

DNA Sequence Identification and Rectification to Intensify the Yield of Agricultural Crops

Janice Magdalene D¹, Charulatha R², Fredolin B Brone³, Darling Jemima D⁴

^{1,2,3}Students, Department of CSE, Sri Krishna College of Technology, Coimbatore, India
Corresponding Author Email: [janicemagdalene\[at\]gmail.com](mailto:janicemagdalene[at]gmail.com)

⁴Assistant Professor, Department of CSE, SKCT, Coimbatore, India

Abstract: Agriculture is the backbone of the Indian economy with a contribution of about 54.6% of Gross product. A specific proportion of it is affected by deadly plant disease like rice burst, wheat burst, strip burst, common root rot, etc. Plant diseases affect biological process like photosynthesis, plant growth, flower growth, fruit development, DNA, etc. which in turn affects the yield of crops. The objective of the work is to detect and rectify plant diseases using modern technologies like bioinformatics, genomics and datafication in order to improve the yield of crops. A new algorithm is developed to identify the defective DNA and hence for the rectification process.

Keywords: Bioinformatics, Datafication, Identification Algorithm, DNA, Genome-editing technology

1. Introduction

According to global Statistics, diseases in plants reduce the crop yield by 20-40%. In India 15-25% of the production is affected by plant diseases. So, detection and rectification of plant diseases plays an important role to increase the yield. In this study, we use genomics, genome-editing technology, bioinformatics, datafication to overcome the diseases caused by genes. Bioinformatic links the technologies from computer science and biology to help us understand and interpret biological data. Bioinformatic tool compares genetic and genomic data and helps for better understanding of molecular biology. Genome editing technology is used to rectify the defective genes of crops that has been studied from genomics to produce high yield of plants. "Robertson, S. and Walker introduced thresholding algorithm that has proven to be critical for classification [1]. According to his studies setting a threshold too high retrieves nothing at all but setting it too low retrieves too many. Genome editing technology is used to rectify the defective genes of crops that has been studied from genomics to produce high yield of plants [2].

The growth of the genome edited crops has to be monitored, tracked and analysed. This process is done by Datafication. "Plant Disease Detection Using Image Processing and Machine Learning" by Pranesh Kulkarni, et.al, detects several diseases of 5 common plants by using threshold algorithm [3]. Detection of unhealthy region of plant leaves using image processing and genetic algorithm is introduced by Vijai Singh, et.al, [4]. Genetic algorithm in image segmentation is employed for detecting leaf diseases. Sharada P. Mohanty et.al, used Deep Learning for Image-Based Plant Disease Detection[5]. Kangquan Yin and Jin-Long Qiu [6] introduced Genome editing for editing genes against bacteria.

In this study, an attempt has been made to identify the defective gene and its rectification through a new algorithm and the relevant efficiency is obtained. A secondary data set is obtained from existing source and the newly developed

CAD system is experimented. The new CAD system easily identifies the defective DNA and signals to rectification process.

2. Statistical Analysis

Figure 1 shows the state-wise soil health that has a significant effect on genetic integrity of crop. Figure 2 shows the proportion of the crops affected by factors like pests, plant disease, weeds, others. The Figure 3 shows the frequency of common symptoms in plant diseases. Source: Cooperative Extension Service University of Kentucky College of Agriculture, Food and Environment, Lexington, KY, 40546.

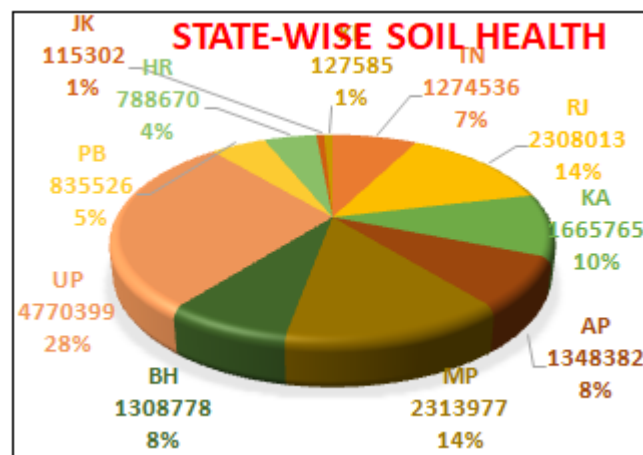


Figure 1: State-wise Soil Health record

Source: Pocket Book of Agricultural Statistics, 2017, Government of India.

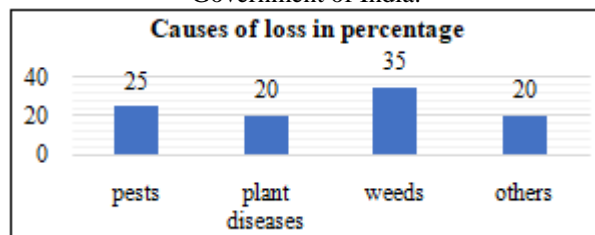


Figure 2: proportion of the crops affected by various factors

Table 1: Common Symptoms in Plant Diseases

Symptom	Description	Fungus	Water molds	Bacteria	Viruses
Blight	Discolour	1	1	1	0
Blotch	spots	1	0	1	0
Bronzing	Bronze colour	1	0	0	0
Canker	Dead region	1	1	1	0
Chlorosis	Yellowish	1	1	1	1
Decline	Plant Tissue	1	1	1	1
Distortation	Irregular shape	1	1	1	1
Gall	Abnormal size	1	0	1	0
Lesion	spots	1	1	1	1
wilt	Dropping of leaves	1	1	1	0

* 1 represents presence and 0 represents absence Source: Oerke, 2006

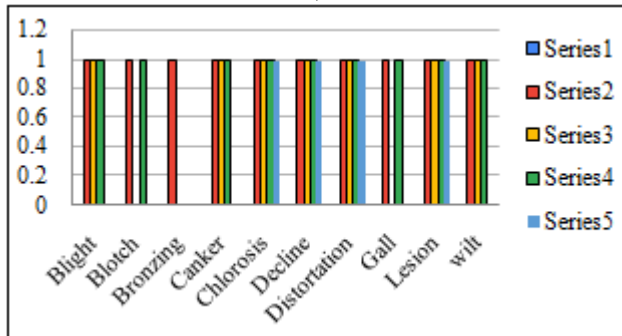


Figure 3: Frequency of common symptoms in plant diseases

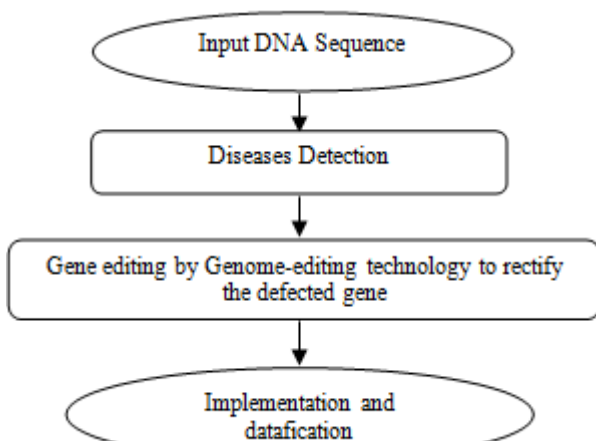


Figure 4: Process Evolution

Table 2: Number of defect DNA simulated over the dataset

Subgroup	1	2	3	4	5	6	7	8	9	10	11	12	Mean
Defects	0	1	0	2	1	0	1	4	0	1	0	0	0.833
UBW	3.57	3.57	3.57	3.6	3.57	3.6	3.6	3.57	3.57	3.57	3.57	3.57	

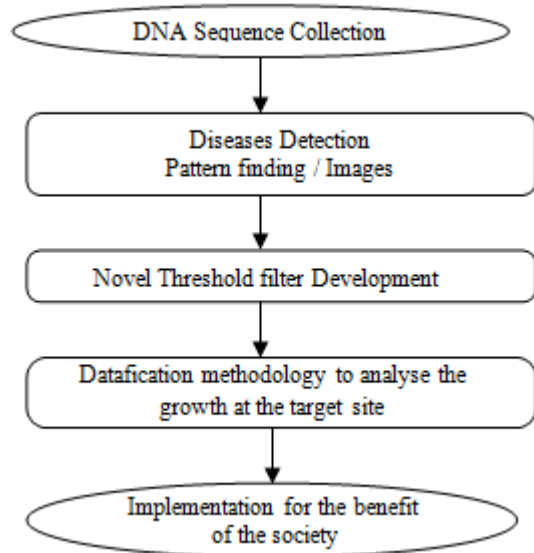


Figure 5: Flow of events to identify and rectify the DNA set

A Novel Algorithm to identify the number defects DNA.
 Step 1: The DNA dataset is categorized into k subgroups.
 Step 2: Let x be Number of defects in the DNA observed during collection of set.
 Step 3: The number of defects in the DNA are very rare and hence it is assumed to follow Poisson Law.
 Step 4: Using Machine Learning Tools average number of defects in the k subgroups and finally the overall mean defects are found.
 Step 5: Now the Computer Aided Diagnostics (CAD) which are newly developed is executed.
 Step 6: CAD determined the Upper Bandwidth and Lower Bandwidth region to identify the limits to monitor and control number of defects in the DNA set.
 Step 7: $UBW = \bar{x} + z_{\alpha} \sqrt{\bar{x}}$ and $LBW = \bar{x} - z_{\alpha} \sqrt{\bar{x}}$.
 Step 8: Now CAD monitors and determine the out of control region if any.
 Step 9: If any subgroup falls outside the control region then immediately alarm is raised by CAD.
 Step 10: The appropriate DNA set is identified and rectified before goes to farmers.

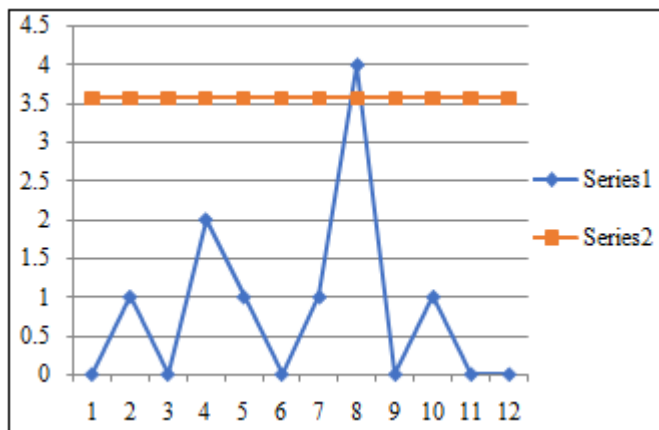


Figure 6: Monitoring Number of Defect DNA

Series1: Number of defect DNA is each subgroup

Series 2: UBW

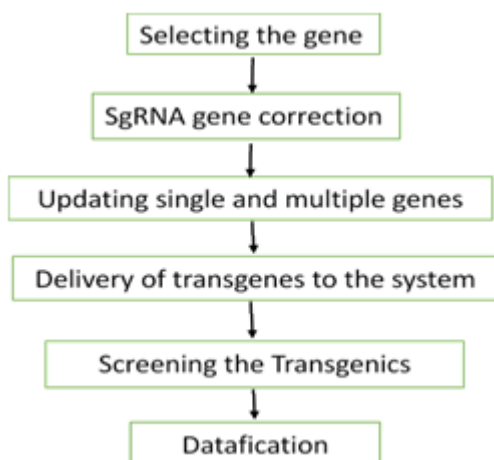


Figure 7: Flow of events for DNA rectification

Fig. 7 describes the DNA rectification process. The site is selected where the gene has to be corrected. Then expression of vector design is done by packing of sgRNA and Cas9 protein with suitable promoters. Then they are delivered into the targeted cells. Now, the screening and conformation of the transgenics is done. Finally further evaluation is done for abiotic stress tolerance and hence the rectification of DNA.

3. Conclusion

In this article a new algorithm is developed to identify the defective DNA set. Figure 6 shows the exact location of defective DNA. It is found that the 8th data set is out of UBW and one can easily identify the defective DNA. Hence the newly developed CAD system identifies the defective one and sends for rectification. Using genome-editing technology the DNA content is rectified in an efficient way.

Acknowledgement:

We thank our college Principal, Management and Staff for encouraging us to do this work.

References

- [1] Robertson, S. and Walker, S. (2000), 'Threshold setting in adaptive filtering', Journal of Documentation.

- [2] Ch Mizas et. al., 'Reconstruction of DNA sequences using genetic algorithms and cellular automata: towards mutation prediction', Biosystems,2008.
- [3] Vijai Singh, Varsha, A K Misra, "Detection of unhealthy region of plant leaves using image processing and genetic algorithm", 2015 International Conference on Advances in Computer Engineering and Applications.
- [4] Pranesh Kulkarni and et al, "Plant Disease Detection Using Image Processing and Machine Learning",2021 Arxiv.
- [5] Sharada P. Mohanty and et al, "Using Deep Learning for Image-Based Plant Disease Detection",2016Frontiers.
- [6] Kangquan Yin and Jin-Long Qiu, "Genome editing for editing genes against bacteria", 2019 National Library of Medicine National Centre for Biotechnology Information.