DETECTION OF ONION LEAF DISEASE USING HYBRIDIZED FEATURE EXTRACTION AND FEATURE SELECTION APPROACH

1st George Oludare Gbadebo
Department of Computer Science
Federal University of Technology
Minna, Nigeria
debourdagr8t@gmail.com

2nd John Kolo Alhassan
Department of Computer Science
Federal University *of Technology*
Minna, Nigeria
jkalhassan@futminna.edu.ng

3rd Oluwaseun A. Ojerinde
Department of Computer Science
Federal University *of Technology*
Minna, Nigeria
o.ojerinde@futminna.edu.ng

2022 5th Information Technology for Education and Development (ITED) | 978-1-6654-9370-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/ITED56637.2022.10051500

Abstract*— Onion (*Allium Cepa*) is one of the most important vegetable and commercial plants that is being grown all around the world for more than 3000 years. Just like several other crop plants, Onion plants too can be attacked by pests and diseases of various kind, this attacks do give rise to low yields, bad quality and of course shortages of this important plants. Visual observation and analysis for detection of onion leaf diseases, if handed over to computing, using Machine Learning techniques, is more efficient, fast, cost saving, consistent, more reliable and highly accurate compare to what any human disease-expert eyes can offer. This work makes use of the prepared datasets of onion leaf digital images, after image preprocessing, some features were extracted/selected using Grey Level Co-occurrence Matrix (GLCM) and Particle Swarm Optimization (PSO) algorithms, the selected/extracted features then fed into classifier algorithms for eventual classification into healthy or unhealthy onion leaf.

Keywords—Leaf Diseases, Onion leaf diseases, Machine Learning, Feature Extraction, Feature Selection.

I. INTRODUCTION

Around one third of the world's agricultural crops are being destroyed by harmful factors. These harms to plants can either come from animals or other pathogens, putting together these harmful factors and if left uncheck, will in no time cut short the food and other farm products supplies to world population. The damages being caused to the plants is estimated to be from 1600 out of 80,000 to 100,000 diseases, 10,000 out of 80,000 insects and almost 30,000 weeds and other harmful substances (viruses, bacteria, nematodes) [1]. Onion plants particularly, have suffered some level of setbacks in its yields, qualities and quantities, which is majorly as a result of diseases encouraged by global earth warming and other related factors [2]. Onion remains one of the major vegetables both as food and for various health remedies applications, but earlier global studies have thus far shown the falls in yield and production of onion plants globally [3]. Applications of Machine Learning (ML) approach in farming systems, precisely in the area of Onion Leaf Diseases Detection, is what this paper looked into. GLCM is very helpful in the identification of different region of image it is applied on and can reveal certain properties about spatial distribution of gray level in texture image. Particle Swarm Optimization, is a swarm intelligence based numerical optimization algorithm capable of solving complex mathematical problems and it also helps find out optimal features based on their high discrimination [4]. Approach here is to combine an optimizing capability of PSO with GLCM, into designing a model for improved digital image feature extraction and aid improved classification eventually.

II. RELATED STUDIES

Proposed in the work of[5], based on the plant leaf disease classification using data from digital image processing, a direct image acquisition using digital image of the leaf plants acquired with the help of good resolution digital camera, then a bit of preprocessing was carried out of the image using color conversion, after which image segmentation was done with aid of K-means clustering, features extraction carried out using Gray Level Co-occurrence Matrix (GLCM) and finally classification of the disease carried out using a variant of SVM. [6] develop a real time disease monitoring system to monitor downy mildew onion disease, where the system is designed to monitor the onion field and acquire images of leafs with digital camera while another part of the system analyze the image using image binarization through numerical thresholding so as to accurately determined the diseased area of the leaf. [7] proposed a method that make use of Artificial Neural Network (ANN) as classification model in Red Onion disease detection. Then final disease classification done by ANN, the acquired data image went through preprocessing steps of image classification and feature extraction using Gray Level Co-occurrence Matrix (GLCM) and the method was able to attain performance accuracy of 91.1%. [8] work was to automate downy mildew disease symptoms detection in Onion plants. Image-based field monitoring system is expected to catch the image and Convolution Neural Network algorithm model ran as classifier for disease detection. The automation performance accuracy attain percentage that ranges between 74.2% - 87.1%, on the onion downy mildew the study is based. [3], [9]–[10], these works in their respective ways look into the worldwide importance, cultivation, yield and production quantities of Onion plants. Improved image features extraction/selection method is promised in this work through the combination of Grey Level Co-occurrence Matrix and Particle Swarm Optimization.

III. METHODOLOGY

A. Proposed Approach and Workflow

The objective of the method proposed in this work, as can be seen in the flow of Research Framework in fig 1, this is the combination of multiple algorithms for feature extraction/selection process, so as to come up with quality digital data which when fed into classifier of choice, it is expected to improve the classifier's performance. This is because it is believed that with better quality extracted features, subsequently better quality of feature can be selected out of this extracted features. The resulted final feature from the extracted/selected feature will be finest data to be fed into the classifier and this in turn is expected to have better

Authorized licensed use limited to: Hohai University Library. Downloaded on February 16,2024 at 22:22:56 UTC from IEEE Xplore. Restrictions apply.

performance as compared to any other process not conducted as such.

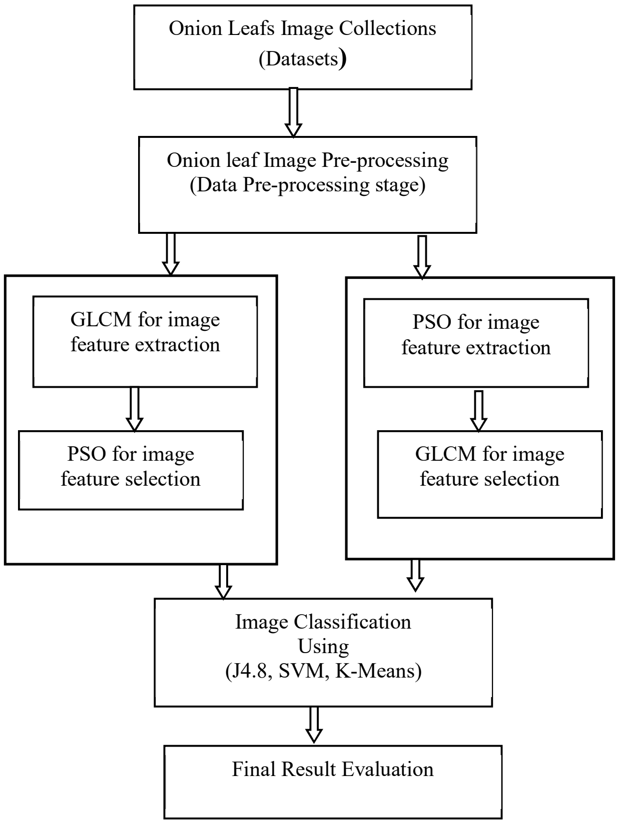


Fig. 1 Research Framework

B. *Data Collection*

As earlier indicated, very little research works were found in the area of Onion leafs diseases detection using machine learning, as compared to some other plants or fruits. Hence, for the purpose of this work, random images of diseased and healthy onion leafs were collected from the internet sites, a handful other snapshots of onion leaf images were also gotten from a Fadama farm in Niger State, Nigeria, with the aid of digital camera, thorough processing were carried out on the image data collected whereby making it suitable for the work. Fig 2 are some of the samples of the onion leaf images gotten to carry out the experiment.

Fig. 2. Onion Leafs Sample images of Healthy and Unhealthy

C. *Preprocessing and Analysis Phase*

This stage involve getting the images that were randomly taken/collected, and this involve things like image resampling, quantization or mapping, removal of data that are considered as noise, a level of filtering was also performed on the images, other necessary pixel corrections made and all images were converted to what is regarded as four color channel image format, being the format on which to continue further analysis. Four channel image here refers to Red, Green, Blue color band and the forth being the Brightness level (Gray).

Suggested for better quality feature extraction here, is Gray level co-occurrence matrix GLCM, while Particle Swarm Optimization (PSO) is employed to optimize the performance of the GLCM. PSO will work best in feature identification, especially on the nature of the datasets available for this research work and possibility is high that GLCM is going to work at its maximum, when its feed is taken up from PSO.

1) *Gray Level Co-occurrence Matrix (GLCM):* GLCM used second-order statistical method of Feature Extraction and it is quite efficient for use in variety of digital image analysis application. Feature Extraction is known to be a process by which useful information is extracted from body of raw data mostly for further analysis towards knowledge building. There has been records of use in image processing and analysis in areas which include but not limited to, biomedical, textile texture analysis, remote sensing, and defect detection purposes in industrial systems. GLCM is one of the best algorithms when it comes to extracting information from data, an important element in feature extraction is Texture, and this is visual pattern that are composed of spatially arranged entities that have characteristic brightness, size, color and shape. In calculating features from texture, the approaches used are, Model, Structural, Filter and Statistical approaches, one of the best and most commonly used of these approaches appear to be Statistical approach, example of which is GLCM. GLCM has two major parameters which form the basis upon which the texture properties of digital images are calculated, these two parameters are; distance and direction. Various combinations of these two parameters are used in calculating the texture properties like: Contrast, Correlation, Homogeneity and Energy. Digital image is the primary source of data for the GLCM model that is up for use in this work, it is from the obtained Onion leaf digital image the model will extract the needed image feature for use.

2) *Analysis for Particle Swarm Optimization:* Particle Swarm Optimization (PSO) is said to be a member of the swarm intelligence algorithm, while swarm algorithms themselves are based on the collective natural behavior of decentralized, self-organized systems of social animals like human being, birds, bees and many such animals that coordinate their actions, based on some guiding principles, towards achieving a goal.

PSO has quite a record of use in the area of feature selection/extraction as well as used to optimize some of other methods used in the area as well as other areas. PSO present a level of simplicity and somewhat a simple framework for solving continuous optimization problems and it has promising feat to ease work. PSO has proven to be great technique that ease computational complexity and aid accuracy.

3) *Implementation of the Proposed Method:* Gray level co-occurrence (GLCM) and Particle Swarm Optimization (PSO) algorithms were combined in the implementation, one is used to extract feature and fed into the other for feature

Authorized licensed use limited to: Hohai University Library. Downloaded on February 16,2024 at 22:22:56 UTC from IEEE Xplore. Restrictions apply.

selection. The two algorithms were in this manner used interchangeably before feeding the resulting data into classifier of choice. The flow of process is shown in the framework there in fig 1. After the reverse use of both algorithms, the final data from interchange combination of GLCM and PSO were fed into selected classifiers and the end results compared. The algorithm that follows, indicates the how the GLCM and PSO pick up each data (d) in body of Datasets (Ds), to produce Selected feature (Sf) for final classification.

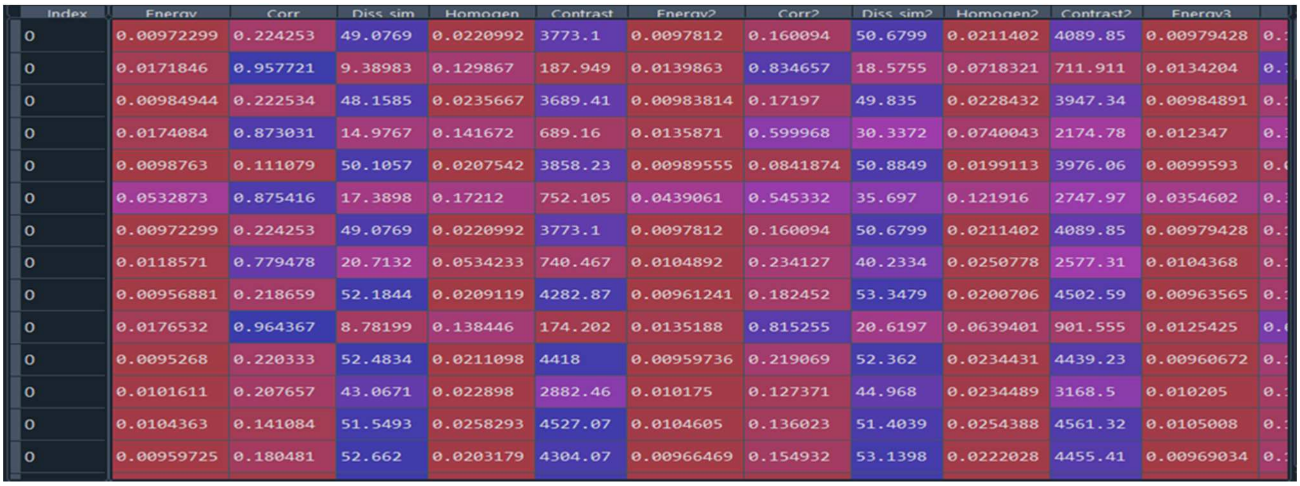
Algorithm 1. Extraction and Selection Implementation

```
// dataset = Ds
// single data in the dataset = d
// Selected feature = Sf
Start
Import Ds
Preprocess Ds
for all ( d in Ds) do:
    use GLCM for feature extraction
    use PSO for feature selection from resulting
    extraction
    use Sf for classification
end for
for all( d in Ds) do :
    use PSO for feature extraction
    use GLCM for feature selection from resulting
    extraction
    use Sf for classification
end for
Stop
```

IV. RESULTS AND DISCUSSION

This section discuss and show how the entire process of the two algorithms used, as well as what the outcome looks like. Gray level co-occurrence(GLCM) and Particle Swarm Optimization (PSO) algorithms were implemented interchangeably, that is, one is used to extract feature and fed into the other for feature selection (and vice versa), before feeding the resulting data into classifier of choice for classification and the end results compared

Fig 3. shown, is the sample of the features extracted based on selected indices.



Index	Energy	Corr	Diss sim	Homogen	Contrast	Energy2	Corr2	Diss sim2	Homogen2	Contrast2	Energy3
0	0.00972299	0.224253	49.0769	0.0220992	3773.1	0.0097812	0.160094	50.6799	0.0211402	4089.85	0.00979428
0	0.0171846	0.957721	9.38983	0.129867	187.949	0.0139863	0.834657	18.5755	0.0718321	711.911	0.0134204
0	0.00984944	0.222534	48.1585	0.0235667	3689.41	0.00983814	0.17197	49.835	0.0228432	3947.34	0.00984891
0	0.0174084	0.873031	14.9767	0.141672	689.16	0.0135871	0.599968	30.3372	0.0740043	2174.78	0.012347
0	0.0098763	0.111079	50.1057	0.0207542	3858.23	0.00989555	0.0841874	50.8849	0.0199113	3976.06	0.0099593
0	0.0532873	0.875416	17.3898	0.17212	752.105	0.0439061	0.545332	35.697	0.121916	2747.97	0.0354602
0	0.00972299	0.224253	49.0769	0.0220992	3773.1	0.0097812	0.160094	50.6799	0.0211402	4089.85	0.00979428
0	0.0118571	0.779478	20.7132	0.0534233	740.467	0.0104892	0.234127	40.2334	0.0250778	2577.31	0.0104368
0	0.00956881	0.218659	52.1844	0.0209119	4282.87	0.00961241	0.182452	53.3479	0.0200706	4502.59	0.00963565
0	0.0176532	0.964367	8.78199	0.138446	174.202	0.0135188	0.815255	20.6197	0.0639401	901.555	0.0125425
0	0.0095268	0.220333	52.4834	0.0211098	4418	0.00959736	0.219069	52.362	0.0234431	4439.23	0.00960672
0	0.0101611	0.207657	43.0671	0.022898	2882.46	0.010175	0.127371	44.968	0.0234489	3168.5	0.010205
0	0.0104363	0.141084	51.5493	0.0258293	4527.07	0.0104605	0.136023	51.4039	0.0254388	4561.32	0.0105008
0	0.00959725	0.180481	52.662	0.0203179	4304.07	0.00966469	0.154932	53.1398	0.0222028	4455.41	0.00969034

Fig 3. Sample of the Extracted Features

A. Experimental Results

1) Implementation: GLCM Extracts Feature and resulted Data fed into PSO for Feature Selection, then classified

Presented here are results obtained when GLCM algorithm is used first, for feature extraction, the resulted dataset has feature selection carried out, using PSO algorithm, and classifiers then ran on the final result. It is important to specify here that, the same datasets were used for the two implementations. The results gotten from one of implementations are as presented in tables and figs here in this section.

Table 1 shows the values of different evaluation matrices the classifier algorithms adopted to present final results. The classifier with the highest overall accuracy value is J4.8 with 100%, follow by SVM and K-means with both having 95.5% in overall accuracy.

Table 1 GLCM into PSO for Feature Extraction and Selection respectively

Classifier	Label	Precision (%)	Recall (%)	F1-Score (%)	G-mean (%)	Accuracy (%)
Decision Tree (J4.8)	**Healthy**	100	100	100	100	**100**
	Unhealthy	100	100	100		
SVM	**Healthy**	91.7	100	95.7	95.5	95.5
	Unhealthy	100	90.9	95.2		
K-means	**Healthy**	90.9	100	92.2	95.8	95.5
	Unhealthy	100	91.7	95.7		

The figs that follows here, reveal distributions of confusion matrix according to values of classifiers during the test phase. Fig 4. is the confusion matrix for the J4.8 (Decision Tree variant) algorithm, and it shows the numbers that were correctly classified and the ones that are wrongly placed. As shown in Fig 4, and from datasets used during the test process of the Decision Tree classifier, 22 healthy and 21 unhealthy onion leafs were all correctly classified and none was

Authorized licensed use limited to: Hohai University Library. Downloaded on February 16,2024 at 22:22:56 UTC from IEEE Xplore. Restrictions apply.

misclassified. Basically, what that referred to is in some cases during the classification stage of some works, depending on how accurate the performance of the model or method used, some healthy samples may mistakenly be classified as unhealthy while unhealthy ones be taken for healthy, but this work is shown to demonstrate high reliability.

Fig 4. Decision Tree Confusion Matrix for GLCM into PSO

SVM and k-means classifiers used, also has their confusion matrix distributions shown in Fig 5. From what can be observed in Fig 5a, healthy leafs classified correctly are 22, and 20 being number of unhealthy leafs, also correctly classified, while only 2 was wrongly classified. A close observation of k-means confusion matrix presented in Fig 5b shows similarities in distribution compare with SVM, it is said to be similar in distribution because it can as well be observed that the number of the healthy onion leafs classified correctly as being healthy as shown in the fig, are 22, just as the number of the leaf therein the section of the leafs classified as unhealthy are in their case are 20 of them, and as well as the number classified wrongly were 2, as it was in SVM, k-means too has the number of leaf it classified wrongly to stand at 2. The occurrence displayed in both of the SVM and K-means were not too surprising, considering that even in the overall accuracy, SVM classifier has considerably high percentage which happens to be 95.5% and K-means classifier was able to pull same 95.5% accuracy. The 95.5% overall value of accuracy is an interpretation that the other indices for measuring confusion matrix, both (SVM and K-means) algorithms surely have approximately same values in all of them.

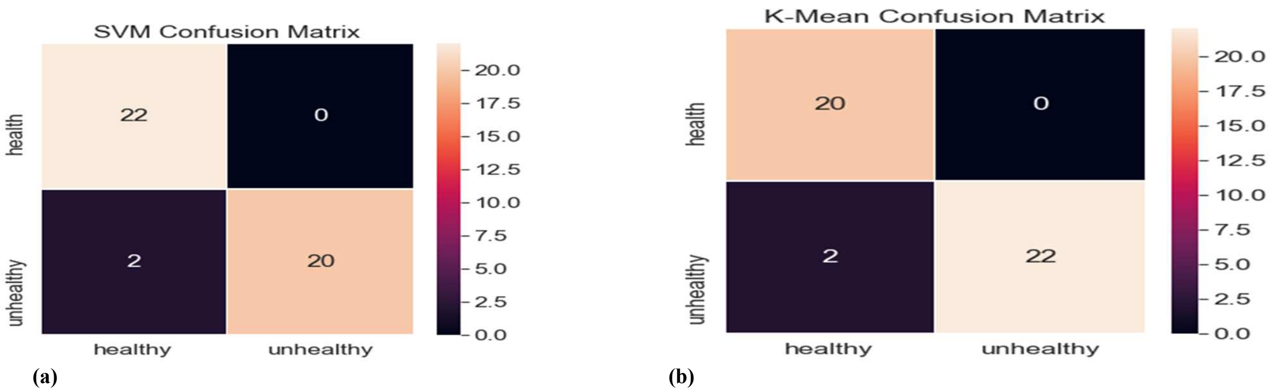


Fig. 5. SVM and K-means Confusion Matrix for GLCM into PSO

B. Discussion of Results

Although, very little work was found in onion plants leaf disease detection so far. Apparently, from the works of [4] - [11], which are more of expert system-based diseases detection methods, the highest possible accuracy in the works combine is just 67%. Meanwhile, in the works of [6] –[8], the ones that are even computer vision and machine learning inclined, yet the highest accuracy that was obtained in all the works is 91.1%. From all obvious indications and in comparison, the proposed technique in this research work performed better. Table 2. gives summarized overview of values of the results produced from this work in comparison with the results of other techniques available to us in other works.

Table 2 Accuracy Values

Technique	PSO – GLCM (*implementation 1*)			GLCM – PSO (*Implementation 2*)			Expert System	GLCM	N/A
Classifier	**J4.8**	**SVM**	**K-Means**	**J4.8**	**SVM**	**K-Means**	**ES**	**ANN**	**DNN**
Accuracy	95.3%	93.2%	86.4%	**100%**	95.5%	95.5%	67%	91.1%	87.2%

Both implementation 1 and 2 in this work, which refer to the proposed technique. It can be seen that among the three classifier algorithms used; Decision Tree (J4.8) has the highest level of accuracy, with the top best of accuracy result occurred in the implementation 2, which makes use of GLCM for feature extraction and PSO for feature selection before feeding into classifier, here J4.8 classifier got us 100% accuracy as top best. As noticed, the technique proposed in this work for extraction/selection can be seen to aid better performance of the final classifier in virtually all instances, except for in the implementation 1 (PSO – GLCM), where K-means classifier algorithm has lower accuracy value of 86.4%.

Fig 6 gives the accuracy comparisons in a vivid glance, where all the accuracy levels can be clearly seen in the bar chart, as the horizontal values represent the implemented techniques that was made use of while the vertical values represent the values of accuracies obtained in the techniques. The technique proposed in this work can be seen to have consistence and conspicuous performance over other works reviewed in the onion leaves diseases detection categories that was reviewed for this work at hand. The method proposed here shown considerably level of supremacy that cannot be overlooked when it comes to early detection of onion disease which many species of onion species do suffer from.

Authorized licensed use limited to: Hohai University Library. Downloaded on February 16,2024 at 22:22:56 UTC from IEEE Xplore. Restrictions apply.

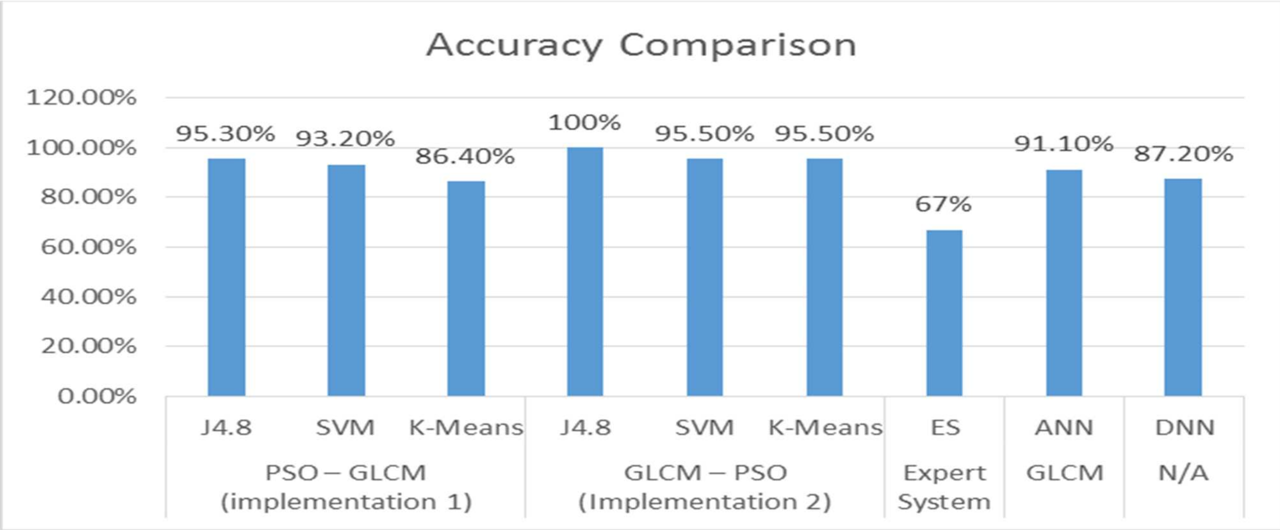


Fig 6. Chart representation of Experimental Accuracies Comparison

From the work of [7], just GLCM was used for image feature extraction. Meanwhile, this research work went a step further to combine GLCM and PSO. Both PSO and GLCM were then used interchangeably for feature extraction/selection, whereby result of one was fed into the other for more refined output, evaluating different grey-levels, where the most efficient output were used for final implementation. The result gotten from this work showed higher degree of accuracy when compare to the works that have been seen so far in the area of onion leaf disease detection. Expert System-based work of [12] and algorithm based works of [6], [8] are other instances for comparisons, and none of these works has percentage of accuracy as high as technique this research work proposed.

C. *Model Deployement*

This model will consist of real time surveillance system comprising of high definition and precision camera, this shall serve as data collection arm, which is meant to take the images of onion plants at intervals and store in the database unit where collected datasets awaits next stage. Another unit of the model responsible for preprocessing of the onion leafs images is preprocessing expert unit, the process here shall include image logging, image cropping, mapping or quantization, this unit of the model shall pick the data from the database, process them and send to the next stage. The last unit of the model is the server side where the GLCM-PSO algorithm module for extraction/selection resides, as well as classification algorithms module, responsible for final classification of collected datasets of onion leafs into healthy or unhealthy. The analysis of the output from the last stage of the model will determine healthy or unhealthy status of the leaf and the percentage will dictates whether it is outright removal or treatment of the affected leafs.

V. CONCLUSION

Conventional onion leaves diseases detection process relying merely on visual observations will most definitely be laden with errors, slow and unreliable. This work has come up with experimentally proven technique on how to improve on the process using Machine Learning, so as to make the process faster, cheaper and with improved accuracy. This work promised achieving such feat with proposed method which involves collection of digital images of onion leaf with the aid of digital camera and preprocess the image to make it suitable for feature extraction/selection. The proposed technique involve hybridization, whereby two feature extraction/selection algorithms (GLCM and PSO) are combined and used, so as to get high quality features for the classifiers. This proposed approach automatically select most useful features for classification accuracies, the method make use of three different classifiers namely; J4.8 (a variant of Decision Tree algorithm), SVM and K-means, for the final classification of onion leaves into healthy and unhealthy categories. The resulted data from feature extraction/selection was differently fed into the classifiers and the outcome improvement was impressive compare to the past works in the area of onion leaf disease detection.

REFERENCES

[1] M. Dianingrum, N. Hermanto, and M. Iqbal, "Expert System for Simulation of Pest and Disease Diagnosis in Onion Plant Using Putty Shafer Method and Rule-Based Approach," vol. 2, no. 1, pp. 1–8, 2019.

[2] F. Hanci, "A Comprehensive Overview of Onion Production : Worldwide and Turkey," vol. 11, no. 9, pp. 17–27, 2018, doi: 10.9790/2380-1109011727.

[3] V. John, S. D. Simon, A. K. Maurya, and A. A. Lal, "Survey of Purple Blotch Disease of Onion (Alternaria porri) of Allahabad District, India," no. October, 2018, doi: 10.20546/ijcmas.2018.710.009.

[4] K. Karadağ, M. E. Tenekeci, R. Taşaltın, and A. Bilgili, "Detection of pepper fusarium disease using machine learning algorithms based on spectral reflectance," *Sustain. Comput. Informatics Syst.*, vol. 28, no. 2018, 2020, doi: 10.1016/j.suscom.2019.01.001.

[5] M. A. Khan, M. I. Lali, and M. Sharif, "An Optimized Method for Segmentation and Classification of Apple Diseases based on Strong Correlation and Genetic Algorithm based Feature Selection," *IEEE Access*, vol. PP, no. c, p. 1, 2019, doi: 10.1109/ACCESS.2019.2908040.

[6] D. Kim, K. Lee, C. Choi, T. Choi, and Y. Kim, "Development of real-time onion disease monitoring system using image acquisition," vol. 5, no. 4, pp. 469–474, 2018.

[7] W. Kim, D. Lee, and Y. Kim, "Machine vision-based automatic disease symptom detection of onion downy mildew," *Comput. Electron. Agric.*, no. July, p. 105099, 2019, doi: 10.1016/j.compag.2019.105099.

Authorized licensed use limited to: Hohai University Library. Downloaded on February 16,2024 at 22:22:56 UTC from IEEE Xplore. Restrictions apply.

[8] W. Kim, D. Lee, and Y. Kim, "Machine vision-based automatic disease symptom detection of onion downy mildew," *Comput. Electron. Agric.*, no. November, p. 105099, 2019, doi: 10.1016/j.compag.2019.105099.

[9] P. R. Meena, G. P. Saraswathy, G. Ramalakshmi, K. H. Mangaleswari, and T. Kaviya, "Detection of leaf diseases and classification using digital image processing," *Proc. 2017 Int. Conf. Innov. Information, Embed. Commun. Syst. ICIIECS 2017*, vol. 2018-Janua, pp. 1–4, 2018, doi: 10.1109/ICIIECS.2017.8275915.

[10] E. E. Pfeufer, B. K. Gugino, P. Pathology, E. Microbiology, and T. Pennsylvania, "Environmental and Management Factors Associated with Bacterial Diseases of Onion in Pennsylvania," vol. 16802, no. November, pp. 2205–2211, 2018, doi: 10.1094/PDIS-11-17-1703-RE.

[11] D. Sitanggang, S. D. Siregar, S. M. F. Situmeang, and E. Indra, "Application of forwardchaining method to diagnosis of onion plant diseases," 2018.

[12] C. Ruth, "Screening And Evaluation Of Onion Varieties Against Fungal Diseases In Onion (Allium Cepa)," vol. 6, no. 5, pp. 135–140, 2017.

Authorized licensed use limited to: Hohai University Library. Downloaded on February 16,2024 at 22:22:56 UTC from IEEE Xplore. Restrictions apply.