
Introduction

- ## Rice
  - Philippines is agriculturally active - Sun et. al.
    - *Rice is heavily traded, increase in production may solidify the slot for global market* - World Rice Production and Trade
  - Symptoms of disease are evident in the leaf - Pascual et. al.
    1. Rice Tungro
    2. Sheath Blight
    3. Rice Blast
    4. Bacterial Leaf Blight
- ## Classification
  - *Traditional method of classification can be time consuming* - Sethy et. al.
  - ### Deep Learning - Liu & Wang
    - Higher Accuracy
    - Higher Computational Cost
    - Low Interpretability
    - Requires Large Dataset
  - ### Classical Approach - Sujatha et. al.
    - Effective
    - Lower Computational Cost
    - Higher Computational Complexity
    - Overfits due to amount of features - Vishnoi et. al., Seays et. al.
- ## Feature Selection
  - **Selects the best feature subset** - Nguyen et. al.
  - Manual Selection is time consuming and biased - Liu & Wang et. al.
  - Can reduce computational complexity - Nguyen et. al.
  - Reduce the amount of features - Seays et. al.
- ## Image Processing
  - **Pre-processing - Ramanjot et. al.**
    - For preparing the images prior to processing, this could be data augmentation, enhancement, denoise, deblur, rotation etc.

- **Segmentation - Sethy et. al.**
  - For separating an object of interest from the rest of image.
    - In the case of rice leaf classification, the object of interest is the rice leaf disease
- **Feature extraction - Sethy et. al.**
  - Features refers to any characteristic. In rice leaf disease classification, this could be the texture, shape, color, etc. of the disease
- **Classification - Corominas**
  - Grouping or clustering image or images

Evolutionary Algorithms

- Effective in different fields - Liu et. al.
- A direction to improving feature selection in plant disease classification - Shrivastava et. al.
- **Swarm Optimization Algorithms - Rostami et. al.**
  - Search and exploit near optimal solutions
  - Employs swarm-based intelligence
- **No Free Lunch Theorem - Zivkovic et. al. & Lapajne et. al.**
  - States that no optimization techniques will perform better than another
  - No one-size-fits-all optimization algorithm that can outperform other algorithms

The performance of these approaches was examined in terms of average accuracy over six data sets taken from the UCI machine learning repository. These data sets are: "Audiology," "Column," "Breast cancer," "Multiple features (Fourier)," "German credit," and "Nursery." To a great extent, the results obtained are consistent with previous research. On the other hand, according to the NFL theorem the tested algorithms should expose the same degree of accuracy. However, this is valid when a sufficiently large number of data sets are available. The authors underline that some common assumptions pertain the data sets. These common assumptions concern the Occam's razor and the independent identical distribution of the samples as well as, mainly, the data-dependent structural properties found in the data sets, that is, determinism and the Pareto principle. Based on these last properties they explain the peculiarities of the data sets and the results concerning the accuracy. Then, it is clear that not all the algorithms perform equally well on all problems.

In addition to the above, the authors perform a number of experiments using kernel machines and especially support vector machines (SVM) as well as deep learning networks. The results obtained show that SVM outperform the other learning algorithms while the performance of deep learning on these small and relatively simple problems is disappointing. In fact while these architectures are designed to handle complex data sets which have inherent abstraction layers they seem to be incapable to cope with simpler data sets with possibly lower data abstraction. This shows that NFL applies even in the case of deep learning which is also subject to limitations as other machine learning algorithms.

In terms of conclusion the authors state that: "*While evaluating the average accuracy ranking for the six data sets, they noticed the effect of the NFL theorem and how assumptions are key to performance.*" Comparing with similar research work they conclude that: "*the data and its pre-processing are as important as, if*

not more so than, the algorithm itself in determining the quality of the model. Data visualization or statistical techniques such as feature selection can be crucial to provide a better fit and obtain simpler and better models."

Source Link

Theoretical

- ## Image Processing
    1. **Acquisition**
       : Obtaining the datasets
    2. **Annotation**
       : Labelling of datasets appropriate to their classification
    3. **Pre-processing**
       : Improves the images for processing
    4. **Extraction**
       : Extracts the features from the image
    5. **Selection**
       : Selects the best features subset from extracted features
    6. **Classification**
       : Groups the input image into the appropriate class
- ## Feature Selection
    1. **Initialization**
       : Original features acts as input
    2. **Subset Generation**
       : Generate a subset from the original features
    3. **Subset Evaluation**
       : Evaluate the fitness of the subset features
    4. **Stop Condition**
       : Criteria for stopping the selection of features.
       - Can be an iteration count
       - Can be a conditional statement
    5. **Selected Subset**
       : Returns the best subset
- ## Wrapper Approach
    1. **Search Algorithm**
       : Refers to the Swarm Optimization technique
    2. **Feature Evaluation**
       : Selection of subset feature using;
    3. **Classification Algorithm**
       : The classifier, will act as the feature function

Conceptual

- **Input**
    - Datasets
    - Feature Selectors
- **Process**
    - Image Processing
    - Wrapper Approach
- **Output**
    - Performance Metrics

SOP

- ## Problems
  - Avoid Overfitting
  - Automatic Feature Selection
  - Comparative Analysis (Because of NFL)
  - Develop Rice Research
- ## Criteria
  1. F-Score
  2. Recall
  3. Precision
  4. Accuracy
  5. True Negative Ratio & True Positive Ratio
- ## Research Questions
  1. Overall performance of tool without feature selector
  2. Overall performance of tool with ACO as feature selector
  3. Difference of performance of tools with ACO as feature selector and without feature selector
  4. Overall performance of tool with PSO as feature selector
  5. Difference of performance of tools with PSO as feature selector and without feature selector
  6. Overall performance of tool with ABC as feature selector
  7. Difference of performance of tools with ABC as feature selector and without feature selector
  8. Difference of performance of tools

Hypothesis

- ## $H_{(o)}$ Null Hypothesis

  > There is no significant difference

- ## $H_{(a)}$ Alternative Hypothesis

  > There is a significant difference

SAL

- ## Datasets - Classification
  - Healthy Rice
  - Rice Tungro
  - Rice Blast
  - Sheath Blight
  - Bacterial Leaf Blight
- ## Techniques
  - No Feature Selector
  - PSO as Feature Selector
  - ABC as Feature Selector

- ACO as Feature Selector
- ## Tool
  - One Tool
  - Select a Model
  - Classifies into one of the Dataset
  - Rice Leaf only
- ## Metrics
  - Performance Metrics
  - ANOVA for hypothesis testing

SOS

1. Researchers
2. Government
3. Agriculture

RRL

- ## Theme#1-3: Rice
  - Rice Leaf Morphology is different from the morphology of other leaf diseases
    - *Rice Morphology*
  - There is a need for rice disease classification
    - *Udayananda et. al. & Sethy et. al.*
  - Early classification can help in prevention
    - *Pascual et. al.*

- ## Theme#4: Image Processing
  - Refers to obtaining valuable information and enhancement of the image
    - *Ramanjot et. al.*
  - In **Rice Leaf Disease** it could help in classification - *Sethy et. al.*
    1. It can help in recognizing the disease part
    2. Measuring infected region
    3. Area and shape of infection
    4. Color of infected area
  - ### Pre-processing
    - To reduce commotion, essentially enhancing the images.
      - *Vijai and Misra*
  - ### Segmentation
    - To obtain the object of interest
      - *Vijai and Misra*
  - ### Extraction
    - Can be utilized to identify the unhealthy part of the disease
      - *Vijai and Misra*
    - Obtains the characteristic of the data

Topic	Author	Methodology	Result
Rice Leaf Disease	Pothen et. al.	SVM (HOG & LBP as Extractor) (Resizing as Pre-processing) (Otsu's thresholding as Segment)	94.6%
Data Augmentation on grape leaf	Liu, Bin et. al.	Xception & GAN for data augmentation (Pre-processing)	98.70%
Sugarcane leaf disease	Padilla et. al.	Linear SVM & NDVI(Pre-processing)	86%
Plant Leaf Disease Detection	Vijai & Misra	Clipping & Smoothing (Pre) (Thresholding as Segmentation) & SVM	95%
Cucumber Leaf Disease Recognition	Hussain et. al.	VGG19 & Exception Transfer Learning (Extract) WOA(FS)	95.2%

Feature extraction	Segmentation	Pre-processing	Survey
GLCM & Color	n/a	Resizing & Enhancement, Denoising, Cropping	Ramanjot et. al.
Color	Fuzzy C mean & Clustering	Color Enhancemnt & Gaussian Filtering	Sethy et. al.

- # Theme#5: Image Classification
  - The goal of classification is to group images into cluster or classes
    - Corominas

Topic	Author	Methodology	Result
Leaf Disease Detection	Zamani et. al.	SVM RBF-SVM ID3 RF	RBF-SVM has the highest accuracy (unspecified)
Rice Leaf Disease	Ahmed et. al.	LR KNN DT Naive Bayes & Correlation Based Feature Selection	KNN w/ 98% accuracy
Plant Leaf Disease Detection	Yogeshwari et. al.	(Extract): GLCM (Segment): Thresholding & Fuzzy C Means PCA for Dimensionality Reduction Deep-CNN	97.43%

- # Theme#6: Machine Learning Techniques
  - ## Classical Approach
    - Classical Approach rely on methods of manual design that can only extract underlying features requiring expertise on subject matter
    - Inferior performance over deep learning
    - Requires manual design that extract the underlying features requiring subject matter expertise
      - Liu & Wang
    - Known for cheaper computation cost while having significant result
      - Sujatha et. al.

- Highest performing plant disease detection model
  - *Shrivastava et. al.*
- Overfits due to proportional growth of features
  - *IBM*

- **Deep Learning Approach**
  - Neural networks are used for data analysis and feature learning
  - High end system requirements
  - High Computational Power
  - Large amount of dataset
  - Poor Interpretability
    - *Liu & Wang*
  - Higher accuracy & Automatic feature learning in the study of *Ramesh & Vaydeki*
  - Unlikely to outperform if dataset lacks quality
    - *Tyagi & Rekha*
  - Lack of resources and poor connectivity of local farmers
    - *Manalo et. al.*
  - Financial setbacks limits the application of artificial intelligence
  - Lack of access to quality dataset may lead to limited performance
    - *Rosales et. al.*
  - ***Implies that Deep Learning may not be applicable to Philippines***
  - ***Lack of interpretability will make it difficult for the farmer to explain how the tool came up with a solution***
- **Improving Classical Approach**
  - Further work can be done by applying various algorithms and improving features
    - *Radovanovic & Dukanovic*
  - Feature Selection is one way to reduce dimensionality maintaining or even improving traditional machine learning by generating an optimal subset of features
  - A wrapper approach is appropriate as they attain general better classification performance as it considers the relationship between the subset, class, and classifier
    - *Nguyen*
  - Feature Selection reduces the number of features
    - *Seays et. al.*

Theme#7: Evolutionary Computation: Feature Selection

- ~~Evolutionary Computation Definition~~
- **Swarm Optimization Techniques**
  - Swarm Optimization techniques leverages swarm intelligence to find solution close to optimal.
  - **PSO**
    - ~~Definition~~
    - Few adjustable parameters
    - May become trapped in local optima
      - *Gad*
  - **ABC**
    - Mimics the behavior of bees in locating food sources, engaging in tasks and to remain unengaged.
    - May get trapped in local optima
      - *Andrushia and Patricia*
  - **ACO**
    - ~~Definition~~
    - Powerful search capability & Rapid convergence
      - *Nayer et. al.*
    - Performant in terms of speed and accuracy

- Ants may converge to suboptimal solution and not explore new solutions
  - *Peng et. al.*
- ## Synthesis of the Study

RD

- ## Quasi-Experimental
  - Dataset will not be randomly distributed into the models
  - Dataset will be split randomly into three sets: Training, Validation, Testing
  - These three sets will be consistent throughout the experimentation
  - Independent = Feature Selectors
  - Dependent = Metrics
- ## Waterfall SLDC
- ## Instrument
  - Experimentation Paper

SA

- ## Process
  1. **Image Acquisition**
     : Images are obtained from Kaggle
  2. **Pre-processing**
     : Images are enhance to get relevant data features
  3. **Extraction**
     : Involves selecting and combining variables from large datasets to identify most relevant features which helps in reducing overfitting
     - *Chatterjee*
  4. **Selection**
     : Selects the most relevant data which can enhance the accuracy of the tool making it easier for classifier to perform effectively
     - *Brownlee*
  5. **Classification**
     : Traditional method is used mainly because SVM is recommended in plant disease due to its high accuracy
     - *Gyanesh*
  6. **Diagnosis**
     : Will show the result of the process which is the classified rice disease

---

SOD

- ## Kaggle
  - ~~Healthy~~
  - Diseases
  - Will be confirmed by expert
- ## ~~IRRI~~
- ## Division

1. Training
   : For training the classifier
2. Validation
   : For tuning Hyperparameters
3. Testing
   : For experimentation and Data generation

DG/DG

- ## Preparation
- ## Experimentation
  - The tool will be utilized for the experimentation
  - Each of the model will be selected and tested with the testing set
  - Results will be compiled and gathered in an experimentation paper
- ## Analyzation

DA

- ## Terms
  - **True Positive**
    - Positive Observation, Positive Result
  - **True Negative**
    - Negative Observation, Negative Result
  - **False Positive**
    - Negative Observation, Positive Result
  - **False Negative**
    - Positive observation, Negative Result
- ## F-Score
  - Combines Precision and Recall
- ## Recall
  - Identifies the positive instances among all actual positive instances
- ## Precision
  - Ratio of correct positive predictions
- ## Accuracy
  - Overall correctness
- ## TNR
  - ~~Definition~~
- ## TPR
  - ~~Definition~~
- ## ANOVA
  - Analyzes the major difference between the techniques as the independent variable and metrics as dependent variable