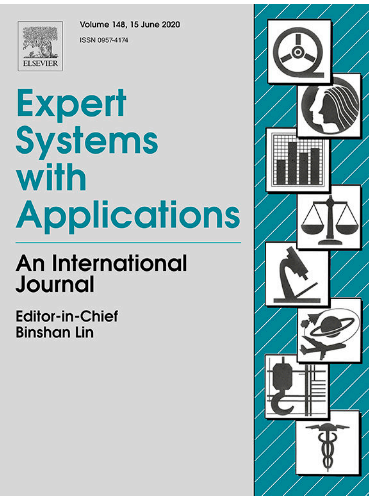


A novel data-centric AI approach based on sensitivity and correlation analyses for multi-organ plant disease classification

Muhammad Hammad Saleem, Fakhia Hammad, Muhammad Taha,
Shivakumara Palaiahnakote, Sadaqat ur Rehman, Mohamad Saraee

PII: S0957-4174(25)01984-0
DOI: https://doi.org/10.1016/j.eswa.2025.128365
Reference: ESWA 128365

To appear in: *Expert Systems with Applications*

Received Date: 7 December 2024
Revised Date: 5 March 2025
Accepted Date: 26 May 2025

Please cite this article as: Saleem, M.H., Hammad, F., Taha, M., Palaiahnakote, S., Rehman, S.u., Saraee, M., A novel data-centric AI approach based on sensitivity and correlation analyses for multi-organ plant disease classification, *Expert Systems with Applications* (2025), doi: https://doi.org/10.1016/j.eswa.2025.128365

A Novel Data-Centric AI Approach Based on Sensitivity and Correlation Analyses for Multi-Organ Plant Disease Classification

Muhammad Hammad Saleem [a,b,*], Fakhia Hammad [c], Muhammad Taha [d], Shivakumara Palaiahnakote [a,b], Sadaqat ur Rehman [a,b], Mohamad Saraee [a,b]

[a] School of Science, Engineering, and Environment, University of Salford, Salford M5 4WT, United Kingdom.

[b] Data Science and Artificial Intelligence Hub, University of Salford, Salford M5 4WT, United Kingdom; m.h.saleem@salford.ac.uk, S.Palaiahnakote@salford.ac.uk, s.rehman15@salford.ac.uk, m.saraee@salford.ac.uk

[c] Department of Biomedical Engineering, Ziauddin University, Karachi 74600, Pakistan; fakhia.hammad@gmail.com

[d] Department of Electrical Engineering, National University of Computer & Emerging Sciences, Karachi 75030, Pakistan; engr.muhammadtaha95@gmail.com

* Corresponding author: m.h.saleem@salford.ac.uk

Abstract: With advancements in deep learning (DL), most research on classification problems has focused on developing or modifying DL models, known as model-centric artificial intelligence (AI) approaches. However, this approach is time-consuming and overlooks the exploration of the available resources and expertise required to address industrial problems. This study proposes a new data-centric AI-based approach by thoroughly investigating dataset complexities, using multi-organ plant disease classification as a case study. To the best of our knowledge, this study is the first to perform comprehensive sensitivity and correlation analyses to evaluate the relationship between dataset complexity exclusion and the accuracy of DL classifier. In contrast to conventional sensitivity analyses which only evaluate changes in model output with respect to input changes, this study introduces a novel Sensitivity Correlation Score (SC-score). The SC-score combines sensitivity and correlation analyses into a single metric formulated as the product of the Absolute Sensitivity Function and Pearson Correlation Coefficient which is normalised for interpretability. This formulation rewards positive sensitivity and strong correlation while neutralising the effects of negative correlation. The SC-score successfully evaluated both the responsiveness and consistency of the performance enhancement of the DL model owing to the elimination of dataset complexities. To demonstrate the robustness of this study, the proposed data-centric DL-based approach was validated on an external testing dataset from diverse agricultural environments and achieved an accuracy improvement of 10.94%. This study demonstrates the strength of data-centric AI in solving industry-oriented problems in real-world applications.

Keywords: Model-centric AI**,** data-centric AI, deep learning, convolutional neural networks, plant disease classification.

1. Introduction

When the dataset is complex owing to diversified images with different practical scenarios, such as uneven illumination effects, the presence of shadows, occlusions, and complex backgrounds, a model-centric approach is usually applied to solve it. In other words, a new/modified Deep Learning (DL) model is developed, hyperparameter tuning is performed, or optimisation algorithms are employed to improve the performance of certain tasks. However, a model-centric AI approach is not always the most feasible solution for addressing industry-oriented problems, where computational resources, time, and expertise are often limited. For instance, farmers in the agricultural sector typically operate under tight budgets and schedules. In this case, timely, accurate, and cost-effective solutions are required to achieve maximum crop productivity and minimise crop loss. Model-centric AI is a resource-intensive approach that requires significant computing power, such as Graphics Processing Units (GPUs), tensor processing units (TPUs), and expertise in Machine Learning (ML) algorithms. Furthermore, the development of ML models is a time-consuming process because of several iterations of training, testing, and fine-tuning. In contrast, this study introduces a new data-centric AI approach that focuses on thoroughly studying the effects of dataset complexity to improve the performance of existing DL models. By excluding these dataset complexities, a data-centric approach not only enhances the performance of the DL model but also ensures its scalability and practicality for real-world deployment. To demonstrate that data-centric DL is an alternative approach for achieving better performance, this study considers multi-organ plant disease classification as a case study. This classification problem was selected because of its real-world application in the agricultural industry, which is currently required to assist farmers in making their products profitable and enhancing crop productivity.

The classification and identification of plant diseases is not a new problem. Several methods have been reported in the literature. With the rapid development of artificial intelligence (AI), the scientific community has actively engaged in deep learning (DL) research to address agricultural problems. In the last decade, several studies on plant disease classification have proven the strength of DL in obtaining accurate and precise outcomes. At the beginning of research on AI-based plant disease classification, DL classifiers were applied to datasets collected in a controlled agricultural environment (Brahimi, et al., 2018; Mohanty, Hughes, & Salathé, 2016; Too, Yujian, Njuki, & Yingchun, 2019). A few studies have curated datasets in real agricultural environments and proven the effectiveness of DL methods under complex agricultural scenarios (Lee, Goëau, Bonnet, & Joly, 2020; Narayanan, et al., 2022; Pandi, et al., 2022; Ravi, Acharya, & Pham, 2022). Moreover, some studies have included different plant organs and generated datasets to further demonstrate the efficiency of DL in detecting plant diseases in real-world scenarios (Saleem, Potgieter, & Arif, 2022a, 2022b). Various state-of-the-art DL models have been trained using scratch- and transfer-learning/fine-tuning methods (Brahimi, Boukhalfa, & Moussaoui, 2017; J. Chen, Chen, Zhang, Sun, & Nanehkaran, 2020). Subsequently, the research community started focusing on the optimization and modification of DL models to improve the accuracy of plant disease classification (Kamal, Yin, Wu, & Wu, 2019; Kaur, Harnal, Gautam, Singh, & Singh, 2023; Tewari & Kumari, 2024). Later, complex and practical agricultural scenarios were considered, and researchers applied DL object detection and segmentation models to locate and segment disease spots on plants (J. Liu & Wang, 2020; K. Zhang, Wu, & Chen, 2021; S. Zhao, Liu, & Wu, 2022) along with their segmented pixels (Mukhopadhyay, Paul, Pal, & De, 2021; Mzoughi & Yahiaoui, 2023; S. Zhang & Zhang, 2023). These studies optimised the DL model to successfully classify and detect plant diseases and are therefore called model-centric AI.

However, some studies have demonstrated the capability of systematically evaluating and improving the quality of datasets to obtain accurate outcomes using a type of AI called Data-centric AI. These studies addressed problems such as overfitting, class imbalance, limited datasets, and dataset variability. In a feedback loop representation, a typical model-centric AI feeds back the output as the accuracy of the DL model for feature extraction to enhance the accuracy of plant disease classification. In contrast, the data-centric AI-based method sends the output back to the data preprocessing step. A diagrammatic representation is shown in Fig. 1, where one can see the difference between model-centric and data-centric AI can be observed.

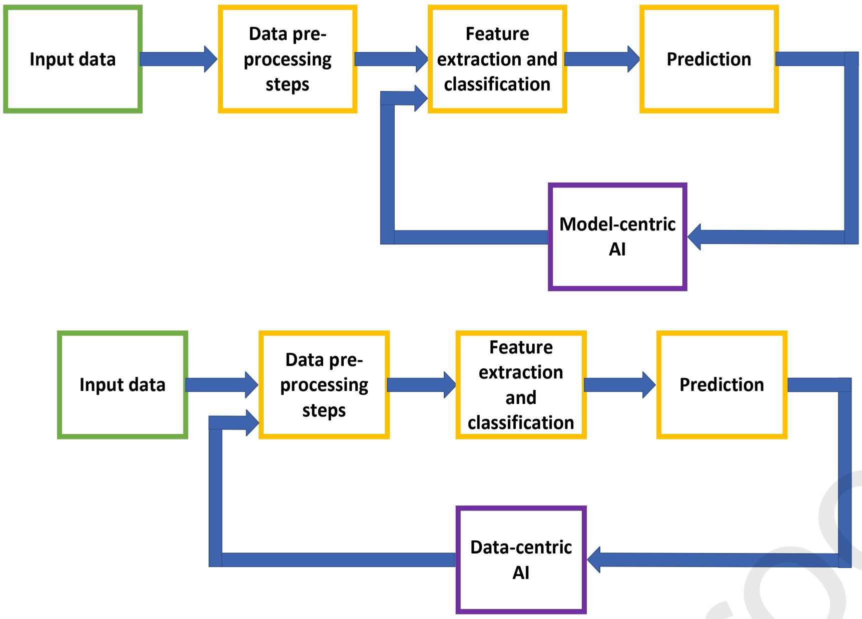


Fig. 1. Difference between model-centric and data-centric AI in feedback loop representation.

Furthermore, studies in the literature have improved the accuracy of plant disease classification by applying augmentation-based methods and studying dataset quality. For example, a study considered individual lesions and spots of plant diseases instead of full-leaf images to classify plant diseases (Barbedo, 2019; Ngugi, Abdelwahab, & Abo-Zahhad, 2023; Sun, Zhang, Yang, & Liu, 2020). Traditional augmentation methods have been used to improve the dataset quantity for plant disease classification (Saleem, et al., 2022b). Various DL-based data augmentation methods have been developed in prior studies. For instance, a modified colour-value distribution-based data augmentation method was presented in a study to expand the dataset and train a DL model to identify important colour-based features (Abayomi-Alli, Damaševičius, Misra, & Maskeliūnas, 2021). A study proposed an image-to-image translation model to address the data bias problem (Min, Kim, Shin, & Shin, 2023). Furthermore, prominent studies have focused on generative adversarial network (GAN)-based methods to augment or enhance plant disease datasets. For instance, a deep convolutional generative adversarial network (DCGAN) method was proposed to detect tomato leaf disease, which not only expanded the dataset but also added diversity to it (Wu, Chen, & Meng, 2020). A study proposed a method called LeafGAN, which consists of an image-to-image translation system along with its attention mechanism (Cap, Uga, Kagiwada, & Iyatomi, 2020). This method transformed relevant areas of the image with different backgrounds, and its performance was compared with that of the Vanilla CycleGAN. The research presented two methods of augmenting datasets, including plant canopy simulation and GANs, to enhance the segmentation of plant diseases (Douarre, Crispim-Junior, Gelibert, Tougne, & Rousseau, 2019).

To achieve high accuracy, DL-based methods have been developed to address the problem of small plant disease datasets. For example, a study proposed a few-shot learning approach for classifying plant diseases by splitting the dataset into source and domain classes. The patterns of plant diseases were learned using Inception V3 through source classes to classify the domain classes. Siamese networks and triplet loss were used and compared with the traditional transfer learning method (Argüeso, et al., 2020). A research focused on the inconsistencies in the annotations of the plant disease dataset and analysed their impact on the performance of the DL model (Dong, et al., 2022). The paper highlighted better outcomes in terms of the mean average precision. A sensitivity analysis was also performed to show the effects of different levels of annotation inconsistencies on average precision. Although the research primarily emphasized dataset inconsistencies, various aspects of data-centric AI could have been further explored, such as the robustness of the approach, consideration of complexities in plant disease in real-world circumstances, and the relationship between one dataset annotation inconsistency and another.

The prominent literature review on data-centric AI for plant disease classification mainly consists of proposing noise filters (Bhujade, Sambhe, & Banerjee, 2024), using traditional data-augmentation methods (Arun & Umamaheswari, 2023), developing Generative Adversarial Networks (GANs) (Sharma, Tripathi, Daga, Nidhi, & Mittal, 2024), creating few-shot learning methods (FSL) (Rezaei, Diepeveen, Laga, Jones, & Sohel, 2024) and applying metaheuristic algorithms (Taji, et al., 2024). These methods mainly focus on removing noise, studying the effects of augmented images, generating synthetic dataset images, addressing the problem of limited datasets to train DL models, and selecting prominent features of plant diseases. In comparison to these approaches, the proposed research focuses on analysing the effects of eliminating dataset complexities to improve the accuracy of the DL model supported by comprehensive sensitivity and correlation analyses.

To the best of the authors' knowledge, none of the previous studies have addressed various research questions in the domain of data-centric AI for plant disease classification. These questions include (a) How does the exclusion of dataset complexities affect the accuracy of DL models in plant disease classification? (b) How sensitive are DL-based plant disease classification models to the rate of change/exclusion of dataset complexity? (c) Is there a correlation between the elimination of dataset complexity and the classification performance of the DL models? (d) How robust is the performance of the DL model after excluding dataset complexities from an external testing dataset that contains images collected in diverse agricultural environments?

This study aims to address these gaps by proposing a novel data-centric AI approach based on sensitivity and correlation analyses. The key objectives of this study are as follows:

- To investigate the impact of dataset complexity on the accuracy of the DL model.
- To examine the sensitivity and correlation between dataset complexity exclusion and the performance enhancement of the DL model.
- To evaluate the robustness and practical aspects of the proposed novel data-centric AI approach on an external dataset that includes images of agricultural conditions different from those of the selected plant disease dataset.
- To measure the combined effect of sensitivity and correlation analyses, identifying the most influential step sizes of dataset complexities exclusion.

To fulfil the research objectives, state-of-the-art DL methods were trained and tested on a plant disease dataset called NZDLPlantDisease-v1 and external testing datasets to select the most accurate model. A class-wise investigation was then performed to identify classes with low accuracy. Subsequently, dataset complexities in low-accuracy classes were identified in real-world scenarios based on image quality measures such as the Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) and Sharpness Index (SI). Next, these complexities are excluded individually and cumulatively (if the exclusion of more than one complexity contributes to improving the accuracy). Sensitivity and correlation analyses were also conducted. For the sensitivity analyses, the absolute sensitivity function was applied to show how the exclusion of dataset complexities affected the classification accuracy for different step sizes. Correlation analysis was performed using Pearson's correlation coefficients. This study proposes a novel performance score, called the sensitivity correlation score (SC-score), to analyse the significance of the proposed approach. Hence, this study provides new insights into DL-based plant disease classification and encourages a shift in the research focus to data-centric AI for agricultural and other real-world problems.

To illustrate the advantage of the proposed Sensitivity Correlation Score over sensitivity and correlation alone, a sample of the sensitivity, correlation, and SC-score is presented in Fig. 2. It can be noted that some of the step sizes of dataset complexity exclusion have a negative sensitivity and positive correlation and vice versa, a few instances gave the SC-score value of '0.' These observations indicate scenarios where the selection of the most affected step size of exclusion of dataset complexity would be crucial; hence, a new performance score would be useful based on combined sensitivity and correlation analysis.

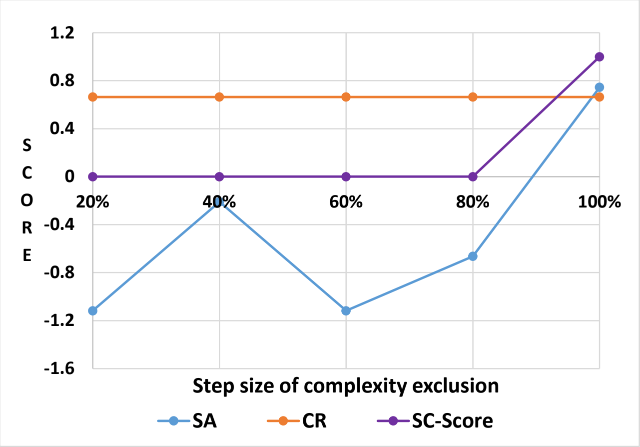


Fig. 2. A sample of absolute sensitivity (SA), correlation (CR), and proposed SC-score after excluding dataset complexity (non-focused object of interest) from plant disease class (pear canker).

The main contributions of this study are as follows:

- A new data-centric DL-based approach is proposed to improve the accuracy of plant disease classification by adding two steps based on a feedback loop representation, including the exclusion of dataset complexities and detailed analyses of changes in accuracy owing to dataset complexity elimination.
- Sensitivity analysis was performed using the absolute sensitivity function based on the ratio of the change in accuracy and step size to eliminate dataset complexity. Detailed Sensitivity analyses demonstrated that the deep learning model is sensitive to certain complexities, both individually and cumulatively. Moreover, comprehensive correlation analyses were performed using the Pearson Correlation Coefficient to extract the relationship between the elimination of dataset complexity and accuracy improvement of the DL model.
- A new performance score, named the Sensitivity Correlation Score (SC-score), is proposed to select the most affected step size by excluding dataset complexity. The robustness of the research was evaluated by testing the proposed novel approach on an external testing dataset in diverse agricultural conditions other than the selected plant disease dataset.

The remainder of this manuscript is organised as follows: Section 2 highlights the related work; Section 3 describes the proposed data-centric DL-based approach; Section 4 presents the experimental results; and Section 5 outlines the conclusions and future directions.

2. Related works

Most of the work on data-centric AI for plant disease classification has focused on applying traditional data augmentation methods, using/advancing Generative Adversarial Network (GANs)-based methods, addressing the problem of limited datasets to achieve high accuracy, and feature selection methods. This section reviews research on these methods, dividing them into four categories.

2.1. Traditional Data Enhancement Methods

Some initial studies on the application of data augmentation methods to classify plant diseases used traditional data augmentation methods. The research used methods such as flipping, random crops, rotations, shifts, and a combination of these methods to augment the dataset. These methods were applied to improve the generalisability and robustness of the DL method; however, an intensive investigation of the effects of these methods has not been discussed (Kamal, et al., 2019). A study explored the effects of individual lesions and spots of plant disease without adding new images to the dataset, which improved the overall accuracy by 12% compared to that achieved by training the DL model using the original images (Barbedo, 2019). A few limitations of this research were pointed out in the paper, such as finding an appropriate number of images to train the features of plant diseases, training and testing datasets taken from the same resource/database, misclassification in the extended dataset owing to class imbalance problems, and the need for a flexible neural network that can be trained on images with variable resolutions. A research evaluated the effects of augmented datasets on the developed DCNN and achieved an improvement of 0.0644 in the validation accuracy. A previous study generated synthetic images to identify important colour features by proposing an image colour histogram transformation technique, resulting in an increase of 3% to 15% accuracy for reduced-quality images (Abayomi-Alli, et al., 2021). An image-to-image translation and attention mechanism-based method was developed in a study that achieved F1 scores of 0.9995, 0.9994, and 0.999 for apple, grape, and potato disease classification, respectively (Min, et al., 2023). A few recent studies have used conventional data augmentation methods for crop disease identification, including shearing, rotation, width and height shifting, horizontal flipping, and zooming (Arun & Umamaheswari, 2023). An article applied augmentation techniques such as random resizing, horizontal flip, transpose, vertical flip, shift scale rotate, and normalised the mean and standard deviation for research on cassava disease (Chhetri, Hohenegger, Fensel, Kasali, & Adekunle, 2023).

2.2. GAN-based Models for Generating Samples

Several researchers have developed and applied various GAN-based methods to augment and enhance plant disease datasets to achieve better accuracy. A Deep Convolutional GAN (DCGAN) was used to classify diseases in tomato leaves. The model was compared with boundary equilibrium generative adversarial networks (BEGAN) and a combination of DCGAN and BEGAN, and a considerably higher accuracy with DCGAN was achieved, demonstrating its effectiveness (Wu, et al., 2020). Another study focused on a GAN model named LeafGAN that contained a generator with degressive channels, followed by the fusion of a dense connectivity strategy, instance normalisation, and deep regret gradient penalty method (B. Liu, Tan, Li, He, & Wang, 2020) , which outperformed DCGAN and Wasserstein GAN (WGAN). A Conditional GAN (CGAN) was used to classify diseases in tomato leaves. Although the research compared the performance of the DL model with synthetic images obtained by CGAN, it could have considered more variants of GAN to further improve the accuracy and robustness of the results (Abbas, Jain, Gour, & Vankudothu, 2021). A study presented a Wide and Deep block (WDBlock)-based GAN known as a Fast WDGAN (FWDGAN) (M. Li, et al., 2022) that combined the depth feature of ResNet and extracted global features of Inception-V1 models. This method performed better than DCGAN, WDGAN, and Self-Attention GAN (SAGAN) in terms of the Fréchet Inception Distance (FID) and ssim-test. One article used a DCGAN and added it to the dataset in different ways, including the original dataset and a background-removed dataset (Huang, et al., 2022). The study also considered an augmented dataset with rotation, noise, and brightness change methods and integrated them into the original dataset and DCGAN, achieving the best results on the dataset containing background-removed images with DCGAN-generated images. Various scenarios were considered to demonstrate the effectiveness of the proposed method. A generative adversarial classified network (GACN) was proposed for plant disease classification, and its effectiveness was compared with state-of-the-art GAN-based methods, such as CGAN, Auxiliary Classifier GAN (ACGAN), multiple fake class GAN (MFC-GAN), balancing GAN (BAGAN), and ControlGAN (Wang & Cao, 2023). Another study used the DCGAN model and compared its performance with that of BEGAN and a combination of these methods; better results were attained by DCGAN only; the significance of the research was shown by comparing the results with those of other studies (Y. Chen & Wu, 2023). An article presented an improved backtracking search (IBS)-optimised GAN to classify diseases on rice leaves and addressed the instability and overfitting problems using a DCGAN (Stephen, Punitha, & Chandrasekar, 2024). A recent study adopted a combination of CNN and Swin Transformer, attained a better FID score compared to LeafGAN and CycleGAN, and named the method SugarcaneGAN (X. Li, et al., 2024). Recent research proposed a crop leaf GAN (ClGAN) with an improved loss function, an encoder-decoder network, and a dynamic correction factor. The novelty of this study was demonstrated by comparing the proposed method with prominent state-of-the-art GAN methods, including DCGAN, WGAN, Wasserstein GAN Gradient Penalty (WGANGP), Information Maximising GAN (InfoGAN), and LeafGAN (Sharma, et al., 2024).

2.3. Models use Limited Datasets

The research community has been actively involved in addressing the effects of limited dataset sizes and improving the diversity of plant disease datasets (Barbedo, 2018). A study used a transfer learning approach to address the limited dataset problem using a vision transformer (ViT), which reduced the computation time (Xu, Yoon, Jeong, & Park, 2022). Most research addressing the limited dataset problem has been resolved using few-shot learning (FSL) methods. For example, a study used ResNet-18 to generate embeddings, which were updated by a transformer, followed by an evaluation of the similarity of the transformed embeddings to the embedding of the target images (Nuthalapati & Tunga, 2021). Although the FSL-based method solved the limited dataset problem, other feature extractors could be trained to further enhance the results. Another FSL-based approach was presented to classify diseases on coffee leaves using TripleNet and ProtoNet in 5-way 1-shot and 5-way 5-shot settings on various feature extractors (Tassis & Krohling, 2022). An article presented a two-stage method based on Faster R-CNN and Siamese Network to detect diseases in strawberry leaves. In this study, various other object detection models were used to compare the performance of the developed method, such as versions of YOLO, Faster R-CNN, and an SSD model (Pan, et al., 2022). An FSL approach was presented in a study that used pretrained weights from the ImageNet dataset and refined them on the PlantCLEF2022 dataset. DenseNet was used as the baseline model, and a Support Vector Machine (SVM) classifier was used to classify plant diseases. Further experiments were performed using the PlantVillage and PDD271 datasets. The study claimed better performance than previous studies at PlantVillage and also achieved good accuracy on the PDD271 dataset (Uskaner Hepsağ, 2024). The study used other DL classifiers, such as GoogLeNet and MnasNet, could have been used, and more DL models could have been trained to further improve the proposed method. A study proposed a method called Zero-shot Transfer Learning by generating synthetic images using methods including data augmentation, normalisation, CNNs, loss functions, and fine-tuning; the discriminative information was preserved by centre-based and triplet-losses using GAN-generated images (Singh & Sanodiya, 2023). A supervised contrastive learning-based FSL method was presented in a recent study in two stages: application of a supervised contrastive learning algorithm to train an encoder with a large number of samples and application of the encoder as a feature extractor of plant disease and implementation of meta-learning training to fulfil the few-shot disease classification tasks by training a nearest-centroid classifier based on distance metrics (Mu, Feng, Yang, Zhang, & Yang, 2024). The novelty of this work was proven by comparing the method with other FSL methods with low GPU resources compared to traditional contrastive learning methods. An FSL method consisting of meta-learning, fine-tuning (PMF), and a novel feature attention (FA) module, stresses the discriminative parts in the image and reduces the impact of complicated backgrounds/undesired objects. ResNet50 and Vision Transformers (ViT) were used as feature extractors (Rezaei, et al., 2024). Another recent study proposed an FSL method based on multi-scale attention fusion with discriminative enhancement using a deep nearest-neighbour neural network (Y. Zhao, Zhang, Wu, Zhang, & Xu, 2024). A bidirectional weighted feature fusion module was created to improve the aggregation of fine-grained features, and an episodic attention module was developed for scene category-relevant attention maps. An additional spacing between category margins was added to improve the original SoftMax loss function, reduce the intra-class distances, and add L2 regularisation constraint terms to stabilise the training process. The proposed method was evaluated one 1-shot and 5-shot tasks.

2.4. Enhancement and Classification Methods

Recently, the community has proposed various methods for feature selection in plant disease classification by proposing novel filters and metaheuristic algorithms, followed by the application of well-known ML and DL models. In this regard, an optimisation-assisted cascaded filtering approach was proposed, and a Gaussian Amended Bilateral filter was designed. The proposed method removed noise from soybean and cotton plant disease datasets in two stages using the Amended Pelican optimisation algorithm and Extended Savitzky-Golay filter and compared the effectiveness of the proposed method with the filters including median, bilateral, and Gaussian filters (Bhujade, et al., 2024). To select the most prominent features of barley leaf disease, the grasshopper optimisation algorithm was used, and a backpropagation neural network was used for the classification of plant diseases (Dorgham, et al., 2024). Another recent study explored the significance of metaheuristic algorithms, including the Binary Dragonfly algorithm (BDA), ant colony optimisation algorithm (ANO), and moth flame optimisation algorithm (MFO), to optimise feature vectors for plant disease classification. The power-law transformation was used to change the intensity values of pixels, contrast-limited adaptive histogram equalisation (CLAHE) was used to improve image contrast, and LAB colour space transformation was used to improve the region of interest (Taji, et al., 2024). A feature method called support vector machine-recursive feature elimination was used to classify paddy leaf images (Dubey & Choubey, 2024). A median filter was used to remove noise, texture, and features were extracted from the green band of RGB, and classification was performed using adaptive bi-long short-term memory. A recent study used a bag-of-features to visually represent diseased leaf features and Speeded-Up Robust Features (SURF) to extract pertinent features of leaf diseases (Bhagat & Kumar, 2023). Salp Swarm Optimisation-based method was proposed in a study to classify diseases in grapes. A median filter was used to remove noise, and a Dilated Residual Network (DRN) and Adam optimiser were used to extract the distinct features of plant diseases (Alsubai, et al., 2023). A recent study used a Quantum Convolutional Neural Network (QCNN) for feature extraction of plant diseases (Anand, Jain, Mittal, & Yadav, 2025). The research optimised the extracted features using evolutionary algorithms, including the Genetic Algorithm (GA), particle swarm optimisation (PSO), constricted particle swarm optimisation (CPSO), ant colony optimisation (ANO), whale optimisation algorithm (WOA), and Modified GA. The optimised features were then applied to traditional ML classifiers for the final classification.

In summary, none of the studies emphasised performing sensitivity and correlation analyses to systematically study the impact of excluding dataset complexities on the accuracy improvement of DL-based plant disease classification. Therefore, the proposed method provides a novel way to fuse the strengths of data-centric AI with sensitivity and correlation analyses. This led to the proposal of a new performance measure called the Sensitivity Correlation Score (SC-score).

3. Proposed Data-Centric Deep Learning-based Approach

From the literature, it was noticed that previous studies did not emphasise performing detailed analyses on the effects of excluding dataset complexities on the accuracy improvement of the DL model for plant disease classification. Therefore, this study presents a novel data-centric AI-based approach for agricultural applications. Comprehensive sensitivity and correlation analyses were performed, and a novel performance score was developed based on these analyses. This study proposed the addition of two steps to general data-centric AI-based research on classifying plant diseases in a feedback loop representation, as shown in Fig. 3. The improved accuracy achieved by the proposed new steps would significantly contribute to the transfer of the research focus to data-centric AI for agricultural applications. These steps demonstrate a thorough analysis of the datasets before proposing any modifications to the DL models to improve feature extraction.

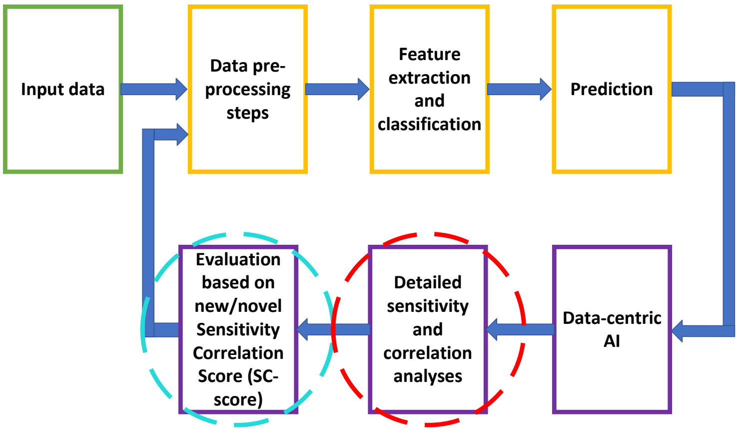


Fig. 3. The proposed addition of detailed sensitivity correlation analyses along with an evaluation based on a novel SC-score.

Because the main objective of the proposed work is to study a data-centric approach to improve plant disease classification, well-known DL models were trained and tested to select the most suitable model based on the highest overall accuracy (OA). Classes that achieved a low score (< 90%) were considered for further investigation and to study the effects of different dataset complexities. The cutoff accuracy was determined empirically. These complexities were then excluded individually and cumulatively (further details on the exclusion of these complexities are provided in the subsections). Subsequently, detailed sensitivity and correlation analyses were performed to thoroughly understand the effectiveness of eliminating dataset complexities. Finally, a novel sensitivity and correlation analysis-based score was developed to evaluate the robustness of the proposed method.

3.1. Theoretical Background

This study on plant disease classification using data-centric AI addresses the limitations of model-centric AI, which focuses on improving dataset quality rather than refining the model architecture and optimisation processes. These limitations can be highlighted mathematically (Equation (1)) using the cross-entropy loss function for a multiclass classification problem for dataset complexities such as complex backgrounds, variations in lighting, and occluded objects.

$$L = -\frac{1}{N}\sum_{i=1}^{N}\sum_{d=1}^{D} y_{id}\, log(p_{id}) \tag{1}$$

Where L-cross entropy loss function, N-total number of plant images in the dataset, D-total number of classes, where each class corresponds to a specific disease (or "healthy" if there is a healthy category), y_{id}-indicates if image i truly belongs to disease category d, and p_{id}-model's predicted probability that image i has disease d.

In the presence of dataset complexity, the predicted probability is likely to be less-accurate. The DL model may extract non-disease features from the image background or poor-quality regions instead of actual plant disease features. Hence, the DL model can assign probabilities based on this irrelevant information. This may misdirect the weights, which may disrupt the extraction of meaningful features. This also slows the training convergence. Therefore, the proposed data-centric AI addresses the limitations of model-centric AI by excluding dataset complexities. This smooths the training process and achieves better generalisation without making complex modifications to the DL models.

3.2. Novel Sensitivity Correlation Score

The proposed data-centric AI-based approach involves comparing DL models in terms of testing accuracy, identifying low-accuracy classes, identifying dataset complexities, excluding dataset complexities, retraining the best-obtained DL model, performing stratified k-fold cross validation, selecting the most-affected dataset complexity, performing comprehensive sensitivity and correlation analyses, and demonstrating the effectiveness of these analyses. Based on the sensitivity and correlation analyses, a new performance score, the Sensitivity Correlation Score (SC-score), was derived. A step-by-step explanation of the proposed methodology is presented in Fig. 4.

3.2.1. Identification of dataset complexities

Dataset complexities were identified by a thorough examination of the training dataset, which could contribute to lowering the accuracy of the models. First, some complexities may be present in a plant disease dataset containing images collected in a real agricultural environment. Dataset-specific complexities were then identified to further enhance the dataset complexities. The general complexities include variable lighting, partial shadows, reflected light, and extremely complex backgrounds, whereas dataset-specific complexities include non-focused objects of interest, different symptom shapes, low spot visibility, and occluded plant organs/disease spots. Sample images for each complexity are shown in Fig. 5(a)–(h), where the unique effect of each dataset complexity can be observed along with the BRISQUE and SI index scores. The dataset complexities are defined as follows:

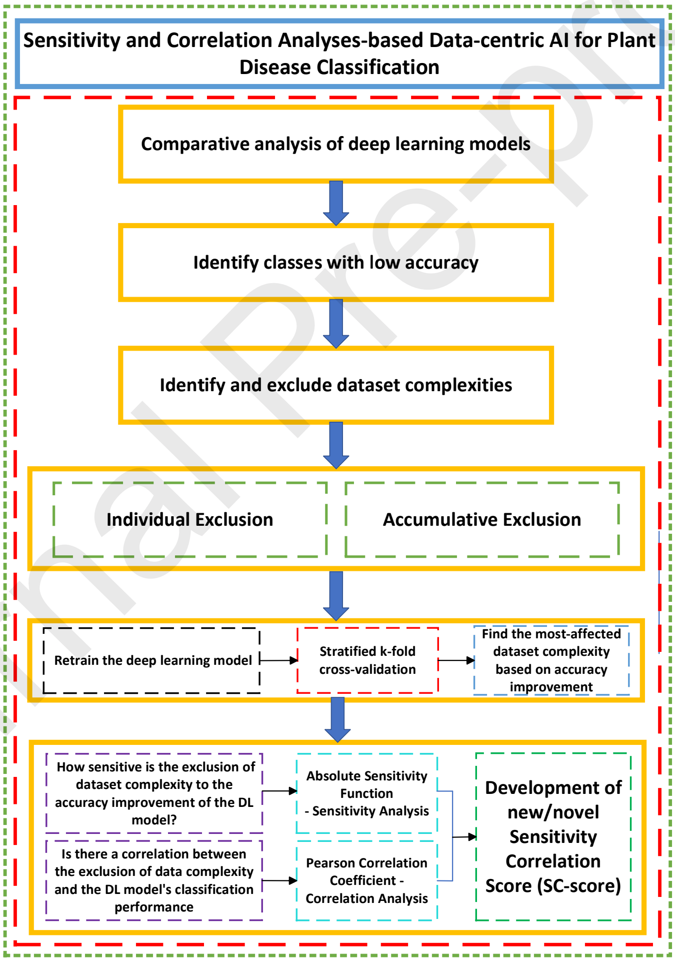


Fig. 4. Detailed representation of the proposed data-centric AI-based approach.

Variable Lighting: Images with varying lighting on the object of interest (leaf and stem) were compared with the surrounding elements. **Partial Shadow:** Images showing shadows on some parts of the leaf, stem, and/or fruit. **Reflected Light:** Images with sunlight reflections. **Extremely Complex Background:** Images with various objects in the background, rather than a few objects such as part of an object of interest, fruits on the ground, grass, etc. **Non-focused Object of Interest:** Images with the object of interest (plant disease spot) are not focused/blurred, and the background is focused. **Different**

symptom shapes: Images with unique disease symptoms, such as a large patch on the leaf instead of spots. **Spot Visibility**: Images with multiple tiny spots. **Occluded plant organs/disease spots:** Images showing obstructions due to other parts of the plant.

The motivation for choosing these complexities is that these adversarial effects are common in plant disease identification. These dataset complexities are identified using image quality measures such as BRISQUE (Blind/Referenceless Image Spatial Quality Evaluator) (Mittal, Moorthy, & Bovik, 2011), which measures natural scene statistics, and SI (Sharpness Index) (Blanchet & Moisan, 2012) , which quantifies image sharpness. The choice of BRISQUE and the SI index for identifying dataset complexities are summarised in Table 1. The threshold value of BRISQUE was set to above 40, suggesting significant distortions that could affect the performance of DL models, and below 1000 for the SI Index, as it indicates significant blurriness or lack of focus on the images.

Table 1. Selection of Image Quality Measures to Identify Dataset Complexities.

Dataset Complexity	**Image Quality Measure**	**Reason for Choice**
Variable Lighting	BRISQUE	Detects unnatural intensity variations caused by lighting changes.
Partial Shadow	BRISQUE	Quantifies distortions caused by shadows and contrast variations.
Reflected Light	BRISQUE	Evaluates overexposure caused by reflected light.
Complex Background	BRISQUE	Detects unnatural patterns or noise introduced by cluttered backgrounds.
Occluded Regions	BRISQUE	Identifies disruptions in natural image structure caused by occlusions.
Non-focused Object of Interest	SI	Measures blurriness or lack of focus in the object of interest.
Different Symptom Shapes	SI	Evaluates the clarity and sharpness of symptom shapes.
Spot Visibility	SI	Quantifies the sharpness and visibility of spots against the background.

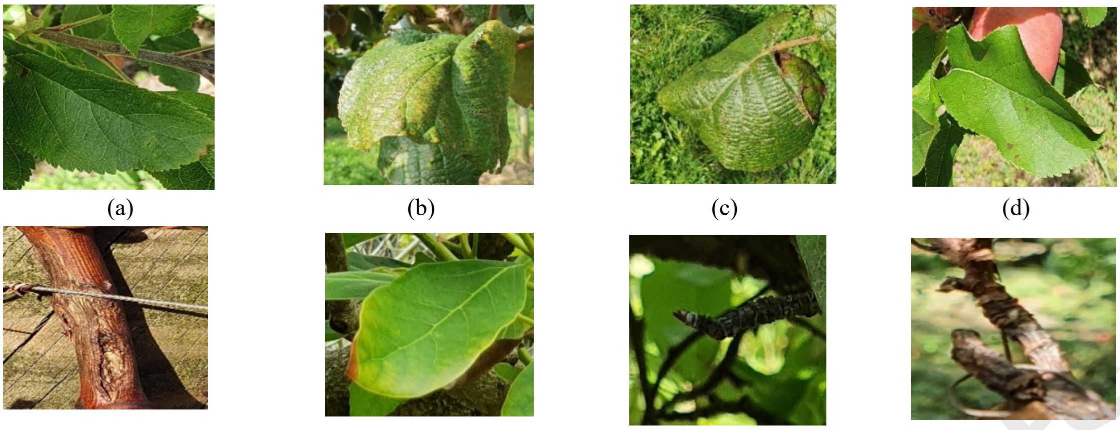


Fig. 5. Examples of identified dataset complexities. (a) Variable lighting (BRISQUE Score: 56.768), (b) Partial shadow (BRISQUE Score: 43.34), (c) Reflected light (BRISQUE Score: 54.979), (d) Extremely complex background (BRISQUE Score: 49.082), (e) Occluded plant organs (BRISQUE Score: 46.250), (f) Different symptom shape (SI Score: 917.424), (g) Spot visibility (SI Score: 497.633) and (h) Non-focused object of interest (SI Score: 457.937)

The next step was to exclude these complexities individually and cumulatively. The former approach excludes images containing the identified dataset complexities from each low-accuracy class. The best DL model was retrained on the dataset obtained after excluding the dataset complexities, and the improvement in the accuracy of the model was noted and compared. Stratified k-fold cross-validation was used to ensure unbiased testing of the DL model. Sensitivity and correlation analyses were then performed, followed by the application of the SC-score (the details of these analyses and the performance score are provided in the following subsections). In the case of performance/accuracy improvement by more than one class complexity, a new dataset was obtained based on the combined exclusion of the most affected dataset complexities. This cumulative exclusion of dataset complexities was performed in pairs of two, three, and so on. The best combination of the exclusion of these dataset complexities is obtained based on a higher class and overall accuracy than the individual complexity removal. A visual representation of the dataset exclusion is shown in Fig. 6. To perform detailed sensitivity and correlation analyses, a stepwise exclusion of dataset complexities was performed based on the number of images belonging to the complexities that should also balance the granularity and computational feasibility. The computational cost was an important factor in selecting the number of step sizes for dataset complexity exclusion, as generating more sub-datasets (due to small step size) proportionally increased the number of times the DL model needed to be trained and tested on each sub-dataset.

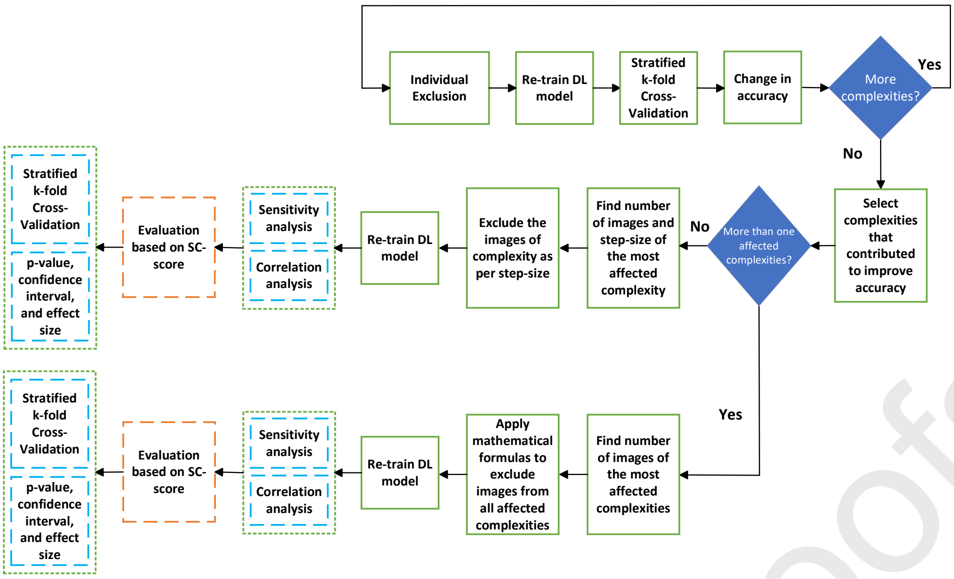


Fig. 6. Process for excluding dataset complexities accumulatively.

The selected step sizes of the dataset complexity exclusion varied from 9.09% to 25% to ensure a meaningful evaluation while avoiding excessive computational overhead. For example, dataset complexities with very few images, such as four images of spot visibility in grapevine black spots, could only exclude one image in each of the four steps, resulting in a step size of 25% (1/4). Similarly, the dataset complexity of spot visibility in kiwifruit bacterial canker had five images, and a step size of 20% (1/5) was considered, as smaller steps would require excluding fractions of images, which was impractical. In the case of a larger number of images belonging to the dataset complexity, such as partial shadow in apple Glomerella leaf spot, which had 22 images, possible options of step sizes were 1, 2, or 11 images. Smaller step sizes, such as one image (4.54%), would create 22 sub-datasets that would require retraining of the DL models 22 times, making it computationally expensive. However, larger step sizes, such as 50% (11/22), would make it difficult to detect changes or improvements in the DL model performance. Hence, a step size of 9.09% (2/22) was chosen as the most feasible option, which generated 11 sub-datasets.

In the case of accumulated complexities, the closest possible step size was considered for the combined exclusion of complexities, based on the number of images. For example, if the variable lighting for a class contains 20 images and an occluded object contains 10 images, then a step size of 2 is considered to exclude the dataset complexities. This means that ten sub-datasets were created. Interestingly, the last five sub-datasets would only have images of variable lighting, as all images related to the occluded object would have been excluded. In this case, the last five datasets would consist of the remaining variable lighting images, and again, two images of an occluded category would be considered randomly for each of the last five datasets. There are a few cases in which the two complexities have images that are not completely invisible to each other. For instance, variable lighting may have 10 images, whereas the occluded image may have only three images. Here, we need to find appropriately perform exclusion. Mathematically, the exclusion of the accumulated dataset complexities is performed using Equations (2)–(8).

$$n_j = round(s * N_j) \qquad (2)$$

$$n_{min} = min(n_j, n_k) \qquad (3)$$

$$K_j = \left\lceil \frac{N_j}{n_j} \right\rceil \quad (4)$$

$$K_k = \left\lceil \frac{N_k}{n_k} \right\rceil \quad (5)$$

$$N_{i,j} = max(0, N_{i-1,j} - n_{min}) \quad (6)$$

$$N_{i,k} = max(0, N_{i-1,k} - n_{min}) \quad (7)$$

$$K_{remaining} = \left\lceil \frac{\min(N_j, N_k)}{n_j} \right\rceil \quad (8)$$

Where N_j-total number of images for complexity j (e.g., "variable lighting"), N_k-total number of images for complexity k (e.g., "occluded object"), n_j-step size for complexity j, n_k-step size for complexity k, s-step size as a fraction (e.g., 10%, 20%), *round*-rounds the value to the nearest integer, n_{min}-combined step size for both complexities; the minimum of both n_j and n_k to ensure images are excluded from both complexities in parallel initially, K_j-number of steps to exclude all images from complexity j, and K_k-number of steps to exclude all images from complexity k, $N_{i,j}$ and $N_{i,k}$ represent the remaining images after each exclusion step, $K_{remaining}$-number of sub-datasets that are created after one complexity is fully excluded.

3.2.2. Sensitivity and Correlation Analyses

One of the main contributions of this study is the performance of sensitivity and correlation analyses. These analyses provide insights into the effects of excluding dataset complexities on the accuracy improvement of the DL model for plant disease classification.

Sensitivity Analysis: Sensitivity analysis (SA) is a method used to determine how changes in the independent variable affect the dependent variable of a model or system. In the context of this study, the step sizes of the dataset complexities were independent variables, and the accuracy change was the dependent variable. Different types of sensitivity analysis functions were applied, including absolute sensitivity analysis (ASA), relative sensitivity analysis (RSA), and semi-relative sensitivity analysis (SRSA), as defined in Equations (9)–(11), respectively.

The ASA measures the absolute change in accuracy relative to the change in step size of the dataset complexity exclusion. In contrast, the RSA scales the ASA by the ratio of the step size dataset complexity to the initial accuracy, and the SRSA scales the ASA by the dataset complexity step size. This indicates that the RSA amplifies the sensitivity values when the accuracy is low and suppresses them when the accuracy is high. Similarly, the SRSA produces higher sensitivity values at larger step sizes, irrespective of the actual step size of the dataset complexity exclusion when the highest accuracy improvement is observed. However, the ASA quantifies the impact of dataset complexity exclusion on accuracy without unnecessary scaling of the accuracy or step size of the dataset complexity exclusion.

As shown in Fig. 7, the ASA function attained its peak value at a 25% step size of dataset complexity exclusion which aligns with the step size of the highest improvement in accuracy. However, the RSA and SRSA peaked at a 100% step size which did not correspond to the optimal performance of the DL model in terms of accuracy enhancement. Hence, the evaluation of sensitivity functions indicates that the ASA provides a more reliable assessment of the effectiveness of dataset complexity exclusion for various step sizes. Therefore, ASA was selected as the primary evaluation metric because of its ability to effectively obtain the optimal step size for dataset complexity exclusion.

$$ASA = \frac{\Delta A}{\Delta Ci} \qquad (9)$$

$$RSA = \frac{\Delta A}{\Delta Ci} * \frac{C}{A} \qquad (10)$$

$$SRSA = \frac{\Delta A}{\Delta Ci} * C \qquad (11)$$

Where *ASA*-absolute sensitivity analysis, *RSA*-relative sensitivity analysis, *SRSA*-semi-relative sensitivity analysis, *A*-current accuracy of the DL model, *C*-current dataset complexity, *Ci*-the i-th step-size of dataset complexity, and Δ*A*/Δ*Ci*-ratio between the difference in accuracy with respect to the i-th step size of complexity.

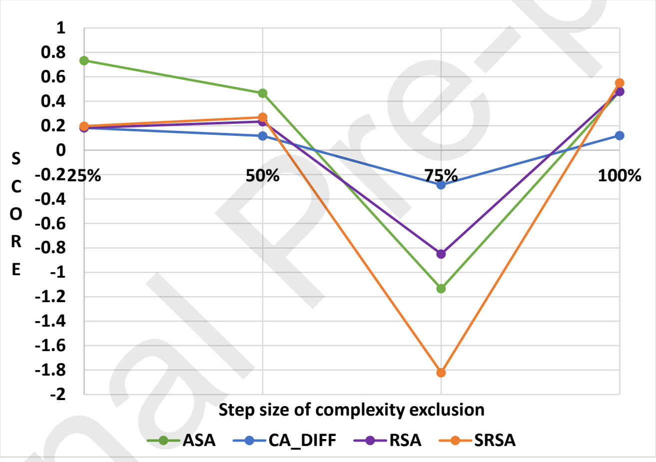


Fig. 7. Absolute (ASA), relative (RSA), and semi-relative (SRSA) sensitivity scores with changes in accuracy (CA_DIFF) after combined exclusion of variable lighting and spot visibility complexities from Apple Black Spot.

Correlation analysis: Correlation analysis is a statistical method used to determine the relationship or correlation between two variables in a dataset. For this study, these variables are the step sizes of the dataset complexities and changes in the classification accuracies of the DL model owing to complexity exclusion. The primary focus of the correlation analysis was to evaluate the strength of the correlation between dataset complexity elimination and the accuracy of the DL model. Therefore, the Pearson Correlation Coefficient (PCC) was used in this research for the following reasons: (a) to interpret how changes in the dataset in terms of excluding complexities are related to changes in accuracy across different step sizes, and (b) to determine the strength of the relationship between dataset complexity elimination and accuracy improvement of the DL model. This information is crucial for understanding how dataset complexity affects model performance. For instance, a strong positive correlation suggests that excluding dataset complexities at different step sizes has a consistent and positive effect on the accuracy of the DL model. While the PCC meets the objectives of the correlation analyses for this research by evaluating whether the correlation strength is strong or weak, it has certain limitations, including its inability to capture non-linear relationships. Other correlation functions, such as Spearman's rank correlation (SRC), assess the monotonic relationship, whether linear or non-linear.

Therefore, the significance of the SRC should be evaluated in follow-up studies which could provide additional insights into determining the nonlinear relationship between dataset complexity exclusion and performance enhancement of the DL model.

To apply PCC, the mathematical equation is defined by Equation (12):

$$PCC = \sum \frac{((X - \mu X)(Y - \mu Y))}{(\sigma X * \sigma Y)} \tag{12}$$

Where *PCC*-Pearson Correlation Coefficient between step size of complexities exclusion and difference in accuracies, *X*-dataset complexities at different percentages of step size, *Y*-corresponding changes in accuracies, μX-mean complexities step sizes, μY-mean change in accuracies, σX-standard deviation of complexities, and σY-standard deviation of accuracy changes.

3.2.3. Sensitivity Correlation Score

This study proposed a novel performance score to comprehensively understand the relationship between excluding dataset complexities and accuracy improvement for plant disease classification. An integrated score of ASA and PCC, named the Sensitivity Correlation Score (SC-score), was proposed in this study.

Criteria of the SC-score: The main criterion for developing the SC-score was to capture both the sensitivity of the DL model's performance to the exclusion of dataset complexity and the consistency of this relationship. The score should indicate that positive sensitivity is more responsive to better classification results because of the elimination of data complexity, and strong correlations (more than 0.5) indicate a more consistent relationship between complexity removal and accuracy enhancement. The formula development considered the following criteria.

- The score should reward both sensitivity and correlation; therefore, the DL model may be sensitive to changes in complexity (ASA) and show a strong correlation (PCC) between complexity and the accuracy.
- The score should guide the selection of the most affected step size for dataset complexity exclusion.
- The score should consider a positive correlation and neutralise the effect of negative correlations because a negative correlation would show an inverse relationship between the accuracy change and step sizes of the dataset complexity exclusion.
- Negative sensitivity should not be cancelled out, as it could show that the accuracy of the DL model decreased after the initial exclusion of the dataset complexities; it may increase after further refining the dataset.
- The step size with a sensitivity score of '0' should not be selected because there would be no accuracy change/improvement or either of the sensitivity was less than 0 (reduction in DL model accuracy) or correlation score was less than 0.5 (weak correlation).

Hence,

- ASA and PCC should be multiplied.
- |PCC| can be used to cancel out the effects of negative correlations and avoid penalisation of the sensitivity score.
- The SC score was normalised to make it more interpretable and consistent. It will also help ensure that both sensitivity and correlation contribute appropriately and select the most affected step size of the dataset complexity exclusion. The normalisation range of [0,1] was chosen to reward only positive correlations, as this study focuses on the strength of the accuracy improvement due to dataset complexity exclusion rather than linear or non-linear relationships between them. The normalisation process ensures that the minimum value of the SC-score is 0 which presents either both or one of the reasons among negative sensitivity and weak/moderate correlation. The maximum

value of the SC-score is 1 which indicates the highest combined effect of positive sensitivity and strong correlation scores.

Based on the above discussion/criteria, the formula for the SC-score is presented in Equations (13) and (14).

$$SC = \begin{cases} ASA * |PCC|, \ if \ ASA > 0 \ and \ |PCC| \ > 0.50 \\ 0, \ Otherwise \end{cases} \tag{13}$$

$$SC - score = \begin{cases} \frac{SC - \min(SC)}{max(SC) - \min(SC)}, \ if \max(SC) \neq \min(SC) \\ 0, \quad if \max(SC) = \min(SC) \end{cases} \tag{14}$$

Where *ASA*-absolute sensitivity analysis function, *PCC*-Pearson Correlation Coefficient, *SC*-SC-score before normalization, *SC-score*-SC-score after normalization

Significance and novelty of the SC-score: The SC-score considers the results obtained from the sensitivity and correlation analyses as a single metric. This integration is significant because it provides a more comprehensive evaluation of the model performance than either analysis alone. Moreover, this score provides a balanced assessment of how a model responds to changes in data complexity (sensitivity) and how consistently it does so (correlation). The SC-score would guide researchers to select the best-suited step size of dataset complexity exclusion, which is both sensitive and consistent in performance/accuracy improvement across different step sizes without the loss of useful information or features. Furthermore, this score neutralises the effect of negative correlations without penalising the effect of the sensitivity score.

4. Results and Discussions

This section presents the results of this research, which consist of selecting the most suitable deep learning model, followed by studying the effects of excluding dataset complexities, an ablation study, and applying sensitivity and correlation analyses along with analyses based on the SC-score, comparative evaluation with state-of-the-art methods, deployment challenges, and limitations of the study. This section provides the limitations of this study.

4.1. Dataset and Evaluation

For this study, a plant disease dataset called NDZLPlantDisease-v1 was used (Saleem, et al., 2022a) from the GitHub repository https://github.com/hsaleem1/NZDLPlantDisease-v1. This dataset was selected because it considers various aspects of a real agricultural environment. The dataset contains healthy and diseased plant leaves and other plant organs, including stems and fruits. The NZDLPlantDisease-v1 dataset comprises 20 healthy and diseased classes from five different fruits, and images were collected under diverse climatic conditions in New Zealand. The dataset was divided into training, validation, and testing sub-datasets at 70%, 20%, and 10%, respectively.

An external testing dataset containing images from random Google searches was used to evaluate the robustness of the proposed method. The external dataset was completely independent of the training and validation datasets to ensure no data leakage. The images in the external dataset were not used during any phase of model training or hyperparameter tuning which confirmed an unbiased evaluation of the generalisation capabilities of the DL models. Successful results on an external dataset will demonstrate the effectiveness of the proposed method in different agricultural environments, rather than considering only the environment used to train the DL models (Saleem, et al., 2022a). Further details of the dataset used in this study are provided in Table 2.

To ensure the reliability of the training process, stratified 5-fold cross-validation was employed in this study because of the class imbalance problem in the selected dataset. This cross-validation technique partitions the data into five folds while preserving the original class distribution of each fold. This ensured an unbiased distribution of the dataset across all the five folds. The average accuracies obtained through these folds were reported in this study for the selection of the best DL model, assessment of the effects of dataset complexity elimination, and accuracy improvement due to the exclusion of complexities in different step sizes. This approach also guarantees unbiased testing and a robust estimate of the proposed approach for the unseen data.

Table 2. Details of NZDLPlantDisease-v1 and External Testing Datasets.

Plant Species	Plant Organs	Healthy and Diseased Classes	Pathogen	Number of Training Images	Number of Validation Images	Number of Testing Images	Number of External Testing Images
Apple	Fruit	Black Rot	Fungus	39	12	6	63
		Healthy	-	132	34	17	44
	Leaf	Black Spot	Fungus	94	20	10	15
		Glomerella Leaf Spot	Fungus	190	62	31	15
		Mosaic Virus	Virus	107	40	20	37
		Healthy	-	152	34	17	15
	Stem	European Canker	Fungus	115	34	17	12
Avocado	Leaf	Algal Leaf Spot	Algae	62	16	8	11
		Branch Canker	Fungus	178	50	25	19
		Healthy	-	163	46	23	30
Grapevine	Cane	Black Spot	Fungus	56	16	8	5
		Healthy	-	115	34	17	85
Kiwifruit	Leaf	Bacterial Canker	Bacteria	264	64	32	14
		Healthy	-	186	54	27	35
Pear	Fruit	Stony Pit	Virus	37	14	7	13
		Healthy	-	155	40	20	19

	Leaf	Fire Blight	Bacteria	78	24	12	20
		Scab	Fungus	142	42	21	38
		Healthy	-	98	40	20	19
	Stem	European Canker	Fungus	69	20	10	11

DL models: After selecting the plant disease dataset, the next step was to select the most suitable DL model. In this regard, various DL classifiers were trained and tested using their pre-trained weights on the ImageNet dataset. To simplify model selection, the version of each model that attained the highest accuracy on ImageNet was retrained on NZDLPlantDisease-v1. The DL models included Xception, VGG16, ResNet152V2, InceptionResNet-V2, MobileNetV2, DenseNet201, NasNetLarge, and EfficientNetV2L (ConvNeXtXLarge and other versions of ConvNeXt could not be trained owing to the memory limitation of the GPU). Their training and validation plots were drawn to ensure that there were no problems, such as overfitting and underfitting. These DL models were compared based on the testing accuracy of NZDLPlantDisease-v1 and external testing datasets to demonstrate the robustness and practical considerations of the proposed methodology.

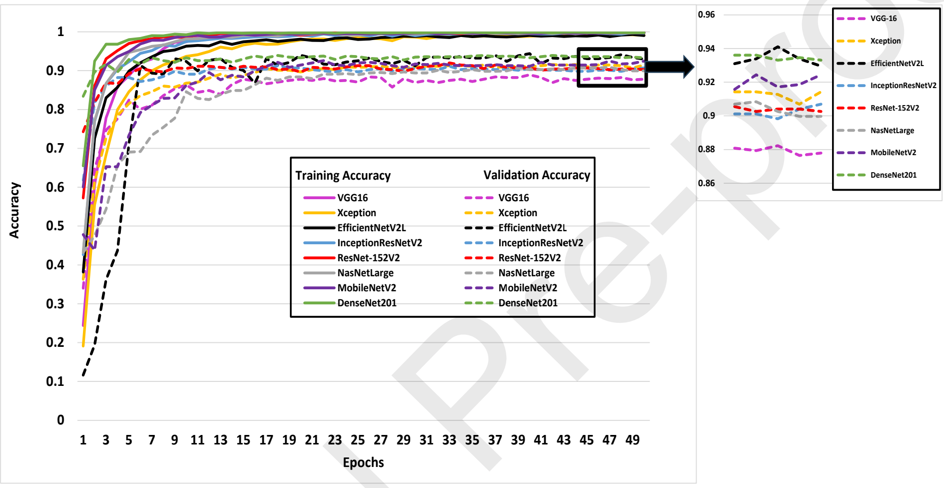
Implementation Details: All DL models were trained using the TensorFlow and Keras frameworks. The hardware included an NVIDIA RTX 4070 GPU with the following specifications: 12GB Memory, 2.48 GHz boost clock, and 5888 CUDA cores. The CuDNN library was imported to accelerate the DL model training. A Stochastic Gradient Descent (SGD) optimiser with momentum was used with the following specifications: learning rate = 0.001, momentum = 0.9, and decay = 0.0005. The hyperparameters were selected using a random search method (Bergstra & Bengio, 2012). All the DL models were trained using a fine-tuning technique with pretrained weights on the ImageNet dataset. The DL models were trained for 50 epochs with a batch size of 16 to achieve training convergence.

4.2. Experiments to Identify the Best DL Model

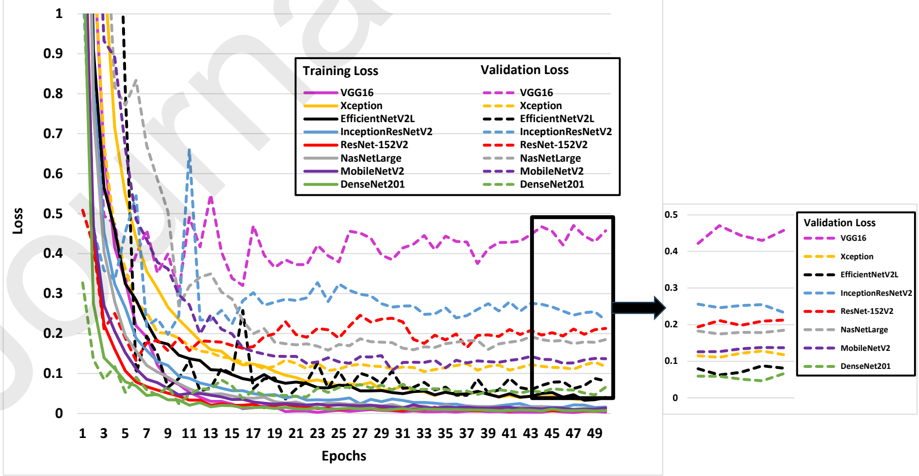
This step was performed to select the best DL model that could be applied to the proposed Data-centric AI-based approach. Eight DL classifiers were trained using the NZDLPlantDisease-v1 dataset. These DL models were tested on the NZDPlantDisease-v1 and external testing datasets. To select the best-performing DL model, various parameters and performance metrics of the models were evaluated. The most optimal choice of the DL model mainly depended on the highest accuracy of both testing datasets.

From the training performance of the DL models presented in Fig. 8 (a), it can be observed that VGG-16 requires the fewest epochs to achieve training convergence. However, MobileNetV2 required the shortest training and testing times in seconds per epoch. The major focus of this study was to improve the accuracy of plant disease classification. Therefore, training time and the number of epochs should be the focus of follow-up studies. It can also be noted that all models achieved a training accuracy of approximately 0.99. However, DenseNet-201 and EfficientNetV2L achieved the highest validation accuracies of 0.9331 and 0.9295, respectively. Similarly, the loss plots (Fig. 8 (b)) of the DL models show that the training loss varies from 0.0031 to 0.0401. However, the lowest validation loss was achieved by DenseNet-201 (0.0660), followed by EfficientNetV2L (0.0812), compared with the validation loss of other DL classifiers. The testing performance of the DL models revealed that DenseNet-201 achieved the highest accuracy, followed by the MobileNetV2 and NasNetLarge models, on the testing dataset of NZDLPlantDisease-v1. However, the highest testing accuracy on the external testing dataset was achieved by DenseNet-201, followed by NasNetLarge and Xception models, as presented in Table 3. The performance of the DenseNet-201 model was validated using the stratified k-fold cross-validation method. The dataset images of each class were folded five times in different combinations of training and validation images to ensure unbiased testing of the DenseNet-201 model. The original class distribution was retained while creating these dataset folds. The testing accuracy on the testing dataset through these folds was 0.9359, 0.9416, 0.9532, 0.9596, and 0.9651, with an average of 0.9511. Similarly, the testing accuracy of the external dataset from the cross-validation method was 0.8090, 0.8122, 0.8225, 0.8292, and 0.8329, with an average of 0.8217.

MobileNetV2 required the lowest GFLOPs and memory size, followed by DenseNet-201. From Table 3 and Fig. 8, it can be concluded that either MobileNetV2 or DenseNet-201 should be selected because of their best performance in terms of testing accuracy, testing time, or memory requirements. As this study mainly focused on improving the accuracy of the DL model using the novel data-centric AI approach, DenseNet-201 was the most optimal choice for this study because it achieved significantly higher testing accuracy on both testing datasets with a bearably higher testing time, GFLOPs, and memory requirement than MobileNetV2. Therefore, DenseNet-201 was used in the subsequent steps of this study.



(a) Accuracy plots



(b) Loss plots

Fig. 8. Training and validation plots of deep learning models.

Table 3. Performance of Deep Learning Models on NZDLPlantDisease-v1 and External Testing Datasets

DL models	Parameters (in Millions)	Number of Epochs	GFLOPS	Memory Size (MB)	Training Time	Testing Time	Training Accuracy	Validation Accuracy	Training Loss	Validation Loss	Testing Accuracy - NZDLPlantDisease-v1	Testing Accuracy – External Testing dataset
VGG16	138.4M	20	30.95	512.53	34	**0.003**	0.9991	0.8779	0.0031	0.4571	0.8818	0.5863
Xception	22.9M	41	9.11	87.91	40	**0.003**	0.9917	0.9142	0.0387	0.1178	0.9116	0.7419
EfficientNetV2L	119.0M	42	24.54	455.78	181	0.019	0.9905	0.9295	0.0401	0.0812	0.9221	0.7380
Inception ResNetV2	55.9M	22	12.95	214.50	43	0.004	0.9967	0.9069	0.0161	0.2334	0.9226	0.5979
ResNet152V2	60.4M	39	21.83	231.45	54	0.006	0.9983	0.9026	0.0080	0.2126	0.9259	0.7289
NASNetLarge	88.9M	35	47.68	341.40	121	0.011	0.9979	0.8997	0.0102	0.1847	0.9312	0.7434
MobileNetV2	3.5M	25	0.6	13.98	15	**0.001**	0.9975	0.9244	0.0135	0.1369	0.9316	0.6398
DenseNet201	20.2M	21	8.59	78.55	42	**0.004**	0.9979	0.9331	0.0075	0.0660	**0.9511**	**0.8217**

4.3. Experiments of Individual and Accumulative Exclusion of Dataset Complexities

In the previous step, DenseNet-201 was selected as the most suitable DL model for this study. As shown in Table 3, the testing accuracy of the DenseNet-201 model was significantly higher for the NZDLPlantDisease-v1 testing dataset (0.9511) than for the external testing dataset (0.8217). The testing accuracy of all individual healthy and diseased classes was noted to determine the classes that attained an accuracy lower than 90% for further analysis. As shown in Table 4, the accuracy of each class was approximately 90% in the testing dataset of NZDLPlantDisease-v1. However, some classes in the external testing dataset did not achieve high accuracy (less than 90%), including apple black spot (A_blk_spot), apple glomeralla leaf spot (A_gl_lf_spot), apple healthy leaves (A_healthy_l), avocado algal leaf spot (Av_alg_lf_spot), avocado bacterial canker (Av_br_canker), grapevine black spot on cane (G_blk_spot_c), kiwifruit bacterial canker (Kf_bac_canker), and pear canker (P_canker). The accuracies ranged from 40% to 78.94%. However, the remaining classes attained high accuracy (89.47 %–100 %).

DenseNet-201 was retrained, followed by the exclusion of individual dataset complexities. Their improvement was observed based on the difference in default accuracy (without any change to the training dataset) and updated accuracy (after excluding the dataset complexities), as shown in Table 5. In the case of accuracy improvement owing to the removal of more than one dataset complexity, an accumulated exclusion of dataset complexities was considered. Because the accuracy of the testing dataset of NZDLPlantDisease-v1 did not significantly increase, the remaining analyses were based on the external testing dataset.

Table 4. Class-wise accuracy of DenseNet-201 models on NZDLPlantDisease-v1 and External testing datasets.

Healthy and disease classes	**Accuracy on NZDLPlantDisease-v1**	**Accuracy on the External Testing Dataset**
Apple black rot (A_blk_rot_f)	1	0.8964
Apply black spot (A_blk_spot)	0.9	**0.75**
Apply European Canker (A_e_canker)	0.9011	0.9133
Apple Glomerella Leaf Spot (A_gl_lf_spot)	0.8967	**0.5909**
Apply healthy fruit (A_healthy_f)	0.9111	0.9233
Apply healthy leaves (A_healthy_l)	0.8823	**0.6**
Apply mosaic virus (A_m_virus)	0.9	0.9108
Avocado Algal Leaf spot (Av_alg_lf_spot)	1	**0.7272**
Avocado branch canker (Av_br_canker)	1	**0.7894**
Avocado healthy leaf (Av_healthy_l)	0.8965	0.9
Grapevine black spot (G_blk_spot_c)	1	**0.4**
Grapevine healthy cane (G_healthy_c)	0.9011	1
Kiwifruit bacterial canker (Kf_bac_canker)	0.9287	**0.7142**
Kiwifruit healthy leaf (Kf_healthy_l)	1	0.9
Pear canker (P_canker)	1	**0.7690**
Pear fire blight (P_fr_blight)	1	0.93
Pear healthy fruit (P_healthy_f)	1	1
Pear healthy leaf (P_healthy_l)	1	0.8947

Pear stony pit (P_s_pit)	1	0.898
Pear scab (P_scab)	0.9047	0.9274

Table 5. Change in accuracy due to individual and cumulative complexity exclusion from low-performing classes.

Classes	Individual and Combined Dataset Complexities	Accuracy on NZDLPlantDisease-v1	Accuracy on External Testing dataset	Change in accuracy on NZDLPlantDisease-v1	Change in accuracy on External Testing Dataset
A_blk_spot	Variable lighting (Var_light)	0.9	0.8	0	0.05
	Spot visibility (Spot_visibility)	0.9	0.8667	0	0.1167
	Different symptom shape (Symp_shape)	0.9	0.8667	0	0.1167
	Extremely Complex background (Comp_back)	0.9	0.7667	0	0.0167
	Var_light_Spot_visibility	0.9	0.8667	0	0.1167
	Var_light_Symp_shape	0.9	0.47	0	-0.28
	Spot_visibility_Symp_shape	0.9	0.6	0	-0.15
	Var_light_Spot_visibility_Symp_shape	0.9	0.47	0	-0.28
	Var_light_Comp_back	0.9	0.6667	0	-0.0833
	Spot_visibility_Comp_back	0.9	0.7333	0	-0.0167
	Symp_shape_Comp_back	0.9	0.6667	0	-0.0833
	Var_light_Spot_visibility_Comp_back	0.9	0.7333	0	-0.0167
	Var_light_Symptom_shape_Comp_back	0.9	0.6667	0	-0.0833
	Spot_visibility_Symptom_shape_Comp_back	0.9	0.7333	0	-0.0167

	Var_light_Spot_visibility_Symptom_shape_Comp_back	0.9	0.5333	0	-0.2167
A_gl_lf_spot	Partial Shadow (Par_shadow)	0.935	1	0.0383	0.4091
	Spot Visibility (Spot_visibility)	0.935	0.5909	0.0383	0
	Variable Lighting (Var_lighting)	1	0.5909	0.1033	0
	Different Symptom Shape (Symp_shape)	0.903	0.9333	0.0063	0.3424
	Non-focused Object of Interest (OOI)	0.9667	0.9333	0.07	0.3424
	Reflected Light (Ref_light)	1	1	0.1033	0.4091
	Occluded Plant Organ/Disease Spots (Occ)	0.935	0.5909	0.0383	0
	Par_Shadow_Symp_Shape	0.935	0.8667	0.0383	0.2758
	Par_Shadow_OOI	0.9667	0.8667	0.07	0.2758
	Par_Shadow_Ref_Light	0.9667	0.8667	0.07	0.2758
	Symp_Shape_OOI	0.9667	0.9333	0.07	0.3424
	Symp_Shape_Ref_Light	0.9667	0.9333	0.07	0.3424
	OOI_Ref_Light	0.9667	0.8667	0.07	0.2758
	Par_Shadow_Symp_Shape_OOI	0.903	0.9333	0.0063	0.3424
	Par_Shadow_Symp_Shape_Ref_Light	0.935	0.8667	0.0383	0.2758
	Par_Shadow_OOI_Ref_Light	0.9667	0.9333	0.07	0.3424
	Symp_Shape_OOI_Ref_Light	1	0.8	0.1033	0.2091
	Par_Symptom_OOI_Ref_Light	0.935	0.8	0.0383	0.2091
A_healthy_l	Variable Lighting (Var_light)	0.7642	0.4667	-0.1181	-0.1333

	Occluded Plant Organ/Disease Spots (Occ)	0.8823	0.6667	0	0.0667
	Spot Visibility (Spot_visibility)	0.7642	0.48	-0.1181	-0.12
	Different Symptom Shape (Symp_shape)	0.7642	0.7333	-0.1181	0.1333
	Extremely Complex Background (Comp_back)	0.8823	0.4667	0	-0.1333
	Occ_Symp_Shape	0.5882	0.4667	-0.2941	-0.1333
Av_alg_lf_spot	Spot Visibility (Spot_visibility)	1	0.8182	0	0.091
	Different Symptom Shape (Symp_shape)	1	0.909	0	0.1818
	Non-focused Object of Interest (OOI)	1	0.7272	0	0
	Spot_Visiblity_Symp_Shape	1	0.636	0	-0.0912
Av_br_canker	Different Symptom Shape (Symp_shape)	1	0.8182	0	0.0288
	Variable Lighting (Var_light)	1	0.6311	0	-0.1583
	Spot Visibility (Spot_visibility)	1	0.7362	0	-0.0532
	Occluded Plant Organ/Disease Spots (Occ)	1	0.8182	0	0.0288
	Non-focused Object of Interest (OOI)	1	0.909	0	0.1196
	Symp_Shape_Occ	1	0.578	0	-0.2114
	Symp_Shape_OOI	1	0.526	0	-0.2634
	Occ_OOI	1	0.736	0	-0.0534
	Symp_Shape_Occ_OOI	1	0.631	0	-0.1584
G_blk_spot_c	Variable Lighting (Var_light)	1	0.4	0	0
	Non-focused Object of Interest (OOI)	1	0.4	0	0

	Occluded Plant Organ/Disease Spots (Occ)	1	0.4	0	0
	Spot Visibility (Spot_visibility)	1	0.6	0	0.2
Kf_bac_canker	Different Symptom Shape (Symp_shape)	1	0.5	0.0713	-0.2142
	Partial Shadow (Par_shadow)	1	0.7421	0.0713	0.0279
	Reflected Light (Ref_light)	1	0.5711	0.0713	-0.1431
	Variable Lighting (Var_light)	0.968	0.7142	0.0393	0
	Spot Visibility (Spot_visiblity)	0.968	0.925	0.0393	0.2108
	Par_Shadow_Spot_Visibility	1	0.571	0.0713	-0.1432
P_canker	Variable Lighting (Var_light)	1	0.4542	0	-0.3148
	Spot Visibility (Spot_visibility)	1	0.3632	0	-0.4058
	Extremely Complex Background (Comp_back)	1	0.6366	0	-0.1324
	Occluded Plant Organ/Disease Spots (Occ)	1	0.7894	0	0.0204
	Non-focused Object of Interest (OOI)	1	0.918	0	0.149
	Different Symptom Shape (Symp_shape)	1	0.7272	0	-0.0418
	Occ_OOI	1	0.5454	0	-0.2236

For instance, complexities, including variable lighting, spot visibility, different symptom shapes, and extremely complex backgrounds, were identified for apple black spots, as shown in Table 5, based on the definitions of dataset complexities provided in Section 3. The exclusion of variable lighting from A_blk_spot improved the accuracy by 5%, whereas the performance of the DL model in the absence of spot visibility and different symptom shapes improved by 11.67%. However, the elimination of the complex background improved the accuracy of A_blk_spot by a marginal margin of 1.67%. As all the identified dataset complexities contributed to improving the accuracy of A_blk_spot, all four identified complexities must be combined in all possible ways. The combined elimination of variable lighting and spot visibility from the A_blk_spot class attained the highest overall accuracy of 86.3% (which was the average of the accuracies obtained by five folds of the stratified k-fold cross-validation method with 87.38%, 86.84%, 86.43%, 85.55%, and 85.30%), which is an improvement of 4.13% compared with the overall accuracy on the default dataset. Another interesting finding was that the class accuracy (CA) of A_blk_spot was 0.8667, which was the same as that achieved by excluding spot visibility and different symptom shapes individually. The overall accuracy was higher with the combined exclusion of variable lighting and spot visibility.

Another interesting finding of this step is the complexity exclusion of the A_gl_lf_spot class. Four dataset complexities achieved better accuracy than the default dataset: partial shadow, different symptom shapes, non-focused objects of interest, and reflected light. The individual elimination of partial shadows and reflected light resulted in an accuracy of 1.00. The combined elimination of different symptom shapes and non-focused objects of interest, different symptom shapes and reflected light, partial shadows along with different symptom shapes and non-focused objects of interest, and partial shadows along with non-focused objects of interest and reflected light also attained a better accuracy of 93.33%. To select the best individual or combined exclusion of dataset complexities, partial shadow was selected because of its higher overall accuracy of 0.8370 (which was the average of the accuracies obtained by five folds of the stratified k-fold cross-validation method with 0.8430, 0.8400, 0.8381, 0.8340, and 0.8298) compared to other exclusions.

For A_healthy_l, only the exclusion of occluded and different symptom shapes improved the class accuracy, as shown in Table 5. However, their combined elimination degraded the accuracy to 0.4667. Different symptom shapes were considered for the next steps of the analysis owing to their highest class and overall accuracy improvement. Similarly, for Av_alg_lf_spot, the removal of spot visibility and different symptom shapes contributed to attaining high accuracies of 81.82% and 90.9%, respectively. However, their combined elimination reduces class accuracy. The different symptom shapes improved the accuracy with the largest margin; hence, they were considered for the next steps.

For Av_br_canker, the elimination of non-focused objects of interest improved the accuracy of DenseNet-121 by 11.96%. The exclusion of spot visibility for G_blk_spot_c attained a better result than all recognised dataset complexities. The exclusion of partial shadows and spot visibility improved Kf_bac_canker classification. However, their combined elimination did not contribute to further improvements in accuracy. Hence, the elimination of spot visibility was considered for the rest of the analysis for Kf_bac_canker because of the improvement in accuracy with the largest difference. For P_canker, the non-focused object of interest attained class accuracy improvement, and a small enhancement was observed when the occluded images were removed. However, other complexities degrade the accuracy of the P_canker. Moreover, the combined elimination of non-focused objects of interest and occlusion also downgraded the performance of the class accuracy by a large margin (22.36 %).

Based on the above discussion, the following observations were made for the next steps of the research:

- The exclusion of some dataset complexities improves the classification accuracy.
- The elimination of some dataset complexities degrades the performance of the DL model.
- Although individual and accumulated/combined elimination of dataset complexities were analysed, individual complexities were found to be the most affected for most of the examined classes.
- Only A_blk_spot attained/retained a higher accuracy after eliminating the complexities of the combined dataset.

The above experimental results show that an increase in the class and overall accuracies improved the feature extraction capability of the DenseNet-121 model owing to the removal of certain dataset complexities, thereby proving the significance of the proposed data-centric AI approach. However, performance degradation was also observed after the exclusion of some of the other dataset complexities, which implies the removal of some important or distinct features of healthy or diseased plant organs from the training dataset. Therefore, further investigation is required to confirm the robustness and validity of the proposed data-centric AI approach. Further examination would also require unfolding the patterns and relationships of the exclusion of dataset complexities to improve the performance of the DL model. Therefore, comprehensive sensitivity and correlation analyses were performed to extract these patterns from the data. Moreover, a new performance score, the SC-score, was developed to integrate the impact and correlation of accuracy improvement with step sizes/instances of the exclusion of the dataset complexities.

4.4. Experiments for Sensitivity and Correlation Analyses

To analyse the impact of dataset complexity on the accuracy improvement of the DL model, this study applied the absolute sensitivity analysis function (SA) by taking the ratio of the difference between class accuracies as the model output and the step sizes of dataset complexity exclusions as the model input. Pearson correlation coefficient was used to perform correlation analyses. Finally, combined sensitivity and correlation analyses were performed based on the newly developed SC-score. To highlight the effectiveness of the proposed SC-score, comprehensive sensitivity, correlation, and SC-score-based analyses were performed to identify the most effective step size of the dataset complexity exclusion and the strength of the association between the accuracy improvement of the DL model and step sizes.

4.4.1. Detailed Sensitivity Analyses: The absolute sensitivity function was applied to the most affected class complexities (from the previous step) for different step sizes. From Fig. 9 (a)–(h), it can be observed that the complexity step sizes ranged from 9.09% to 25% based on the number of images in the dataset. Owing to the negative change/degradation in the accuracy for different complexity step sizes, the sensitivity scores also decreased; however, they increased with the positive change/improvement in the accuracy. There were a few instances in which the class accuracy improved, but the overall accuracy decreased. For instance, the sensitivity score was positive for G_blk_spot after excluding spot visibility for class accuracy; however, the overall accuracy was reduced by 25%, 50%, and 100%, respectively. This implies that DenseNet-201 improved the accuracy of the G_blk_spot class, and the overall discrimination and classification between the classes was reduced. However, at a size step of 75%, both CA and OA improved.

Some important observations can be made for different symptom shape complexities for Av_alg_lf_spot. The best results in terms of the highest sensitivity score were achieved with the exclusion of 30% and 100% of the data for class accuracy. However, the sensitivity score for overall accuracy was highest at 30% and 70% complexity exclusion rather than at full elimination. Another important point is noted for Kf_bac_canker, where there is a slight difference in CA for the 20% and 100% complexity exclusions; the most impactful step size for OA is 100% exclusion for better OA; therefore, complete elimination of the identified complexity can be considered for OA.

As discussed in the previous subsection, A_blk_spot was the only class with a combination of complexities that achieved better accuracy. Therefore, the effects of different complexity step sizes were studied for both variable lighting and visibility. Subsequently, the most affected complexity was used to apply sensitivity analyses at different step sizes of the combined/accumulated complexity exclusion. This approach attained a significantly high overall accuracy of 93.11% on an external testing dataset.

While analysing the complexity of the non-focused object of interest for P_canker, the complete elimination (100%) of the complexity was found to be the most effective method for improving the CA and OA. For A_healthy_l, the step sizes of 28.56%, 71.40%, and 99.96% showed improvements in both CA and OA. The 5th step size (71.4%) was found to be the most affected in eliminating different symptom shapes from A_healthy_l owing to the highest sensitivity scores for CA and OA. The full elimination of the non-focused object of interest for Av_br_canker was found to be the most affected step size. The partial shadow for A_gl_lf_spot had a positive effect on all step sizes for better CA and was positive for the full removal of the complexity for better OA.

4.4.2. Detailed Correlation Analyses: Correlation analysis was performed to understand the relationship between the accuracy changes of the class and the overall accuracy with respect to the exclusion of dataset complexities at different step sizes. From Fig. 9 (a)–(h), it can be seen that Av_br_canker has the strongest correlation between the complexity step sizes and class accuracy change owing to the exclusion of the images belonging to the non-focused object of interest. However, the strongest positive correlation was observed for the overall accuracy when spot visibility was removed from G_blk_spot. P_canker also has a strong and almost same correlation in terms of CA and OA.

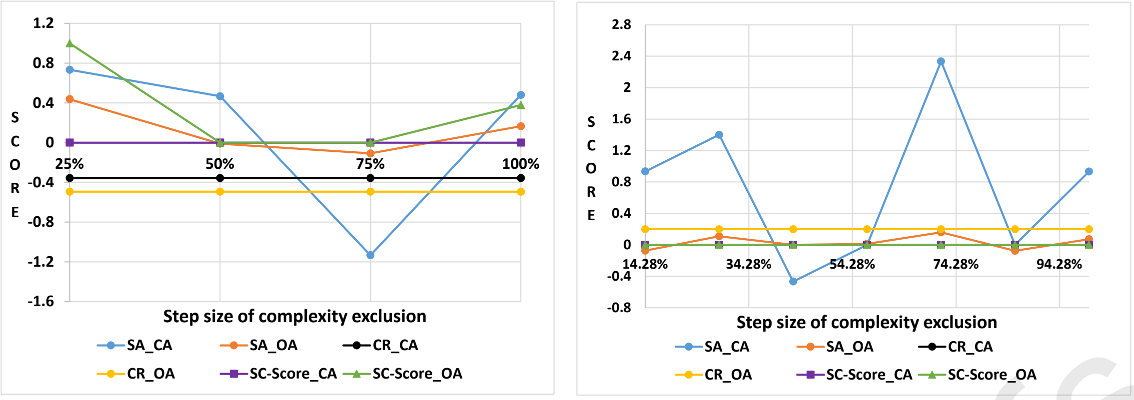
The complexities associated with classes such as G_blk_spot, Kf_bac_canker, and A_gl_lf_spot showed a moderate positive correlation (10-50%) with better CA. Av_alg_lf_spot had a weak positive correlation (0-10), A_healthy_l had almost no correlation, and A_blk_spot showed a moderate negative correlation for CA. In terms of the correlation of OA with the dataset complexity step sizes, Av_br_canker, Kf_bac_canker, and A_gl_lf_spot had a strong positive correlation (more than 50%). The classes Av_alg_lf_spot and A_healthy_l have moderate correlations. Similar to the correlation for CA, A_blk_spot also had a strong negative correlation with OA.

4.4.3. Detailed SC-Score-based Analyses: From the sensitivity and correlation analyses, an exact indication of the most sensitive and correlated step size was not obtained, and a combined score was required. The absolute sensitivity and correlation scores were calculated according to the proposed SC-score presented in Section 3.2.3, and the SC-score was applied. The results presented in Fig. 9 (a)–(h) show the sensitivity, correlation, and SC-scores for the class and overall accuracies of all selected healthy and diseased classes with their most affected dataset complexity. Based on the experimental results, the following conclusions were drawn:

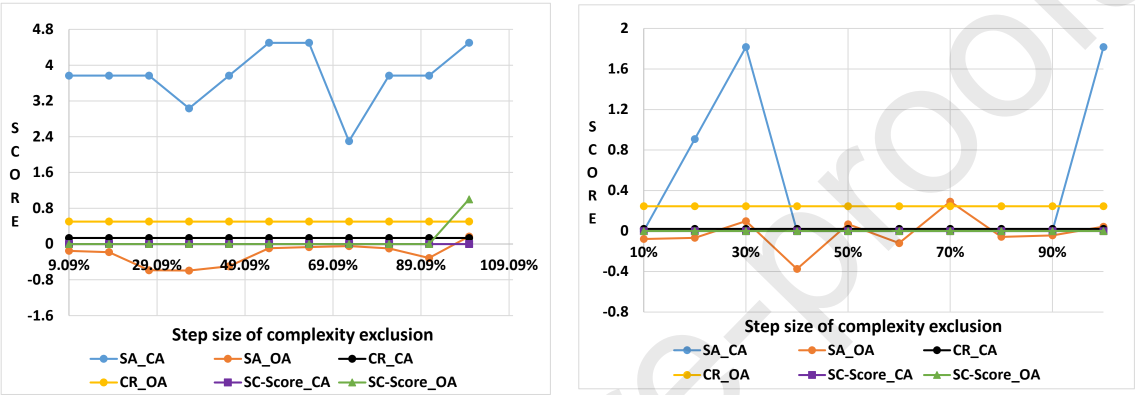
- To address the sensitivity and correlation of different portions/sections of the complexities on the accuracy of the DL model, different instances of the complexities are analysed. For instance, when the non-focused object of interest (OOI) was excluded from P_canker, an SC-score of 1.0 was obtained at a 100% step size for the improved CA owing to high values of sensitivity and correlation scores. Therefore, all training images belonging to the OOI for P_canker were removed from the dataset to improve the class accuracy from 0.769 to 0.918. Similarly, an SC score of 1.0 was attained with the full elimination of OOI for improved OA.
- Similarly, for Av_br_canker, the most affected step size of the exclusion of OOI was 100% to improve the class accuracy from 0.7894 to 0.909. Similarly, improved overall accuracy was observed at a 100% step size of dataset complexity exclusion, and an SC-score of 1 was obtained owing to positive sensitivity and high correlation scores. Hence, all images belonging to dataset complexity should be removed from Av_br_canker.
- For G_blk_spot_c, 3/4th (75%) of the spot visibility could have been removed to classify all testing images as the OA was improved with a small difference of 0.0064. The sensitivity score was small but greater than zero, and the correlation score was high. Therefore, this elimination should be considered at a 75% step size of the dataset complexity exclusion.
- Similarly, for Kf_bac_cnaker and A_gl_lf_spot, a positive sensitivity and strong correlation were observed for improved OA of 83.4% and 83.7%, respectively, at a 100% step size of dataset complexity elimination. Therefore, the removal of dataset complexities should be considered.
- For the class including Av_alg_lf_spot and A_healthy_l, the sensitive step sizes of the complexities are 30%, 100%, and 71.4%, respectively, for class accuracies, the correlation scores are weak. Hence, the SC-scores are 0 for these classes which suggests avoiding dataset complexity exclusion from these classes. Similarly, for overall accuracies, the sensitivity scores were positive at 70% and 71.40% step sizes, and the correlation was moderate; therefore, the SC-score was 0, and dataset complexity elimination should not be considered.
- The SC-score also answered how to determine which complexity steps are the most correlated and sensitive to better accuracy in the case of the combined exclusion of complexities. For A_blk_spot, when variable lighting and spot visibility were excluded, the most sensitive and correlated step size was 25. The SC scores at 25% and 100% step sizes were 1.0 and 0.3775, respectively, for improved OA. Hence, the dataset complexities should be removed at a 25% step size, as both sensitivity and correlation scores attained high scores when OA was considered for evaluating the SC-score.
- The overall accuracy (OA) improved from 82.17% to 93.11% (an average of accuracies in five folds of training-validation datasets as per the stratified k-fold cross-validation method as 93.58%, 93.32%, 92.83%, 92.64%, and 93.18%) after removing 25% of images belonging to variable lighting and spot visibility from A_blk_spot. To validate that this improvement was not due to random variation, various statistical measures were used based on the results of the stratified 5-fold cross-validation. The mean improvement in overall accuracy across the five folds was 10.94%, with individual improvements of 11.41%, 11.15%, 10.66%, 10.47%, and 11.01%. The improvement was

statistically significant ($p < 0.001$), with a 95% confidence interval of 10.64% to 11.24% in terms of the accuracy improvement compared with the initial accuracy of 82.17%. Additionally, the effect size (Cohen's d) for the accuracy improvement was 32.42 which indicates a very large effect. These results confirm that the improvement in accuracy is meaningful and not due to random variations.

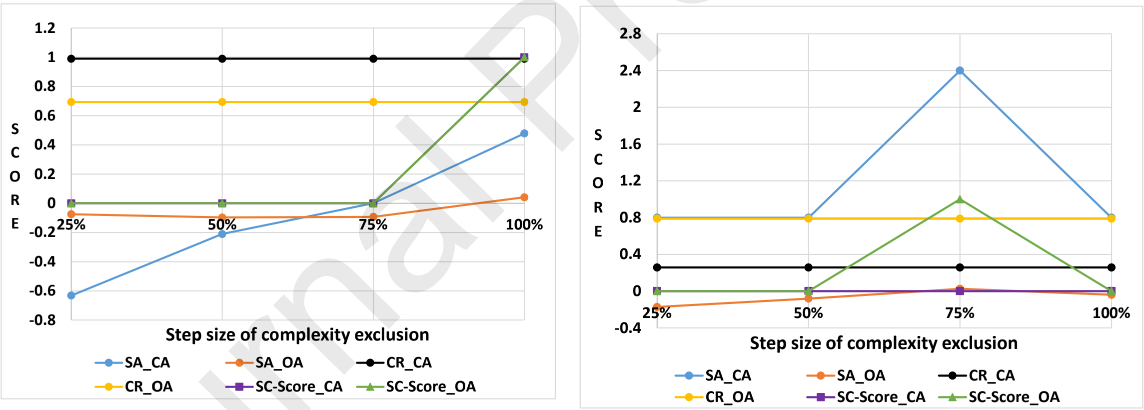
- To determine the implications of the SC-score, it captures both the responsiveness to changes in data complexity (sensitivity) and the consistency of this relationship (correlation) in a single metric. Therefore, the selection of the most affected step size of complexity exclusion considers not only the extent to which they are sensitive to changes in complexity, but also a consistent relationship between complexity removal and accuracy enhancement.



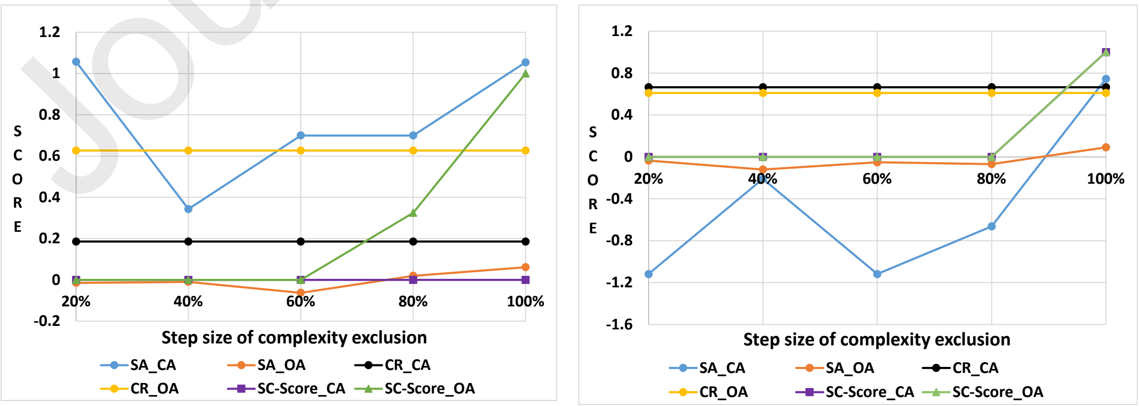
(a) Analysis for Apple Black Spot (b) Analysis for Apple Healthy Leaves



(c) Analysis for Apple Glomeralla Leaf Spot (d) Analysis for Avocado Algal Leaf Spot



(e) Analysis for Avocado Branch Canker (f) Analysis for Grapevine Black Spot



(g) Analysis for Kiwifruit Bacterial Canker (h) Analysis for Pear Canker

Fig. 9. Absolute Sensitivity (SA_CA; SA_OA), correlation (CR_CA; CR_CA), and SC-Score (SC-Score_CA; SC-Score_CA) plots for the eight selected healthy and disease classes for class accuracy (CA) and overall accuracy (OA).

4.5. Comparative Analysis with State-of-the-art Methods

The effectiveness of the proposed data-centric AI approach was evaluated by comparing its performance with the latest DL approaches (Table 6). In this regard, a recently proposed DL model named ResdenseNet was selected which was developed using the MobileNet, DenseNet, and ResNet architectures for plant disease classification problems (Nagpal & Goel, 2025). Another recent study was considered for the comparative evaluation of the proposed approach, which was based on a metaheuristic algorithm for the optimal feature selection of plant diseases (Taji, et al., 2024). This approach was based on methods including the Binary Dragonfly algorithm (BDA), ResNet-18, Shannon entropy, and F-tree classifier.

The results presented in Table 6 show that the testing accuracy of the proposed data-centric AI approach attained a significantly higher accuracy of 98.51% on a testing dataset of NZDLPlantDisease-v1 and 93.11% on an external testing dataset than all other state-of-the-art approaches. Although model-centric AI (ResDenseNet) is a computationally inexpensive solution in terms of GLOPS, memory requirements, and training/testing time, the accuracy achieved by this method was considerably lower than that of the proposed approach. This demonstrates the effectiveness of the proposed data-centric AI approach compared with the latest model-centric AI approach in terms of high testing accuracy. The proposed approach was compared with another recent state-of-the-art approach, which focused on feature selection using a metaheuristic algorithm. The testing accuracy attained by the meta-heuristic-based method was considerably good (91.10%) on the testing dataset of the NZDLPlantDisease-v1 dataset, and it attained a low accuracy of 72.86% on an external testing dataset. Comparative evaluations with the latest state-of-the-art methods demonstrated the effectiveness of the proposed data-centric AI approach.

Table 6. Comparison of the proposed data-centric AI approach with the state-of-the-art methods

DL models	**Parameters (in Millions)**	**Number of Epochs**	**GFLOPS**	**Memory Size (MB)**	**Training Time**	**Testing Time**	**Training Accuracy**	**Validation Accuracy**	**Training Loss**	**Validation Loss**	**Testing Accuracy-NZDLPlantDisease-v1**	**Testing Accuracy – External Testing dataset**
DenseNet-201 (best DL Model)	20.2M	21	8.59	78.55	42	0.0044	0.9979	0.9331	0.0075	0.0660	0.9511	0.8217
ResDenseNet (Nagpal & Goel, 2025)	0.72M	42	2.34	3.63	12	0.0015	0.9424	0.9031	0.1772	0.2981	0.8456	0.6461
Meta-heuristic algorithm (Taji, et al., 2024)	11.7M	24	7.73	98.37	21	0.0031	0.9979	0.9067	0.0084	0.3701	0.9110	0.7286
Proposed Data-Centric AI Approach	20.2M	21	8.59	78.55	42	0.0044	0.9985	0.9582	0.0036	0.0432	**0.9851**	**0.9311**

4.6. Ablation Study

The ablation study shows that the exclusion of dataset complexities based on correlation analysis-only and sensitivity analysis-only was not sufficient to select the most impactful step size of the dataset complexity exclusion.

Analysis of dataset complexity exclusion based on sensitivity function only: In some cases, the impact of dataset complexity exclusion on the accuracy enhancement of the DL model was the same for the sensitivity analysis as that of the correlation analysis. For example, the ASA value due to the exclusion of different symptom shapes from Av_alg_lf_spot at 70% step size was 0.291, and the PCC value was 0.2451 for overall accuracy, which shows the similar behaviour of the sensitivity and correlation analyses. However, some cases of exclusion of dataset complexities provided a high score for the sensitivity function, and the correlation showed a weak strength. For example, the value of the ASA function was 2.3340 at a 71.40% step size of exclusion of different symptom shapes from the apple healthy leaves class, and the PCC was -0.0005 for class accuracy. Therefore, if sensitivity analysis was only considered to select the most optimal step size of the dataset complexity elimination, 71.40% of different symptom shapes would have been eliminated which does not show consistency in the performance improvement of the class.

Analysis of dataset complexity exclusion based on correlation function-only: A correlation analysis was performed to evaluate the strength of the accuracy improvement of the DL model based on the step sizes of dataset complexity exclusion. For instance, a strong correlation was observed when the non-focused object of interest was excluded from the pear canker class which also validates the result obtained by the sensitivity function owing to its high value at a 100% step size for better class accuracy, as shown in Fig. 9. However, there were few cases in which correlation analyses provided a strong correlation in terms of a high value of the correlation function, indicating a significant association between the change in accuracy of the DL model and the elimination of dataset complexity. In contrast, the sensitivity function attained a low (negative) score at some instances of dataset complexity exclusion. For instance, the value of PCC after excluding the complexity of spot visibility from Kiwifruit Bacterial Canker (Kf_bac_canker) was 0.6268, but the ASA value was -0.0145 (Fig. 9) on a step size of 20% with a difference of -0.0029 in overall accuracy. Hence, if the analysis of the dataset complexity exclusion was based solely on the correlation analysis, then the elimination of the spot visibility from the Kf_bac_canker would have been selected at a 20% step size, depicting an incorrect selection of the most impactful step size of dataset complexity exclusion due to performance degradation at that step size.

Analysis of dataset complexity exclusion based on SC-Score: The SC-score was proposed to evaluate the combined effect of sensitivity and correlation analyses. For instance, the normalised SC-score at a 25% step size of the accumulated exclusion of variable lighting and spot visibility from apple black spot was 1.00, resulting from high ASA and PCC values which indicated that the 25% elimination of the dataset complexity should be considered to improve the overall accuracy of the DL model.

4.7. Limitations of the Study

Deployment Challenges: This study demonstrates the strength of the data-centric AI approach in solving industry-oriented problems in real-world applications. However, the successful integration of this methodology into agricultural systems requires addressing several deployment challenges. First, dataset maintenance is important to ensure the long-term effectiveness of the proposed methodology because of the diverse nature of the agricultural environment in terms of plant species, diseases, and environmental conditions. Therefore, new classes and training images must be added to the dataset to maintain the accuracy and robustness of the proposed method. Incremental training methods can be used instead of training the model from scratch with each increase/change in the training dataset. In the case of performance degradation, new dataset images should be thoroughly investigated by identifying and excluding the complexities in the dataset images based on non-parametric quality measures. Once dataset complexities are identified and excluded from the training dataset, the proposed methodology consists of sensitivity, correlation analysis, and SC-score which are post-processing steps that do not contribute to increasing the computation time of the proposed data-centric AI approach.

Second, the computational feasibility for end users (farmers) must be considered in the model. Although deep learning models are computationally intensive, deploying lightweight models can reduce the need for high-end hardware. Cloud-based solutions can also provide scalable and cost-effective access to computational resources to enable farmers to use the proposed approach without significant additional infrastructure investments.

Third, adaptation to various plant species and environments is essential for proving the robustness of this approach. Although this study was conducted on five different plant species and agricultural environments to test the proposed approach, greater diversity in plant species and agricultural environments should be considered in follow-up studies.

Other Challenges: Although study attained high accuracy after excluding the dataset complexities in different step sizes, there are still a few limitations that can be addressed in follow-up research. First, this study considered manual identification and exclusion of dataset complexities. This process is exhaustive for large datasets and should be automated using DL-based methods. Next, the proposed data-centric AI approach was evaluated using the NZDLPlantDisease-v1 and an external testing dataset. Further investigations should be conducted using other datasets of plant diseases. This can be achieved using the weights obtained from the DL model to classify the plant diseases. Moreover, the significance of the proposed approach was assessed using the newly developed scores, and additional analyses were performed using other or advanced performance scores.

This score was also designed to be interpreted by non-technical users in agricultural contexts. Interestingly, no weight factor was considered in the score, which can be added in future studies. Furthermore, this study is based on the classification of plant diseases, and the detection (localisation and classification) and segmentation of plant diseases should also be considered in future studies. These tasks present more challenges and enhance research on data-centric AI for plant disease recognition. Moreover, based on the methodology proposed in this study, sensitivity and correlation analyses were performed on the best DL model. This limitation can be addressed by testing the proposed approach on other advanced and latest DL architectures. Moreover, the analysis of linear and nonlinear relationships was not considered; hence, other correlation functions, such as Spearman's rank correlation, should be considered in the future.

The computational cost of the proposed methodology depends on three steps: (1) quantifying the complexity of the dataset, (2) training the deep learning models, and (3) evaluating the performance of the model using the SC score. Quantification of dataset complexity involves preprocessing the images of training datasets using statistical measurements that are linearly scaled by the dataset size. Training deep learning models is the most computationally demanding step in the process, and this demand increases significantly as the number of training samples increases. This is because larger datasets require more iterations to achieve training convergence, which leads to higher memory usage, longer processing times, and increased costs in terms of computational resources (e.g. GPU/CPU usage and energy consumption). To address these issues, distributed training on GPUs or TPUs can significantly reduce the training time for larger datasets. Cloud-based solutions, such as AWS, Google Cloud, and Azure, can provide scalable infrastructure for both data storage and computation. Furthermore, efficient data preprocessing pipelines with parallelism can reduce the calculations required to quantify dataset complexity. The steps to assess the performance of the proposed approach using the SC score are relatively simple and cost ineffective. Future research should examine the computational requirements of the proposed data-centric AI approaches for larger datasets.

5. Conclusion and Future Work

This study aimed to improve the accuracy of plant disease classification using a data-centric AI approach based on sensitivity and correlation analysis. After the comparative analysis of the DL classifiers, DenseNet-201 achieved the highest accuracy on the NZDLPlantDisease-v1 and an external testing dataset. The overall accuracy (OA) of the DenseNet-201 model on the testing dataset of NZDLPlantDisease-v1 was 95.11%. However, the OA of DenseNet-201 for the external testing dataset was 82.17%. The accuracy of all healthy and disease classes in the NZDLPlantDisease-v1 dataset was greater than 90%. However, eight classes achieved an accuracy of less than 90% when DenseNet-201 was tested using an external dataset. Thus, an improvement in the accuracy of the DenseNet-201 model for these classes was aimed at performing comprehensive analyses. The complexities of the dataset for these eight classes were identified and excluded from the training datasets. The accuracy of some classes improved to more than 90%; however, some did not achieve the desired accuracy.

To address this problem, a new sensitivity-and correlation analysis-based approach is proposed in this study. In this regard, the dataset complexities were excluded for different step sizes, and the most sensitive and correlated step sizes were obtained based on the improvement in the class and overall accuracies. Improvements of 10.94% and 3.40% in overall accuracies were observed on the selected plant disease dataset (NZDLPlantDisease-v1) and an external testing dataset. Moreover, a novel Sensitivity Correlation Score (SC-score) was proposed to evaluate sensitivity and correlation analyses using a single metric. This score revealed both the responsiveness and consistency of the performance improvement of the DL model owing to the exclusion of the dataset complexities. Hence, a high SC-score also confirmed that there was no loss of important information/features while removing the dataset complexities.

In the future, the automation of the exclusion of dataset complexities will be focused. Moreover, DL-based detection and segmentation models can be tested using the proposed methodology to enhance the localisation, identification, and segmentation of diseases in plants. In future work, we will focus on optimising dataset complexity exclusion strategies and extending the use of the proposed data-centric AI approach to other domains, such as medical image analysis for disease detection, remote sensing for environmental monitoring, and autonomous crop monitoring for precision agriculture. Moreover, a hybrid model and data-centric AI approach can be proposed by integrating DL models with the proposed data-centric AI approach. In addition, the proposed SC-score creates various opportunities for further research to improve the performance of DL models for plant disease classification, such as the automatic exclusion of dataset complexities by setting a threshold for excluding datasets or data points that are too complex or inconsistent. An value of the SC-score below a certain threshold can be automatically flagged for exclusion or further review.

CRediT authorship contribution statement

Muhammad Hammad Saleem: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Visualization, Data curation, Writing – original draft, Writing – review and editing. **Fakhia Hammad:** Data curation, Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft. **Muhammad Taha:** Data curation, Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft. **Shivakumara Palaiahnakote:** Methodology, Validation, Visualization, Investigation, Writing – review and editing. **Sadaqat ur Rehman:** Writing – review and editing. **Mohamad Saraee:** Writing – review and editing.

Declaration of Competing 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.

Data availability

We have used a publicly available dataset from the GitHub repository https://github.com/hsaleem1/NZDLPlantDisease-v1.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Abayomi-Alli, O. O., Damaševičius, R., Misra, S., & Maskeliūnas, R. (2021). Cassava disease recognition from low-quality images using enhanced data augmentation model and deep learning. *Expert Systems, 38*, e12746.

Abbas, A., Jain, S., Gour, M., & Vankudothu, S. (2021). Tomato plant disease detection using transfer learning with C-GAN synthetic images. *Computers and Electronics in Agriculture, 187*, 106279.

Alsubai, S., Dutta, A. K., Alkhayyat, A. H., Jaber, M. M., Abbas, A. H., & Kumar, A. (2023). Hybrid deep learning with improved Salp swarm optimization based multi-class grape disease classification model. *Computers and Electrical Engineering, 108*, 108733.

Anand, K., Jain, B., Mittal, H., & Yadav, V. K. (2025). QEFS: A novel plant disease prediction approach using quantum-inspired evolutionary feature selection. *Applied Intelligence, 55*, 1-23.

Argüeso, D., Picon, A., Irusta, U., Medela, A., San-Emeterio, M. G., Bereciartua, A., & Alvarez-Gila, A. (2020). Few-Shot Learning approach for plant disease classification using images taken in the field. *Computers and Electronics in Agriculture, 175*, 105542.

Arun, R. A., & Umamaheswari, S. (2023). Effective multi-crop disease detection using pruned complete concatenated deep learning model. *Expert Systems with Applications, 213*, 118905.

Barbedo, J. G. A. (2018). Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. *Computers and Electronics in Agriculture, 153*, 46-53.

Barbedo, J. G. A. (2019). Plant disease identification from individual lesions and spots using deep learning. *Biosystems engineering, 180*, 96-107.

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of machine learning research, 13*.

Bhagat, M., & Kumar, D. (2023). Efficient feature selection using BoWs and SURF method for leaf disease identification. *Multimedia Tools and Applications, 82*, 28187-28211.

Bhujade, V. G., Sambhe, V., & Banerjee, B. (2024). Digital image noise removal towards soybean and cotton plant disease using image processing filters. *Expert Systems with Applications, 246*, 123031.

Blanchet, G., & Moisan, L. (2012). An explicit sharpness index related to global phase coherence. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1065-1068): IEEE.

Brahimi, M., Arsenovic, M., Laraba, S., Sladojevic, S., Boukhalfa, K., & Moussaoui, A. (2018). Deep learning for plant diseases: detection and saliency map visualisation. *Human and machine learning: Visible, explainable, trustworthy and transparent*, 93-117.

Brahimi, M., Boukhalfa, K., & Moussaoui, A. (2017). Deep learning for tomato diseases: classification and symptoms visualization. *Applied Artificial Intelligence, 31*, 299-315.

Cap, Q. H., Uga, H., Kagiwada, S., & Iyatomi, H. (2020). Leafgan: An effective data augmentation method for practical plant disease diagnosis. *IEEE Transactions on Automation Science and Engineering, 19*, 1258-1267.

Chen, J., Chen, J., Zhang, D., Sun, Y., & Nanehkaran, Y. A. (2020). Using deep transfer learning for image-based plant disease identification. *Computers and Electronics in Agriculture, 173*, 105393.

Chen, Y., & Wu, Q. (2023). Grape leaf disease identification with sparse data via generative adversarial networks and convolutional neural networks. *Precision Agriculture, 24*, 235-253.

Chhetri, T. R., Hohenegger, A., Fensel, A., Kasali, M. A., & Adekunle, A. A. (2023). Towards improving prediction accuracy and user-level explainability using deep learning and knowledge graphs: A study on cassava disease. *Expert Systems with Applications, 233*, 120955.

Dong, J., Lee, J., Fuentes, A., Xu, M., Yoon, S., Lee, M. H., & Park, D. S. (2022). Data-centric annotation analysis for plant disease detection: Strategy, consistency, and performance. *Frontiers in plant science, 13*, 1037655.

Dorgham, O., Abu-Shareah, G., Alzubi, O., Al Shaqsi, J., Aburass, S., & Al-Betar, M. (2024). Grasshopper Optimization Algorithm and Neural Network Classifier for Detection and Classification of Barley Leaf Diseases. *IEEE Open Journal of the Computer Society*.

Douarre, C., Crispim-Junior, C. F., Gelibert, A., Tougne, L., & Rousseau, D. (2019). Novel data augmentation strategies to boost supervised segmentation of plant disease. *Computers and Electronics in Agriculture, 165*, 104967.

Dubey, R. K., & Choubey, D. K. (2024). An efficient adaptive feature selection with deep learning model-based paddy plant leaf disease classification. *Multimedia Tools and Applications, 83*, 22639-22661.

Huang, Y., Li, R., Wei, X., Wang, Z., Ge, T., & Qiao, X. (2022). Evaluating data augmentation effects on the recognition of sugarcane leaf spot. *Agriculture, 12*, 1997.

Kamal, K., Yin, Z., Wu, M., & Wu, Z. (2019). Depthwise separable convolution architectures for plant disease classification. *Computers and Electronics in Agriculture, 165*, 104948.

Kaur, P., Harnal, S., Gautam, V., Singh, M. P., & Singh, S. P. (2023). A novel transfer deep learning method for detection and classification of plant leaf disease. *Journal of Ambient Intelligence and Humanized Computing, 14*, 12407-12424.

Lee, S. H., Goëau, H., Bonnet, P., & Joly, A. (2020). New perspectives on plant disease characterization based on deep learning. *Computers and Electronics in Agriculture, 170*, 105220.

Li, M., Zhou, G., Chen, A., Yi, J., Lu, C., He, M., & Hu, Y. (2022). FWDGAN-based data augmentation for tomato leaf disease identification. *Computers and Electronics in Agriculture, 194*, 106779.

Li, X., Li, X., Zhang, M., Dong, Q., Zhang, G., Wang, Z., & Wei, P. (2024). SugarcaneGAN: A novel dataset generating approach for sugarcane leaf diseases based on lightweight hybrid CNN-Transformer network. *Computers and Electronics in Agriculture, 219*, 108762.

Liu, B., Tan, C., Li, S., He, J., & Wang, H. (2020). A data augmentation method based on generative adversarial networks for grape leaf disease identification. *IEEE Access, 8*, 102188-102198.

Liu, J., & Wang, X. (2020). Tomato diseases and pests detection based on improved Yolo V3 convolutional neural network. *Frontiers in plant science, 11*, 898.

Min, B., Kim, T., Shin, D., & Shin, D. (2023). Data augmentation method for plant leaf disease recognition. *Applied Sciences, 13*, 1465.

Mittal, A., Moorthy, A. K., & Bovik, A. C. (2011). Blind/referenceless image spatial quality evaluator. In *2011 conference record of the forty fifth asilomar conference on signals, systems and computers (ASILOMAR)* (pp. 723-727): IEEE.

Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in plant science, 7*, 1419.

Mu, J., Feng, Q., Yang, J., Zhang, J., & Yang, S. (2024). Few-shot disease recognition algorithm based on supervised contrastive learning. *Frontiers in plant science, 15*, 1341831.

Mukhopadhyay, S., Paul, M., Pal, R., & De, D. (2021). Tea leaf disease detection using multi-objective image segmentation. *Multimedia Tools and Applications, 80*, 753-771.

Mzoughi, O., & Yahiaoui, I. (2023). Deep learning-based segmentation for disease identification. *Ecological Informatics, 75*, 102000.

Nagpal, J., & Goel, L. (2025). ResdenseNet: a lightweight dense ResNet enhanced with depthwise separable convolutions and its applications for early plant disease classification. *Neural Computing and Applications*, 1-22.

Narayanan, K. L., Krishnan, R. S., Robinson, Y. H., Julie, E. G., Vimal, S., Saravanan, V., & Kaliappan, M. (2022). Banana plant disease classification using hybrid convolutional neural network. *Computational Intelligence and Neuroscience, 2022*, 9153699.

Ngugi, L. C., Abdelwahab, M., & Abo-Zahhad, M. (2023). A new approach to learning and recognizing leaf diseases from individual lesions using convolutional neural networks. *Information Processing in Agriculture, 10*, 11-27.

Nuthalapati, S. V., & Tunga, A. (2021). Multi-domain few-shot learning and dataset for agricultural applications. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 1399-1408).

Pan, J., Xia, L., Wu, Q., Guo, Y., Chen, Y., & Tian, X. (2022). Automatic strawberry leaf scorch severity estimation via faster R-CNN and few-shot learning. *Ecological Informatics, 70*, 101706.

Pandi, S. S., Senthilselvi, A., Gitanjali, J., ArivuSelvan, K., Gopal, J., & Vellingiri, J. (2022). Rice plant disease classification using dilated convolutional neural network with global average pooling. *Ecological Modelling, 474*, 110166.

Ravi, V., Acharya, V., & Pham, T. D. (2022). Attention deep learning-based large-scale learning classifier for Cassava leaf disease classification. *Expert Systems, 39*, e12862.

Rezaei, M., Diepeveen, D., Laga, H., Jones, M. G., & Sohel, F. (2024). Plant disease recognition in a low data scenario using few-shot learning. *Computers and Electronics in Agriculture, 219*, 108812.

Saleem, M. H., Potgieter, J., & Arif, K. M. (2022a). A performance-optimized deep learning-based plant disease detection approach for horticultural crops of New Zealand. *IEEE Access, 10*, 89798-89822.

Saleem, M. H., Potgieter, J., & Arif, K. M. (2022b). A weight optimization-based transfer learning approach for plant disease detection of New Zealand vegetables. *Frontiers in plant science, 13*, 1008079.

Sharma, V., Tripathi, A. K., Daga, P., Nidhi, M., & Mittal, H. (2024). ClGanNet: A novel method for maize leaf disease identification using ClGan and deep CNN. *Signal Processing: Image Communication, 120*, 117074.

Singh, R. S. R., & Sanodiya, R. K. (2023). Zero-Shot Transfer Learning Framework for Plant Leaf Disease Classification. *IEEE Access*.

Stephen, A., Punitha, A., & Chandrasekar, A. (2024). Optimal deep generative adversarial network and convolutional neural network for rice leaf disease prediction. *The Visual Computer, 40*, 919-936.

Sun, R., Zhang, M., Yang, K., & Liu, J. (2020). Data enhancement for plant disease classification using generated lesions. *Applied Sciences, 10*, 466.

Taji, K., Sohail, A., Shahzad, T., Khan, B. S., Khan, M. A., & Ouahada, K. (2024). An ensemble hybrid framework: a comparative analysis of metaheuristic algorithms for ensemble hybrid CNN features for plants disease classification. *IEEE Access*.

Tassis, L. M., & Krohling, R. A. (2022). Few-shot learning for biotic stress classification of coffee leaves. *Artificial Intelligence in Agriculture, 6*, 55-67.

Tewari, A. S., & Kumari, P. (2024). Lightweight modified attention based deep learning model for cassava leaf diseases classification. *Multimedia Tools and Applications, 83*, 57983-58007.

Too, E. C., Yujian, L., Njuki, S., & Yingchun, L. (2019). A comparative study of fine-tuning deep learning models for plant disease identification. *Computers and Electronics in Agriculture, 161*, 272-279.

Uskaner Hepsağ, P. (2024). Efficient plant disease identification using few-shot learning: a transfer learning approach. *Multimedia Tools and Applications, 83*, 58293-58308.

Wang, X., & Cao, W. (2023). GACN: generative adversarial classified network for balancing plant disease dataset and plant disease recognition. *Sensors, 23*, 6844.

Wu, Q., Chen, Y., & Meng, J. (2020). DCGAN-based data augmentation for tomato leaf disease identification. *IEEE Access, 8*, 98716-98728.

Xu, M., Yoon, S., Jeong, Y., & Park, D. S. (2022). Transfer learning for versatile plant disease recognition with limited data. *Frontiers in plant science, 13*, 1010981.

Zhang, K., Wu, Q., & Chen, Y. (2021). Detecting soybean leaf disease from synthetic image using multi-feature fusion faster R-CNN. *Computers and Electronics in Agriculture, 183*, 106064.

Zhang, S., & Zhang, C. (2023). Modified U-Net for plant diseased leaf image segmentation. *Computers and Electronics in Agriculture, 204*, 107511.

Zhao, S., Liu, J., & Wu, S. (2022). Multiple disease detection method for greenhouse-cultivated strawberry based on multiscale feature fusion Faster R_CNN. *Computers and Electronics in Agriculture, 199*, 107176.

Zhao, Y., Zhang, Z., Wu, N., Zhang, Z., & Xu, X. (2024). MAFDE-DN4: Improved Few-shot plant disease classification method based on Deep Nearest Neighbor Neural Network. *Computers and Electronics in Agriculture, 226*, 109373.

ORCID Information of all authors

Muhammad Hammad Saleem: 0000-0002-3625-3021

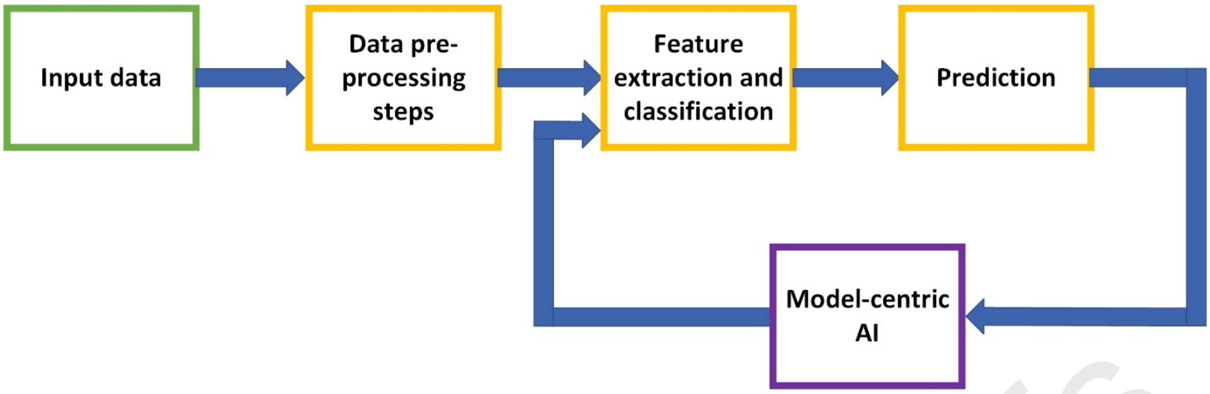
Fakhia Hammad: 0009-0005-9465-7453

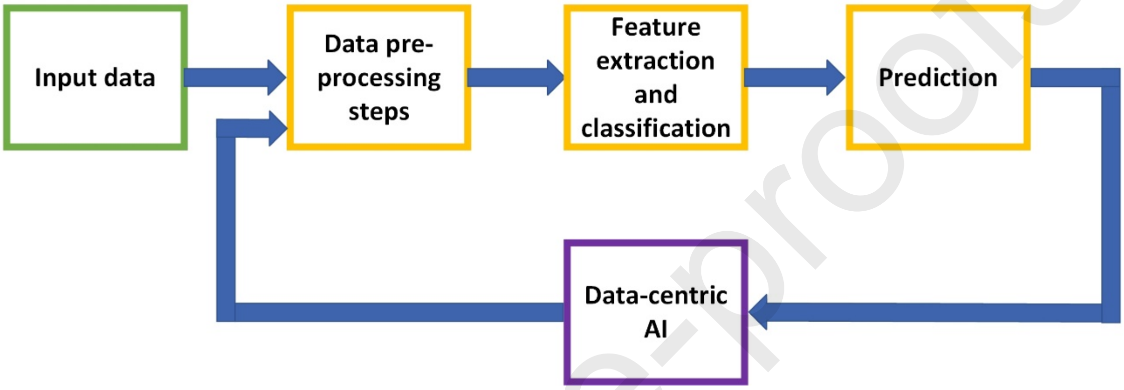
Muhammad Taha: 0009-0002-2290-8872

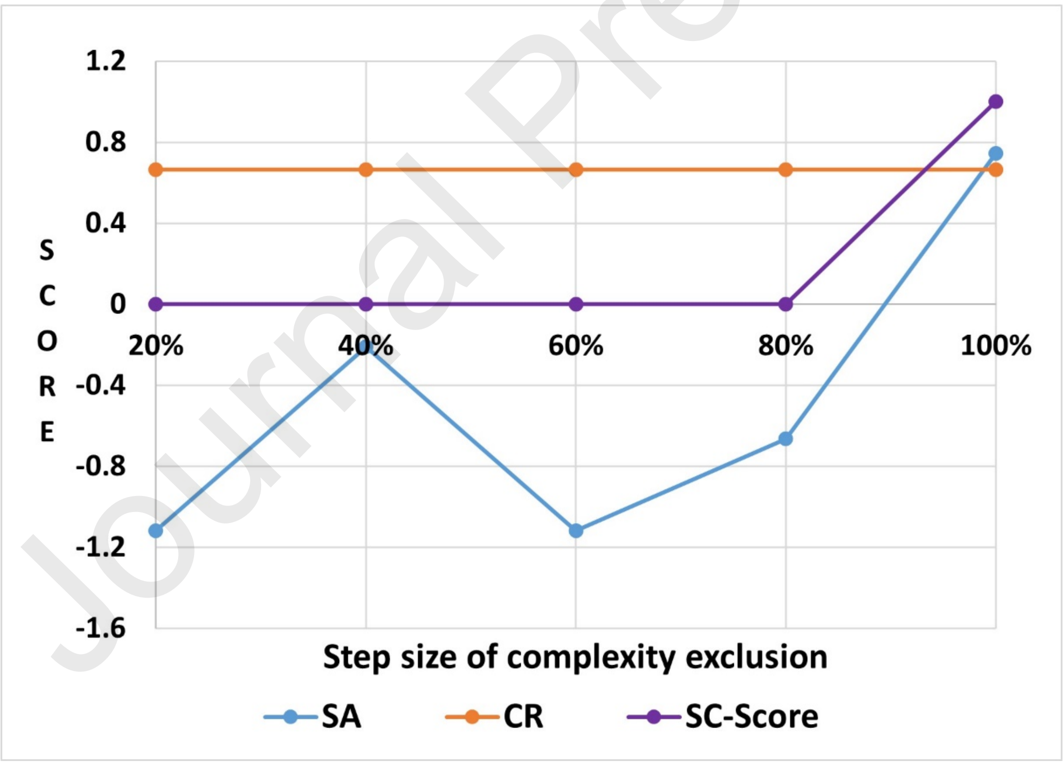
Shivakumara Palaiahnakote: 0000-0001-9026-4613

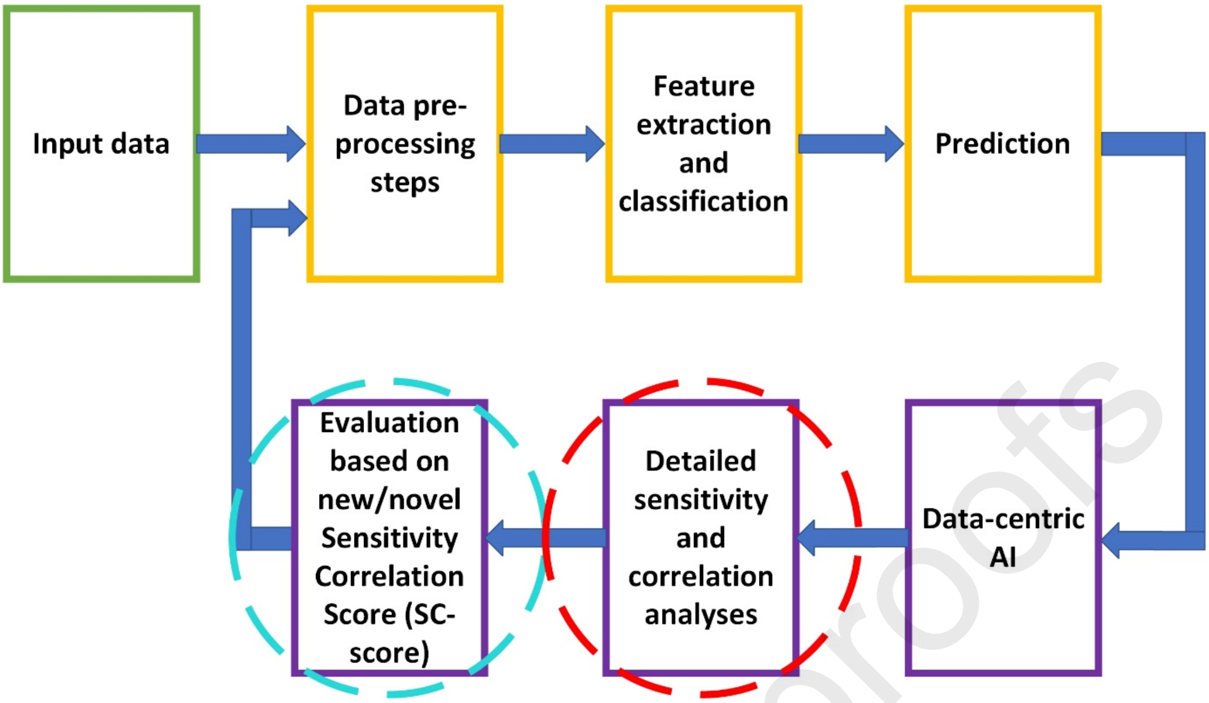
Sadaqat ur Rehman: 0000-0001-7823-3814

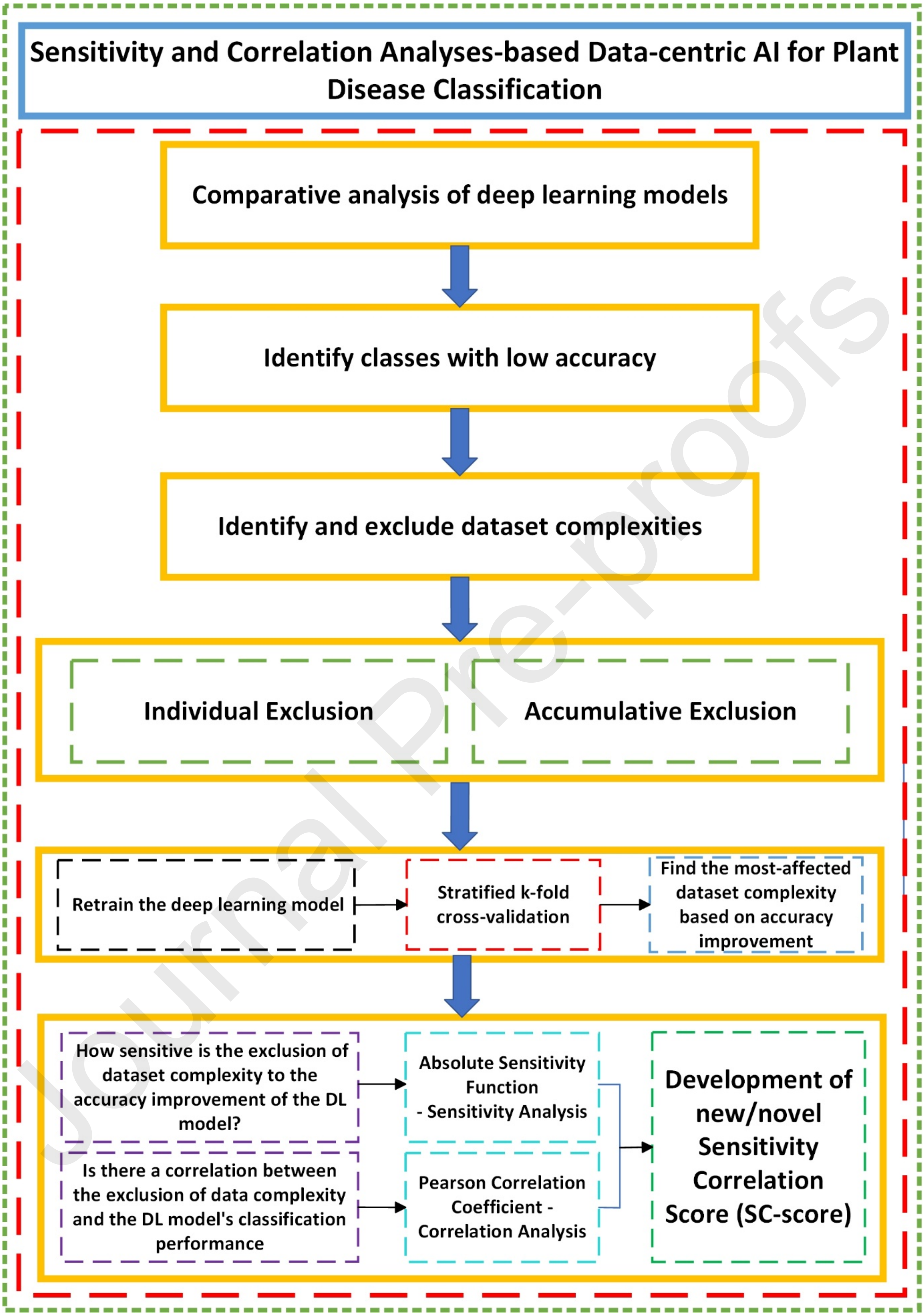
Mohamad Saraee: 0000-0002-3283-1912

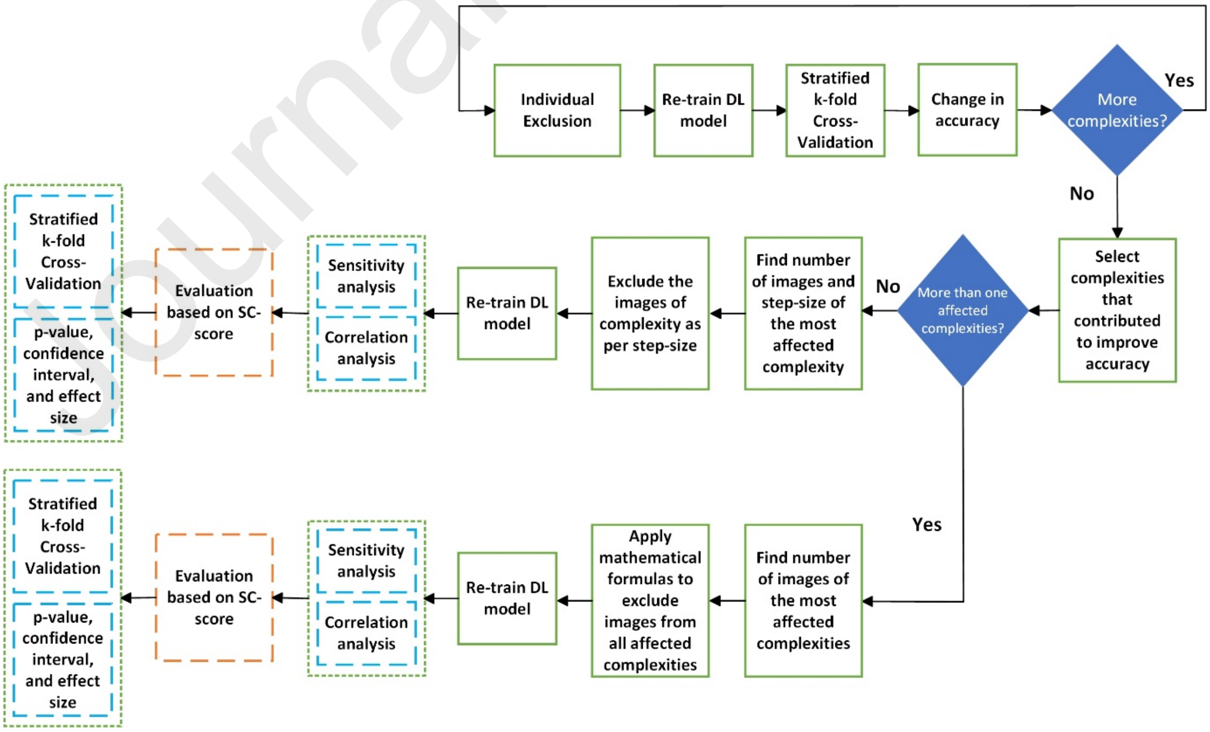

Input data
Data pre-processing steps
Feature extraction and classification
Prediction
Model-centric AI

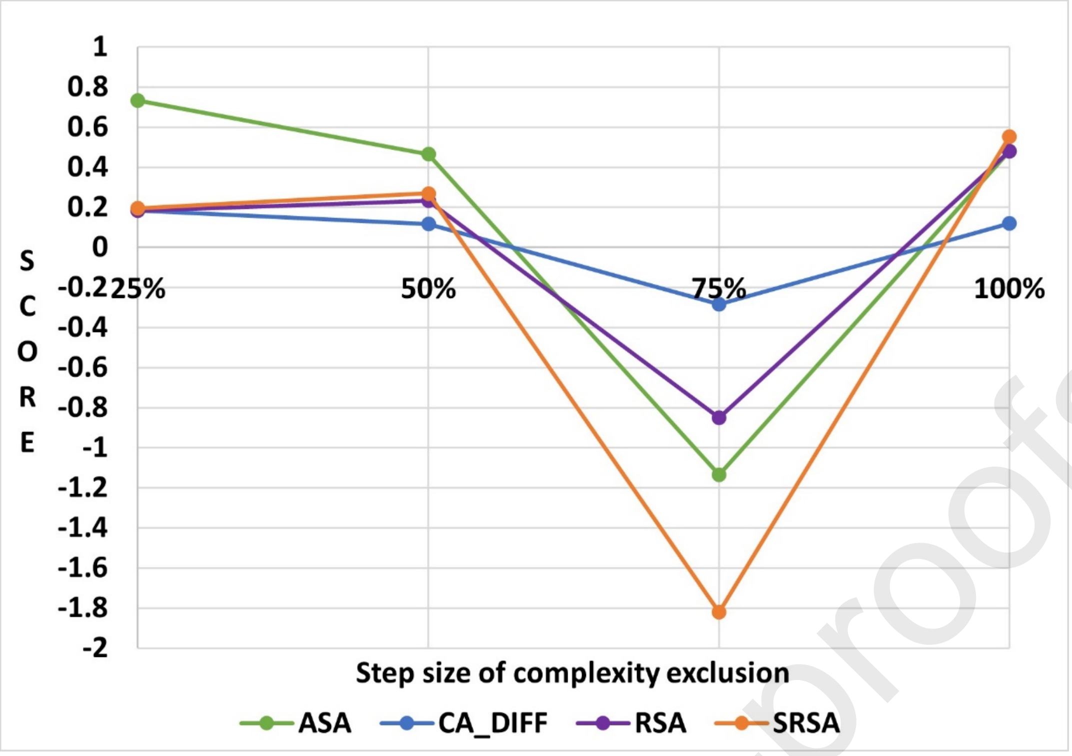

Input data
Data pre-processing steps
Feature extraction and classification
Prediction
Data-centric AI

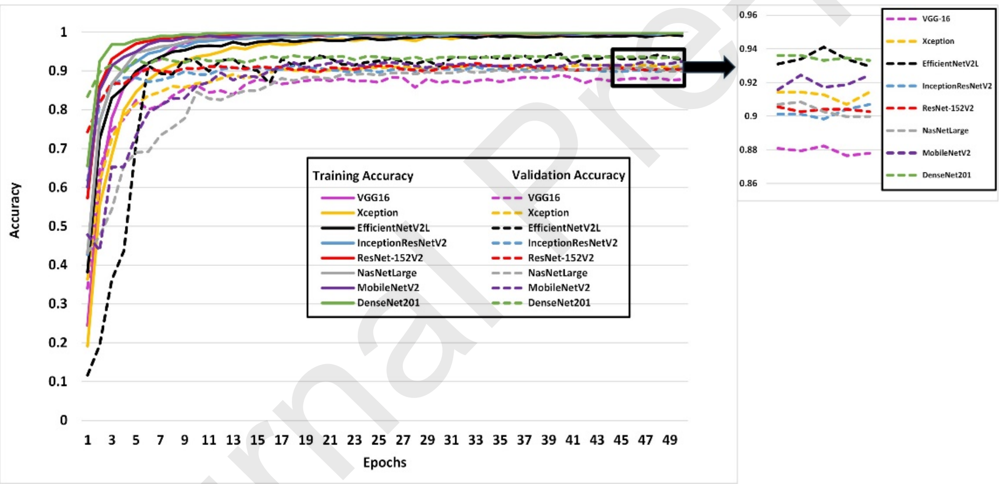

SCORE
1.2
0.8
0.4
0
-0.4
-0.8
-1.2
-1.6
20%
40%
60%
80%
100%
Step size of complexity exclusion
SA
CR
SC-Score

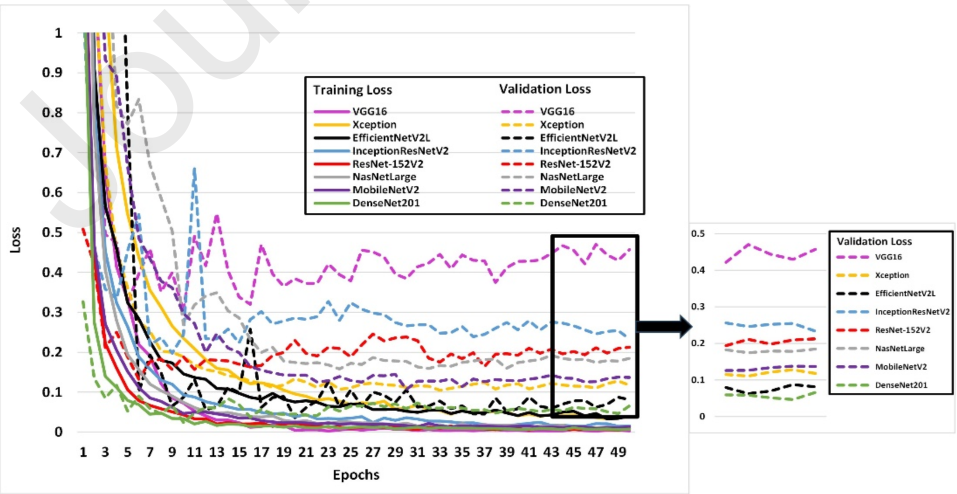

Input data
Data pre-processing steps
Feature extraction and classification
Prediction
Data-centric AI
Detailed sensitivity and correlation analyses
Evaluation based on new/novel Sensitivity Correlation Score (SC-score)

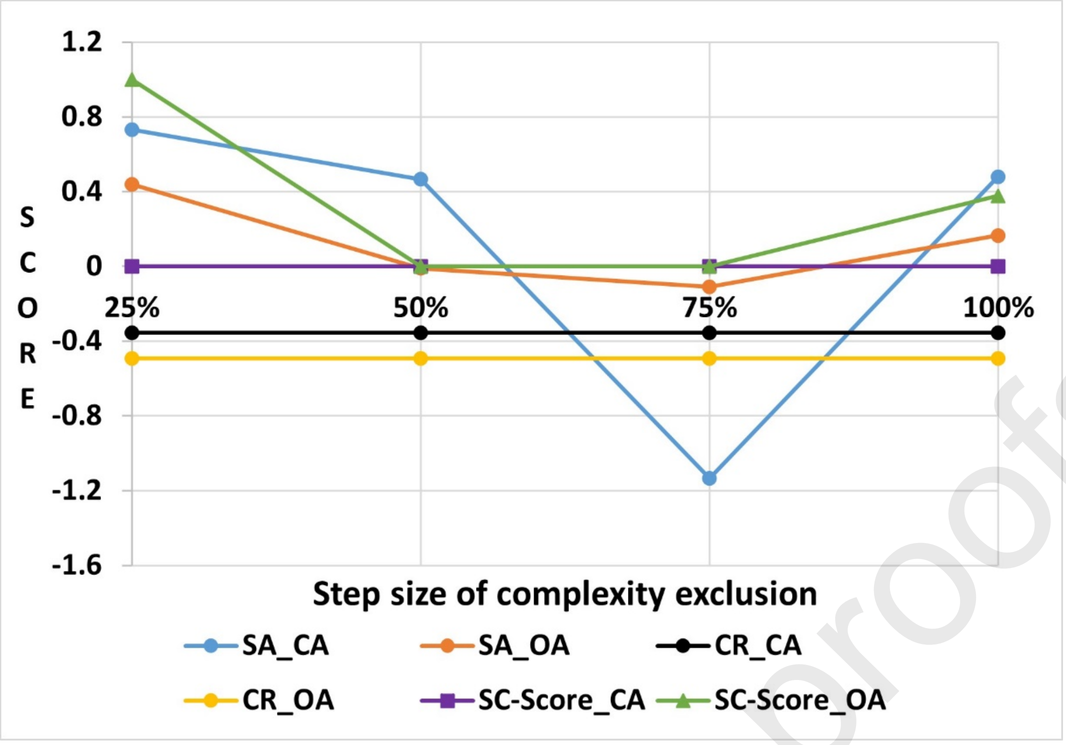

Sensitivity and Correlation Analyses-based Data-centric AI for Plant Disease Classification
Comparative analysis of deep learning models
Identify classes with low accuracy
Identify and exclude dataset complexities
Individual Exclusion
Accumulative Exclusion
Retrain the deep learning model
Stratified k-fold cross-validation
Find the most-affected dataset complexity based on accuracy improvement
How sensitive is the exclusion of dataset complexity to the accuracy improvement of the DL model?
Absolute Sensitivity Function - Sensitivity Analysis
Is there a correlation between the exclusion of data complexity and the DL model's classification performance
Pearson Correlation Coefficient - Correlation Analysis
Development of new/novel Sensitivity Correlation Score (SC-score)

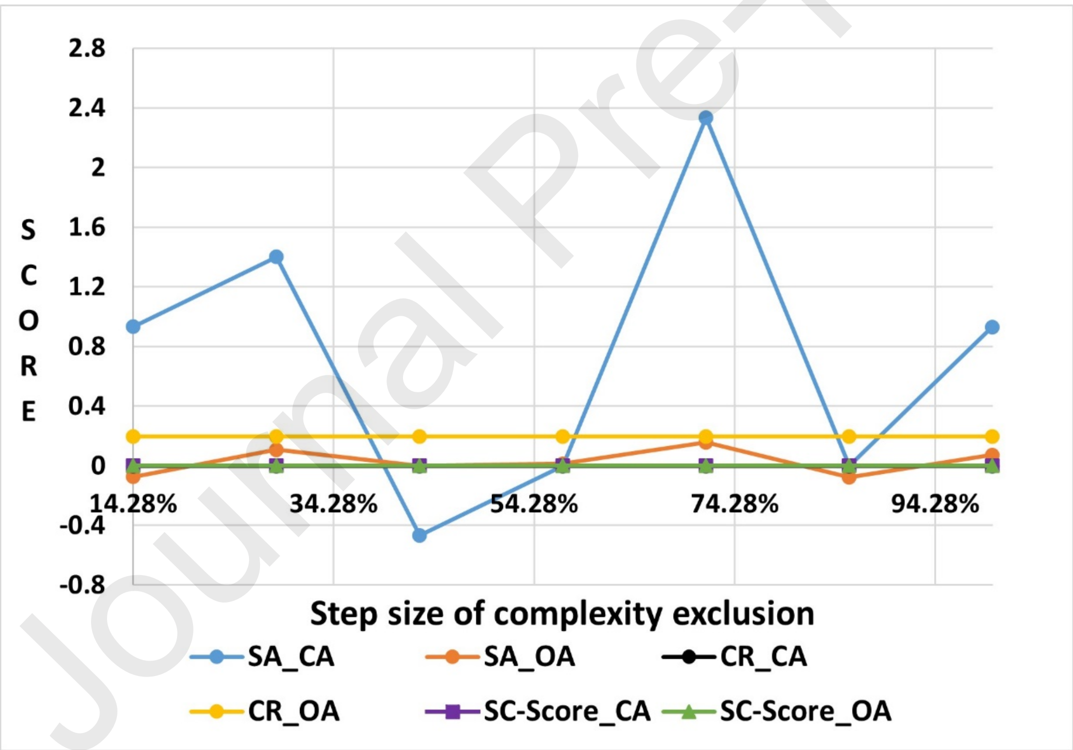

Individual Exclusion
Re-train DL model
Stratified k-fold Cross-Validation
Change in accuracy
More complexities?
Yes
No
Select complexities that contributed to improve accuracy
More than one affected complexities?
No
Find number of images and step-size of the most affected complexity
Exclude the images of complexity as per step-size
Re-train DL model
Sensitivity analysis
Correlation analysis
Evaluation based on SC-score
Stratified k-fold Cross-Validation
p-value, confidence interval, and effect size
Yes
Find number of images of the most affected complexities
Apply mathematical formulas to exclude images from all affected complexities
Re-train DL model
Sensitivity analysis
Correlation analysis
Evaluation based on SC-score
Stratified k-fold Cross-Validation
p-value, confidence interval, and effect size

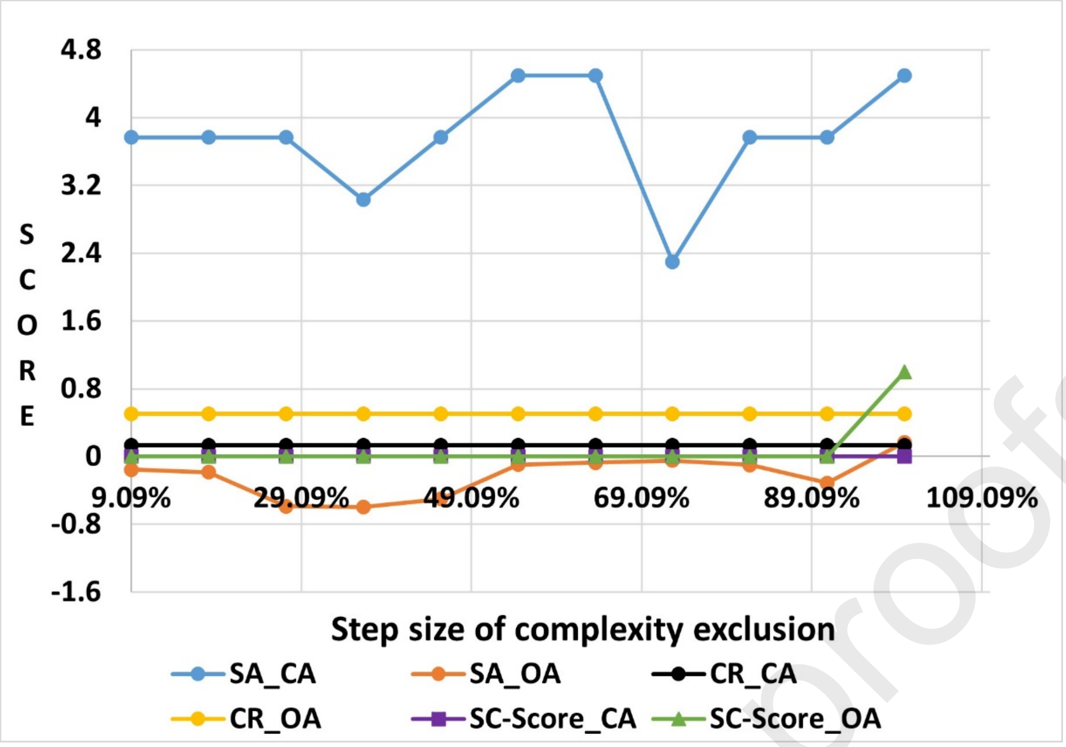

1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-1.2
-1.4
-1.6
-1.8
-2
SCORE
25%
50%
75%
100%
Step size of complexity exclusion
ASA
CA_DIFF
RSA
SRSA

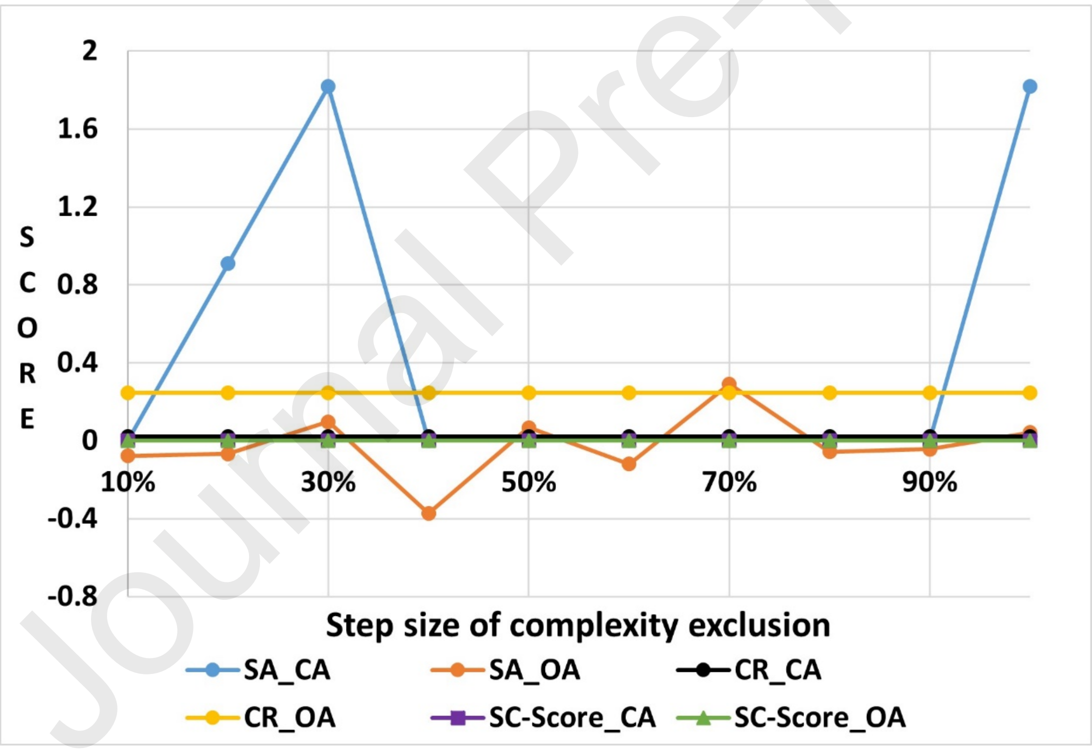

Accuracy
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49
Epochs
Training Accuracy
VGG16
Xception
EfficientNetV2L
InceptionResNetV2
ResNet-152V2
NasNetLarge
MobileNetV2
DenseNet201
Validation Accuracy
VGG16
Xception
EfficientNetV2L
InceptionResNetV2
ResNet-152V2
NasNetLarge
MobileNetV2
DenseNet201
0.96
0.94
0.92
0.9
0.88
0.86
VGG-16
Xception
EfficientNetV2L
InceptionResNetV2
ResNet-152V2
NasNetLarge
MobileNetV2
DenseNet201

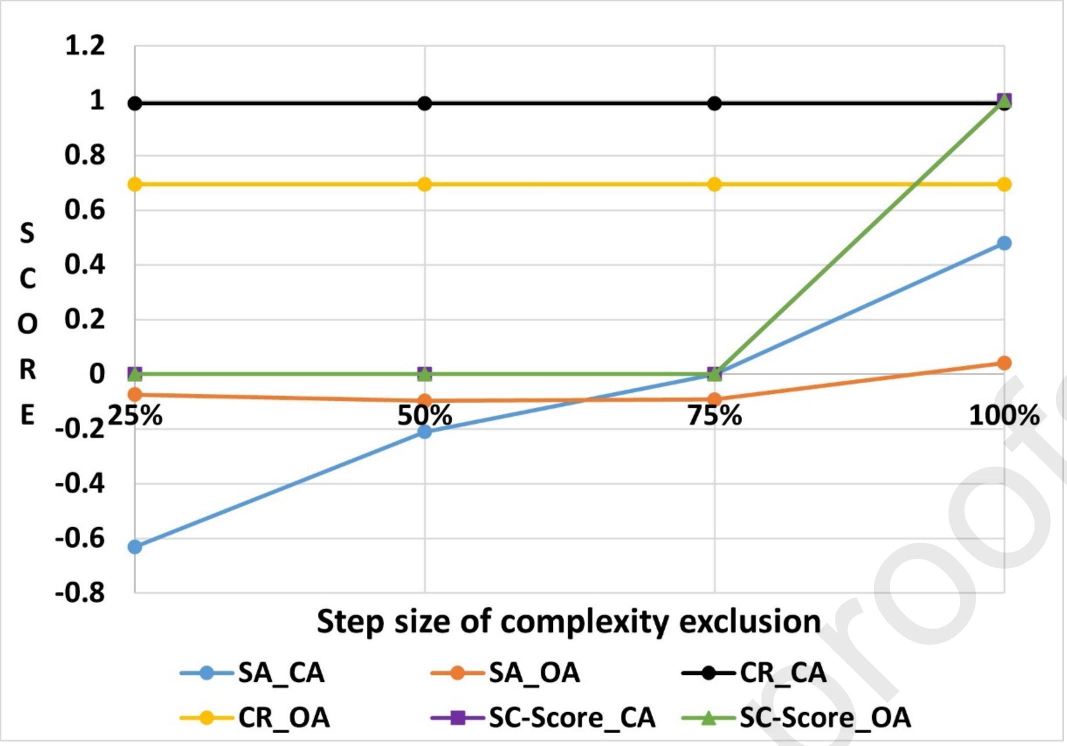

Loss
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49
Epochs
Training Loss
VGG16
Xception
EfficientNetV2L
InceptionResNetV2
ResNet-152V2
NasNetLarge
MobileNetV2
DenseNet201
Validation Loss
VGG16
Xception
EfficientNetV2L
InceptionResNetV2
ResNet-152V2
NasNetLarge
MobileNetV2
DenseNet201
0.5
0.4
0.3
0.2
0.1
0
Validation Loss
VGG16
Xception
EfficientNetV2L
InceptionResNetV2
ResNet-152V2
NasNetLarge
MobileNetV2
DenseNet201

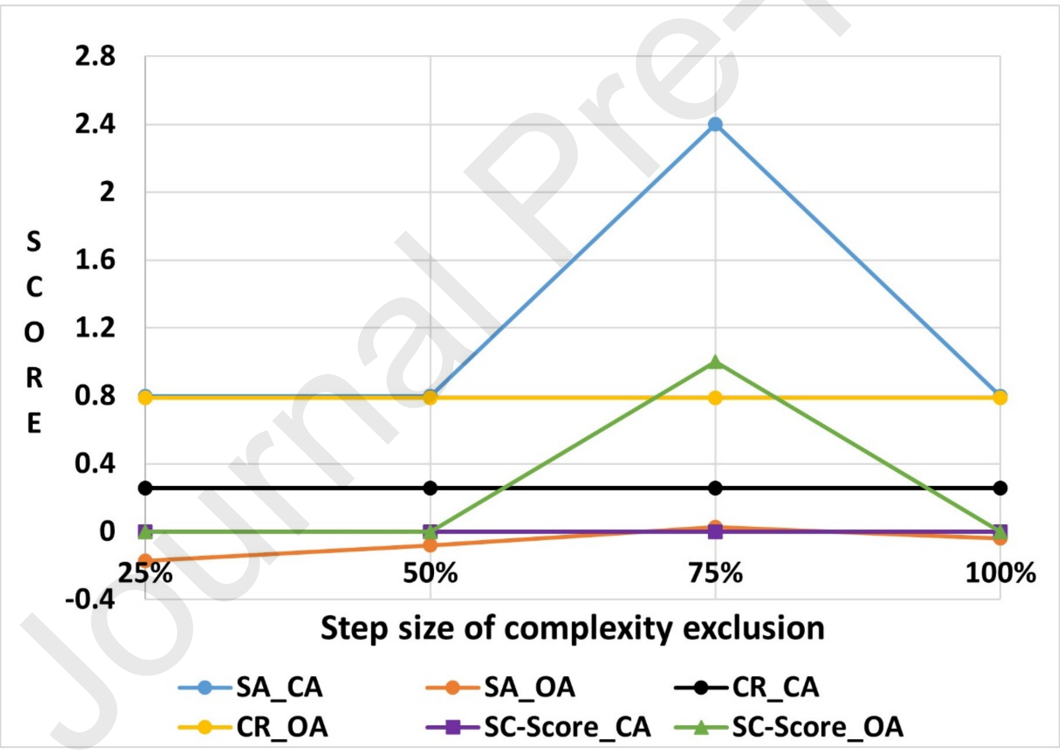

1.2
0.8
0.4
0
-0.4
-0.8
-1.2
-1.6
SCORE
25%
50%
75%
100%
Step size of complexity exclusion
SA_CA
SA_OA
CR_CA
CR_OA
SC-Score_CA
SC-Score_OA

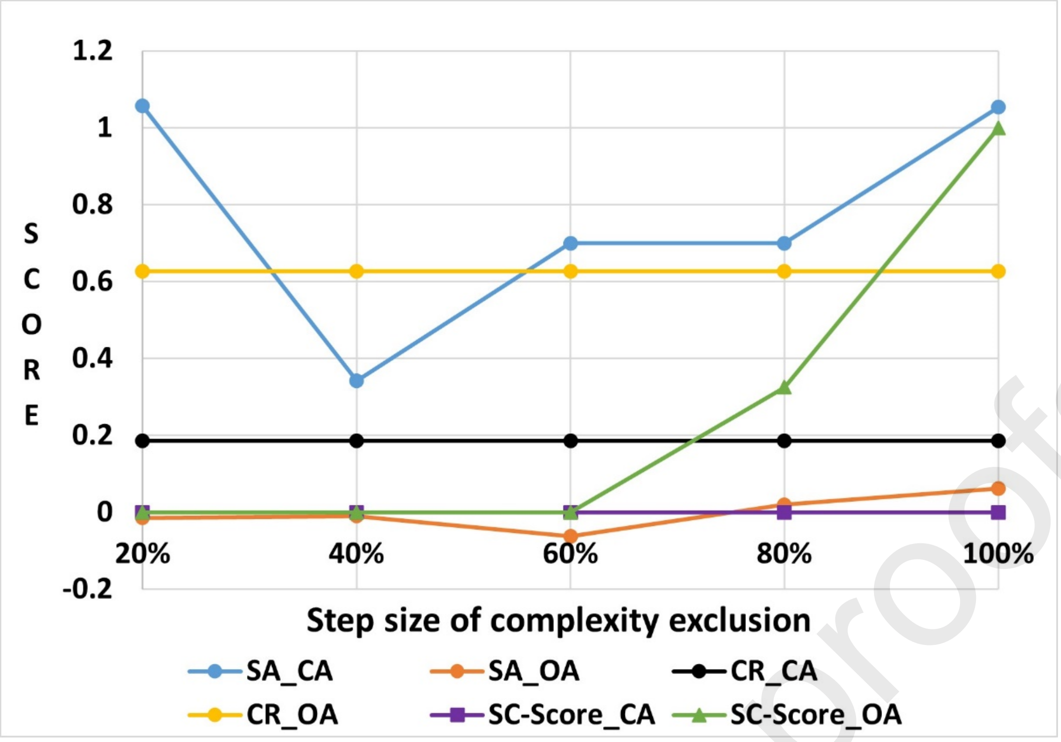

2.8
2.4
2
1.6
1.2
0.8
0.4
0
-0.4
-0.8
SCORE
14.28%
34.28%
54.28%
74.28%
94.28%
Step size of complexity exclusion
SA_CA
SA_OA
CR_CA
CR_OA
SC-Score_CA
SC-Score_OA

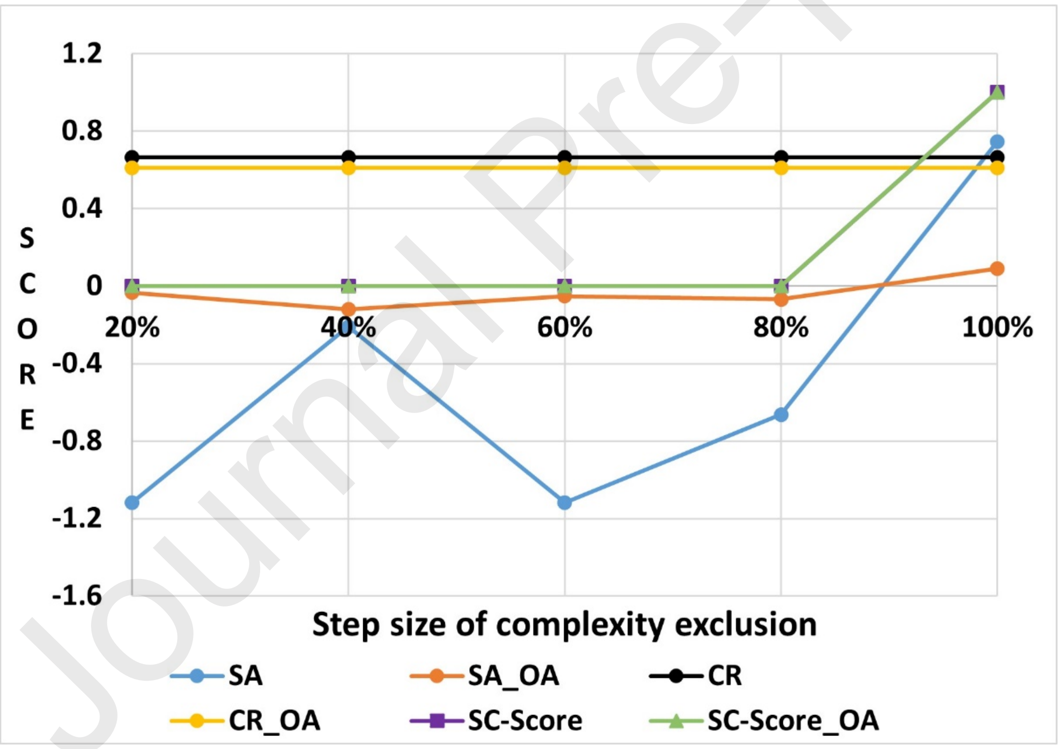

4.8
4
3.2
2.4
1.6
0.8
0
-0.8
-1.6
SCORE
9.09%
29.09%
49.09%
69.09%
89.09%
109.09%
Step size of complexity exclusion
SA_CA
SA_OA
CR_CA
CR_OA
SC-Score_CA
SC-Score_OA


2
1.6
1.2
0.8
0.4
0
-0.4
-0.8
SCORE
10%
30%
50%
70%
90%
Step size of complexity exclusion
SA_CA
SA_OA
CR_CA
CR_OA
SC-Score_CA
SC-Score_OA


1.2
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
SCORE
25%
50%
75%
100%
Step size of complexity exclusion
SA_CA
SA_OA
CR_CA
CR_OA
SC-Score_CA
SC-Score_OA


2.8
2.4
2
1.6
1.2
0.8
0.4
0
-0.4
SCORE
25%
50%
75%
100%
Step size of complexity exclusion
SA_CA
SA_OA
CR_CA
CR_OA
SC-Score_CA
SC-Score_OA


1.2
1
0.8
0.6
0.4
0.2
0
-0.2
SCORE
20%
40%
60%
80%
100%
Step size of complexity exclusion
SA_CA
SA_OA
CR_CA
CR_OA
SC-Score_CA
SC-Score_OA


1.2
0.8
0.4
0
-0.4
-0.8
-1.2
-1.6
SCORE
20%
40%
60%
80%
100%
Step size of complexity exclusion
SA
SA_OA
CR
CR_OA
SC-Score
SC-Score_OA