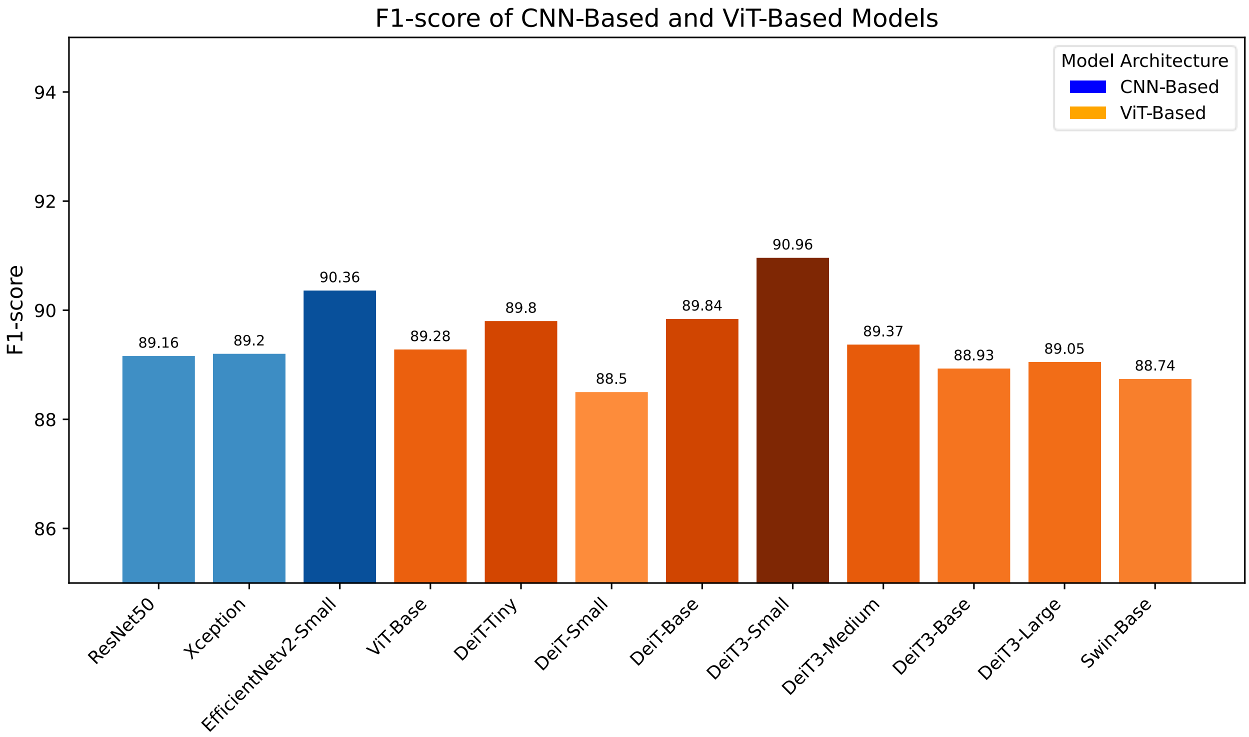


**Fig. 4.** F1-Score of CNN-Based and ViT-Based models

Accordingly, Figure 4, when the F1-score values are examined, it is seen that both CNN-Based and ViT-Based models are successful in the tasks. F1-score values of CNN-Based models generally range between 89.16 and 90.36. Among these models, the EfficientNetv2-Small model exhibits the highest performance with an F1-score of 90.36. F1-score values of ViT-Based models vary between 88.50 and 90.96. The DeiT3-Small model stands out as the highest performing ViT-Based model with an F1-score of 90.96.

In general, it appears that ViT-Based models have high F1-score values and a few models perform best. However, it is observed that CNN-Based models also obtain quite competitive results. As a result, it can be said that both CNN-Based and ViT-Based models can be used successfully in various tasks and prove their performance with F1-score values.

The confusion matrix is a metric table utilized to assess how well a classification model performs. It demonstrates the correlation between the predicted classes by the model and the actual classes. Commonly applied in classification tasks, the confusion matrix serves as a foundation for computing the model's accuracy, sensitivity, specificity, and performance metrics like recall and F1-score.

The number of true positives (TP) refers to the number of positive examples that the model predicted correctly. The number of true negatives (TN) refers to the number of negative examples that the model correctly predicted. The number of false positives (FP) refers to the number of examples that the model predicted as positive but were negative. The number of false negatives (FN) refers to the number of examples that the model predicted as negative but were positive. These four values indicate how accurately or incorrectly the model predicted each class. Confusion matrix is very important for understanding the performance of the model and is used in the development and tuning of classification models. As seen in Figure 5, the DeiT3-Small and DeiT-Tiny models have high TP rates and low FP and FN errors. Differences were observed between classes; Some classes are generally predicted with high accuracy across all models, while some classes are predicted with low accuracy across all models. Each model has its strengths and weaknesses; Which model to use may depend on the performance on a particular class or feature set.

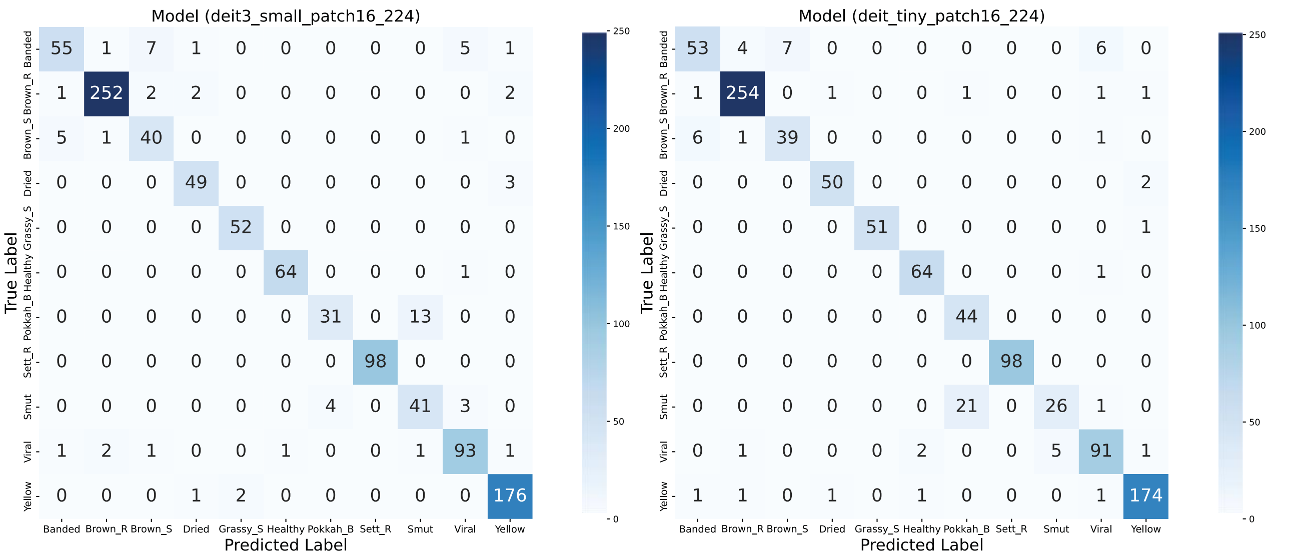


**Fig. 5.** Confusion matrix for DeiT3-Small and DeiT-Tiny

## 5. **Discussion and Conclusion**

Recently, deep learning methods have become popular for image processing. In this study, the classification of sugarcane leaf images belonging to 11 classes was examined on the Sugarcane Leaf Dataset. For this purpose, Vit and CNN models were examined. In this review, CNN models Resnet50, Xception, EfficientNetv2-Small and ViT models as well as Vit-Base and Swin models as well as Deit architecture models were examined. The highest accuracy rate was found in the DeiT3-Small (0.9379) model. Even models with low accuracy such as EfficientNet-b5 (0.9053) and EfficientNetv2-Large (0.9014) achieved high accuracy. However, it is seen that other models still provide high accuracy, with DeiT-Small (0.9221) and DeiT3-Base (0.9221) being the models with the worst results. In addition to these results, it was concluded that DeiT-Tiny achieved high accuracy (0.9310) with the lowest parameter (5.53m). In conclusion, the study shows the performance differences of different model architectures. Each model has strengths and weaknesses, so consideration should be given to which model will perform best in a particular class or feature set. For more realistic results, it may be recommended to work with larger data sets. Additionally, it would be advisable to validate the proposed model in live application in future studies.

### **Author Contributions**

Conceptualization, İ.K.; methodology, İ.K and İ.P.; software, İ.P.; validation, İ.K and İ.P.; investigation, İ.K. and İ.P.; resources, İ.K.; writing—original draft preparation, İ.K.; writing—review and editing, İ.K. and İ.P.; visualization, İ.K.; supervision, İ.P. All authors have read and agreed to the published version of the manuscript.

### **Funding**

This research received no external funding.

## Data Availability Statement

We used a public dataset which Sugarcane Leaf Image Dataset. For access https://data.mendeley.com/datasets/9twjtv92vk/1

## Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgment

This research was not funded by any grant.

## References

[1] Mueller, D. S., Wise, K. A., Sisson, A. J., Allen, T. W., Bergstrom, G. C., Bosley, D. B., Bradley, C. A., Broders, K. D., Byamukama, E., Chilvers, M. I., Collins, A., Faske, T. R., Friskop, A. J., Heiniger, R. W., Hollier, C. A., Hooker, D. C., Isakeit, T., Jackson-Ziems, T. A., Jardine, D. J., … Warner, F. (2016). Corn Yield Loss Estimates Due to Diseases in the United States and Ontario, Canada from 2012 to 2015. Plant Health Progress, 17(3), 211–222. https://doi.org/10.1094/PHP-RS-16-0030

[2] Fowler, H. J., Kilsby, C. G., & O'Connell, P. E. (2003). Modeling the impacts of climatic change and variability on the reliability, resilience, and vulnerability of a water resource system. Water Resources Research, 39(8). https://doi.org/10.1029/2002WR001778

[3] Kunduracıoğlu, İ. (2018). Examining the interface of lego mindstorms ev3 robot programming. Journal of Educational Technology and Online Learning, 1(1), 28-46.. https://doi.org/10.31681/jetol.372826

[4] Kunduracıoğlu, İ., & Durak, G. (2018). A CONTENT ANALYSIS ON GAMIFICATION. European Journal of Open Education and E-Learning Studies, 3(2). https://doi.org/10.5281/zenodo.2228190

[5] Javaid, M., Haleem, A., Singh, R. P., & Suman, R. (2022). Enhancing smart farming through the applications of Agriculture 4.0 technologies. International Journal of Intelligent Networks, 3, 150–164. https://doi.org/10.1016/j.ijin.2022.09.004

[6] Arjunagi, S., & Patil, N. B. (2023). Optimized convolutional neural network for identification of maize leaf diseases with adaptive ageist spider monkey optimization model. International Journal of Information Technology, 15(2), 877–891. https://doi.org/10.1007/s41870-021-00657-3

[7] He, J., Liu, T., Li, L., Hu, Y., & Zhou, G. (2023). MFaster R-CNN for Maize Leaf Diseases Detection Based on Machine Vision. Arabian Journal for Science and Engineering, 48(2), 1437–1449. https://doi.org/10.1007/s13369-022-06851-0

[8] Viswanathan, R., & Rao, G. P. (2011). Disease Scenario and Management of Major Sugarcane Diseases in India. Sugar Tech, 13(4), 336–353. https://doi.org/10.1007/s12355-011-0102-4

[9] Bashir, S., & Sharma, N. (2012). Remote Area Plant Disease Detection Using Image Processing. In IOSR Journal of Electronics and Communication Engineering, 2(6), 31-34. http://dx.doi.org/10.9790/2834-0263134

[10] Taherei Ghazvinei, P., Darvishi, H. H., Mosavi, A., Bin Wan Yusof, K., Alizamir, M., Shamshirband, S., & Chau, K. W. (2018). Sugarcane growth prediction based on meteorological parameters using extreme learning machine and artificial neural network. Engineering Applications of Computational Fluid Mechanics, 12(1), 738–749. https://doi.org/10.1080/19942060.2018.1526119

[11] Hamuda, E., Mc Ginley, B., Glavin, M., & Jones, E. (2017). Automatic crop detection under field conditions using the HSV colour space and morphological operations. Computers and Electronics in Agriculture, 133, 97–107. https://doi.org/10.1016/j.compag.2016.11.021

[12] Akbarzadeh, S., Paap, A., Ahderom, S., Apopei, B., & Alameh, K. (2018). Plant discrimination by Support Vector Machine classifier based on spectral reflectance. Computers and Electronics in Agriculture, 148, 250–258. https://doi.org/10.1016/j.compag.2018.03.026

[13] Hoang Trong, V., Gwang-hyun, Y., Thanh Vu, D., & Jin-young, K. (2020). Late fusion of multimodal deep neural networks for weeds classification. Computers and Electronics in Agriculture, 175, 105506. https://doi.org/10.1016/j.compag.2020.105506

[14] Bhosle, K., & Ahirwadkar, B. (2021). Deep learning convolutional neural network (cnn) for cotton, mulberry and sugarcane classification using hyperspectral remote sensing data. Journal of Integrated Science and Technology, 9(2), 70-74.

[15] Kunduracioglu, I., & Pacal, I. (2024). Advancements in deep learning for accurate classification of grape leaves and diagnosis of grape diseases. Journal of Plant Diseases and Protection, 131(3), 1061-1080. https://doi.org/10.1007/s41348-024-00896-z

[16] Khan, A., Yadav, M. S., & Ahmad, S. (2017). Image processing based disease detection for sugarcane leaves. International Journal of Advance Research, Ideas and Innovations in Technology, 3(4), 497-502. https://www.jetir.org/view?paper=JETIR1806177

[17] Arivazhagan, S., Shebiah, R. N., Ananthi, S., & Varthini, S. V. (2013). Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features. Agricultural Engineering International: CIGR Journal, 15(1), 211-217.

[18] Hashemi-Beni, L., Gebrehiwot, A., Karimoddini, A., Shahbazi, A., & Dorbu, F. (2022). Deep Convolutional Neural Networks for Weeds and Crops Discrimination From UAS Imagery. Frontiers in Remote Sensing, 3. https://doi.org/10.3389/frsen.2022.755939

[19] Vezıroglu, E., Pacal, I., & Coşkunçay, A. (2023). Derin Evrişimli Sinir Ağları Kullanılarak Pirinç Hastalıklarının Sınıflandırılması. Journal of the Institute of Science and Technology, 13(2), 792-814. https://doi.org/10.21597/jist.1265769

[20] Kılıçarslan, S., & Pacal, I. (2023). Domates Yapraklarında Hastalık Tespiti İçin Transfer Öğrenme Metotlarının Kullanılması. Mühendislik Bilimleri ve Araştırmaları Dergisi, 5(2), 215-222. https://doi.org/10.46387/bjesr.1273729

[21] Pacal, I. (2024). Enhancing crop productivity and sustainability through disease identification in maize leaves: Exploiting a large dataset with an advanced vision transformer model. Expert Systems with Applications, 238, 122099. https://doi.org/10.1016/j.eswa.2023.122099

[22] Srinivas, B., Satheesh, P., Rama Santosh Naidu, P., & Neelima, U. (2021). Prediction of guava plant diseases using deep learning. In ICCCE 2020: Proceedings of the 3rd International Conference on Communications and Cyber Physical Engineering (pp. 1495-1505). Springer Singapore. https://doi.org/10.1007/978-981-15-7961-5_135

[23] Karmokar, B. C., Ullah, M. S., Siddiquee, M. K., & Alam, K. M. R. (2015). Tea leaf diseases recognition using neural network ensemble. International Journal of Computer Applications, 114(17), 27-30. https://doi.org/10.5120/20071-1993

[24] Baranwal, S., Khandelwal, S., & Arora, A. (2019). Deep Learning Convolutional Neural Network for Apple Leaves Disease Detection. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3351641

[25] Goluguri, N. V. R. R., Devi, K. S., & Srinivasan, P. (2021). Rice-net: an efficient artificial fish swarm optimization applied deep convolutional neural network model for identifying the Oryza sativa diseases. Neural Computing and Applications, 33(11), 5869–5884. https://doi.org/10.1007/s00521-020-05364-x

[26] Militante, S. V., & Gerardo, B. D. (2019, December). Detecting sugarcane diseases through adaptive deep learning models of convolutional neural network. In 2019 IEEE 6th International Conference on Engineering Technologies and Applied Sciences (ICETAS) (pp. 1-5). IEEE. https://doi.org/10.1109/ICETAS48360.2019.9117332

[27] Chen, W., Ju, C., Li, Y., Hu, S., & Qiao, X. (2021). Sugarcane Stem Node Recognition in Field by Deep Learning Combining Data Expansion. Applied Sciences, 11(18), 8663. https://doi.org/10.3390/app11188663

[28] Wang, W., Li, C., Wang, K., Tang, L., Ndiluau, P. F., & Cao, Y. (2022). Sugarcane stem node detection and localization for cutting using deep learning. Frontiers in Plant Science, 13. https://doi.org/10.3389/fpls.2022.1089961

[29] Borhani, Y., Khoramdel, J., & Najafi, E. (2022). A deep learning based approach for automated plant disease classification using vision transformer. Scientific Reports, 12(1), 11554. https://doi.org/10.1038/s41598-022-15163-0

[30] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929. http://arxiv.org/abs/2010.11929

[31] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. https://doi.org/10.1038/nature14539

[32] Huang, Y. L., & Fan, L. T. (1993). Artificial intelligence for waste minimization in the process industry. Computers in Industry, 22(2), 117–128. https://doi.org/10.1016/0166-3615(93)90059-A

[33] Zhong, J., Yang, B., Huang, G., Zhong, F., & Chen, Z. (2016). Remote Sensing Image Fusion with Convolutional Neural Network. Sensing and Imaging, 17(1), 1–16. https://doi.org/10.1007/S11220-016-0135-6/TABLES/4

[34] Jha, K., Doshi, A., Patel, P., & Shah, M. (2019). A comprehensive review on automation in agriculture using artificial intelligence. Artificial Intelligence in Agriculture, 2, 1–12. https://doi.org/10.1016/j.aiia.2019.05.004

[35] Li, X., Li, X., Zhang, S., Zhang, G., Zhang, M., & Shang, H. (2023). SLViT: Shuffle-convolution-based lightweight Vision transformer for effective diagnosis of sugarcane leaf diseases. Journal of King Saud University - Computer and Information Sciences, 35(6), 101401. https://doi.org/10.1016/j.jksuci.2022.09.013

[36] Thite, S., Suryawanshi, Y., Patil, K., & Chumchu, P. (2024). Sugarcane leaf dataset: A dataset for disease detection and classification for machine learning applications. Data in Brief, 53. https://doi.org/10.1016/j.dib.2024.110268