Bio-Info Cloud: Gadget for Generalization Flaw Discovery among Crops for Efficient Growth & Yield

Kumaravel .N

M.E Student, Dept. of C.S.E, Arunai Engineering College, Thiruvannamalai, India

ABSTRACT: Bioinformatics is a interdisciplinary field that develop methods and software tool for understanding biological data. It is a combination of computer science, statistics, mathematics and engineering to analyze and interpret biological data. Agriculture is the largest economic sector and plays a major role in socio-economic development of India. Agriculture in India, it is the means of livelihood of almost two thirds of the workforce. Crop Cultivation depends on rainfall, quality of soil and climatic conditions. Crop diseases will reduces both Quality and Quality of yield. A major challenging task in agriculture is past and pre-detection of disease affect among loops. This paper considers this Challenges and various proposed solution are given out with the inclusion of various algorithm and methods. And a information regarding the crops was given to the farmer periodically with an help of a emerging service providing technology. An cloud infrastructure was given with three major services which will monitor the activity of farmer and provide required information to increases the yield of the crops.

KEYWORDS: Bioinformatics, Quality of Yield, Image Processing, Predetection,Genome.

I. INTRODUCTION

Bioinformatics combine many fields such as computer science, statistics, mathematics, and engineering to analyze and interpret biological data. Bioinformatics has become an imperative part of many areas of biology. In experimental molecular biology, bioinformatics technique such as image processing and signal processing permit extraction of useful consequences from large amounts of raw data. Bioinformatics tools assist in the comparison of genetic and genomic data and more ordinarily in the understanding of evolutionary aspects of molecular biology. A major challenging task in agriculture is past and pre-detection of disease affect among loops. This paper considers this Challenges and various proposed solution are given out with the inclusion of various algorithm and methods.Image segmentation is an imperative topic in image processing task. It can produce the progress of the complete recognition which is becoming more and more significant in the agricultural automation. For instance, in the field of crop disease automatic recognition. According to crop leaf image, the disease type can be established with the image processing method and expert's knowledge. However, the segmentation's accuracy not satisfy the current practical needs. Thresholding, boundary-based, region-based, and hybrid techniques are used .Thresholding is the delegate method and depends on the facet map and threshold value. In the proposed work, we have focused on premature detection of fungal disease from visual symptoms .This plain-speaking work in the real world consists of the tasks like image acquisition,pre-procesing of image, feature selection, development of methodologies for identification of fungal disease symptoms pretentious on different agriculture/horticulture crops, and finally development of architecture for the computer vision systems. As a first step, image of fungal disease symptoms exaggerated on different agriculture /horticulture crops are collected from the department of plant pathology, university of agriculture science, dharwar summarizes some of the image processing methods proposed for fungal disease detection. In this Case Based Reasoning (smith- waterman) algorithm in bioinformatics and image processing is used. A line of attack in which knowledge and inferences are derived directly from historical cases/examples. Analogical reasoning smith-waterman algorithm, determining the outcome of a problem with the use of DNA sequence. A Procedure for drawing the conclusion about a problem by using the past experience directly. Inductive learning, A contraption learning approach in which rules are inferred from the historical data. Algorithm is a compilation of a set of database .It is based on human information (HIP) model in some problem areas. Thinking about how human process information.Remember preceding case/recall similar cases

&modify to fit a new situation. Human experts depend heavily on past experiences when solving new problems. Image processing is a indicative pattern closely related to the human way of reasoning and acting in daily situations when facing novel problems. bioinformatics uses old experiences to solve novel problems, based on the following sentence, known as the bioinformatics assumption: Similar problems have similar solutions. A Circumstances experienced in the way that it has been captured and erudite is referred to as *past/ previous case* and is stored in the case base. A latest situation asking for a solution forms the description of a *new/target case*. An important part of the canny edge detection and histogram methodology is its learning skill, which comes as a natural result of its problem-solving process: the case base is updated every time a new skill is obtained. This knowledge can be reused when needed without implementing the whole process from scratch or to accentuate a methodology that should be avoided in a similar situation. Therefore, case based reasoners are able to progress their problem-solving performance over time. The crops defect solving and knowledge process can be described as a cyclical process comprising four process.

• *retrieve* the most significant cases,

• *reuse* the appreciative provided to the new problem,

• *revise* the result obtained, and

• *retain* the parts of the new solution that are likely to be used for upcoming purposes. In adding together to the knowledge obtained from previous cases, there's also area-dependent knowledge supporting the system process.Cases can be viewed as collected of two parts: the Problem description and the trouble solution.

II. RELATED WORK

India is an agricultural country where in concerning 70% of the population depends on agriculture. Farmers have wide range of assortment to select suitable Fruit and Vegetable crops. However, the cultivation of the crops for best possible capitulate and quality produce is highly technical. It can be enhanced by the abet of technological support. The management of perennial fruit crops requires close monitor especially for the supervision of diseases that can affect production extensively and consequently the post-harvest life. It is estimated that 2014 plant disease losses in Georgia (UK) is approximately \$549.84 million. Of this amount, around 195 million USD was depleted on controlling the diseases, and the respite is the value of damage caused by the diseases. Automatic detection of plant diseases is an fundamental research topic as it may ascertain benefits in monitoring large fields of crops, and thus manually detect the symptoms of diseases as soon as they come into sight on plant leaves. Therefore looking for fast, automatic, less expensive and accurate method to perceive plant disease cases is of great reasonable significance .

III. PROPOSED SYSTEM

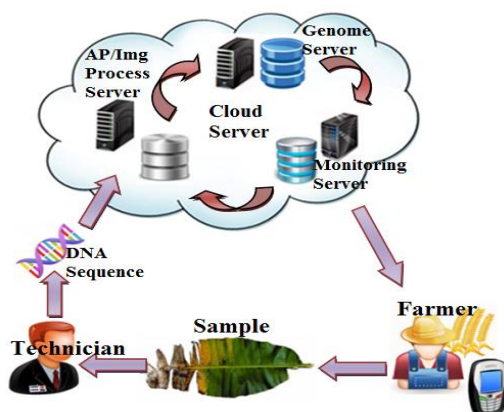In this paper five modules are used to describe the system architecture.

Fig 1.System Architecture

MODULES:
   • Soil minerals and nutrition identification.

- Crop weakness identification.
- Disease identification.
- Image comparison for diseases identification
- Information providing monitoring server .

SOIL MINERALS AND NUTRITION IDENTIFICATION:

In the soil, all mineral elements are combination of mass flow and diffusion, and all strategies designed to improve mineral gaining by crops affect one or both of these processes . Diffusion of mineral elements is determined by the concentration gradient between the root surface and soil solution , which are often dictate by interactions with soil mineral surfaces. It operate over tiny distances, and is especially important for the macronutrients P and K that often limit crop production . The heap flow of a mineral element to the root surface is determined by its concentration in the soil solution and the transpiration movement of water to the root surface. It is important for the attainment of indispensable mineral elements with large concentration in the soil solution, such as N, K, S, Ca, Mg and Cl in agricultural soil, and for mineral elements that are mandatory in fairly small quantities by plants, such as Fe, Mn, Zn, Cu, Ni, B and Mo. If essential mineral elements are not present, or not available, in the soil, then they must be supplied as fertilizers to facilitate crop production and guarantee produce quality. Many agronomic strategies can be adopt to increase the effectiveness with which inorganic and organic fertilizers are used to provide the essential mineral elements required by crops will optimize the chemistry, quantity, placement, and timing of fertilizer applications .
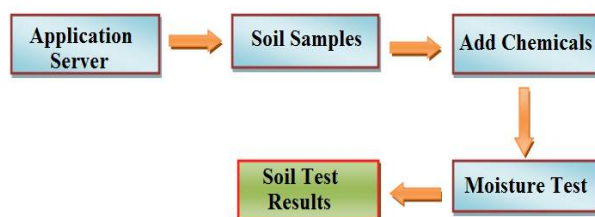
Fig1.a.Soil minerals and nutrients identification

CROP WEAKNESS IDENTIFICATION:
(1)Our main aim are all find the "similar" sequences in databases with query sequence.
(2)The definition of "similar" are approximately same. Blast normally reflect on the number same letters and gaps. Search engine also use the amount of same words, distance between the sequence to calculate such "similarity".
(3)Comparing very long sequence or multiple Comparing generally pertain agreed version algorithm based on basic method.
(4)They also have the similar evaluation metrics. Blast will apply Sensitivity and selectivity to estimate alignment algorithm. Correspondingly, search engine make use of recall and precision. Such evaluation metrics are almost same.
(5)The blast compares the sequence by letters, but search engine make use of words. Additionally, blast and related upgrading versions utilize position to index letters. Search engine index document, in general a short sequence. For long set of sequence, they divide the long text into short document. So the outcome of search engine are series of documents. Results of Blast are segments of sequences.But seek engine result show stature is usually the dynamic abstract of document, which is also the combination of match segments of result manuscript.
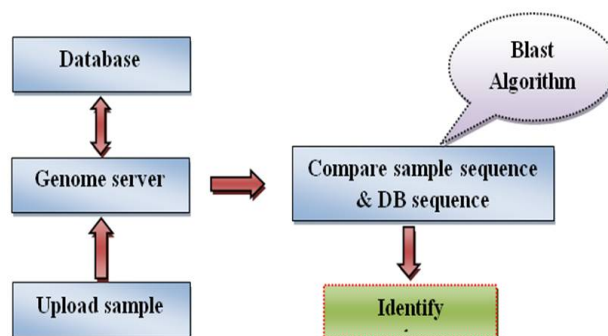
Fig 1.b.Crop Weakness Identification

DISEASE IDENTIFICATION WHOLE GENOME:

Genome analysis consists of at least four steps. We depict these four steps in . In the DNA sequencing step information preset in DNA molecules is made digitally clear by translating DNA macromolecules into strings of A's, C's,' G's, and T's that encode the nucleobases Adenine, Cytosine, Guanine, and Thymine, respectively. State-of-the-art DNA sequence are capable to sequence billions of bases in short time, which increase the total of genome data to analyse dramatically. DNA sequencers are not capable to sequence complete DNA molecules at once, but only in small, overlapping pieces. These DNA piece is called reads and have a small length of in relation to 100 base pairs that is rather small compared to the size of a whole genome such as the barley genome with 5 billion base pair. Sequentially to perform genome analyses, in a second step, DNA molecules must be restored after DNA sequencing from the short reads. After alignment, analysis of genomes starts. As human or plant genomes are enormous, only interesting sites in the genome are analysed at first. To discover such interesting sites difference between sample genomes and the use reference genome are determined and a variant is called in case a difference is reliable. Such differences can be base exchange at single genome sites, so called single nucleotide polymorphisms (SNPs), or insertions and deletions of base regarding the reference genome . At variant sites, further downstream analysis takes place that a lot require more data sources to determine the effect of a certain variant. For example, a mutation in a certain barley gene influences the number of spikes of a barley plant . Often, downstream analysis is support by revelation tools such as IGV that permit scientists to visually inspect and navigate through the genome.
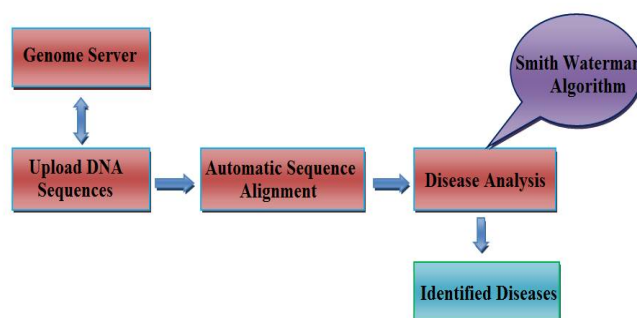

Fig.1.c.Disease Identification Whole Genome

IMAGE COMPARISON FOR DISEASES IDENTIFICATION:
Step 1: Read test image and database image.
Step 2: Apply Canny Edge Detection Algorithm on RGB image
Step 3: Retrieve Strong Edges of image (Healthy or Infected) Step 4: Plotted Edge Histogram.
Step 5: Plotted Color Histogram.
Step 6: Separate the layers of testing image into Red, Green and Blue Histogram.

Step 7: Calculate the difference in histograms of both the images (Testing and Database Images).
Step 8: Results shows the disease of plant by comparing the testing image with database image.
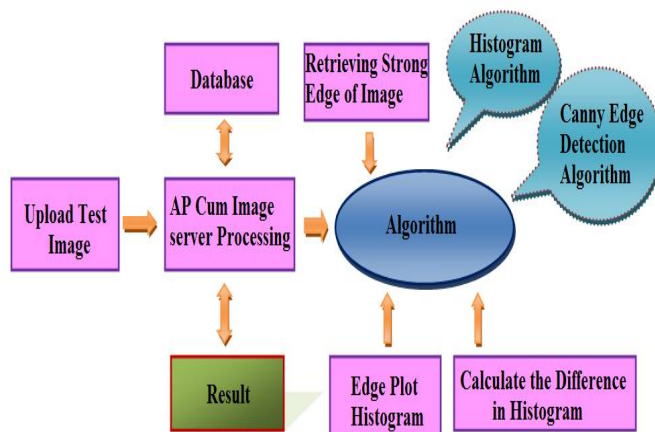

Fig.1.d.Image Comparison for Diseases Identification

INFORMATION PROVIDING MONITORING SERVER:

This server plays an important roles in the Agro-cloud because the related information regarding the cultivation was delivered to the farmer periodically based on the dynamic registration involved by them and the server connection was done by this server based on the past query, automated service prediction, the service may be of any cultivation related issues.
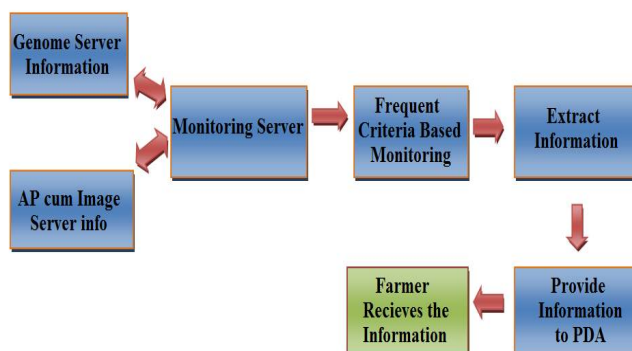

Fig1.e. Information Providing Monitoring Server

1) *Monitoring Service:*

Monitoring service is online state monitoring, which continuously tracks certain states of information according to a farmers query deployable within the cloud. The information from various servers like weather forecasting server, crop information server, server having an information about various plant disease etc., are periodically updated to cloud storage database by using this service. If monitoring services detect to send a notification to the farmer's PDA can be done with the help of an information Providence system.

2) *Information Providence:*

Provided information was related with new crops which may not be already known by the by the farmer and gives the knowledge of how new crops can be cultivated and what are all the mandatory steps involved was given out, pest related information (i.e.) the information regards common crop affecting disease and beyond the knowledge crop affecting diseases, notification of weather forecasting information and precautions for a natural hazards was delivered to the farmers to her personal assistance device.

In the previous few years soil testing, disease identification for a crop will done in separate areas. Due to this quality and quantity of the crop will be decreased. Interaction and knowledge of the farmer will not receive properly. Now a days, agriculture is became an endangered field. In order to improve the quality and quantity of the crop by combining two test (soil test, DNA sequence test).While doing the soil test, minerals and nutrients are present in the correct proportion will be checked. By this we can identify the quality of the soil.DNA sequence test will identify the disease in the crop at the early stage. We can retrieve this disease by DNA sequence alignment [Smith waterman algorithm, Blast algorithm].We can improve the quality of the crop by identifying the disease within a short period of time using image comparison(Canny edge detection algorithm& Histogram algorithm).Along with this finally we use monitoring server, here information about the crop will be sent to the farmer, which will be very helpful for identify the disease & provides the recovery information about the disease.
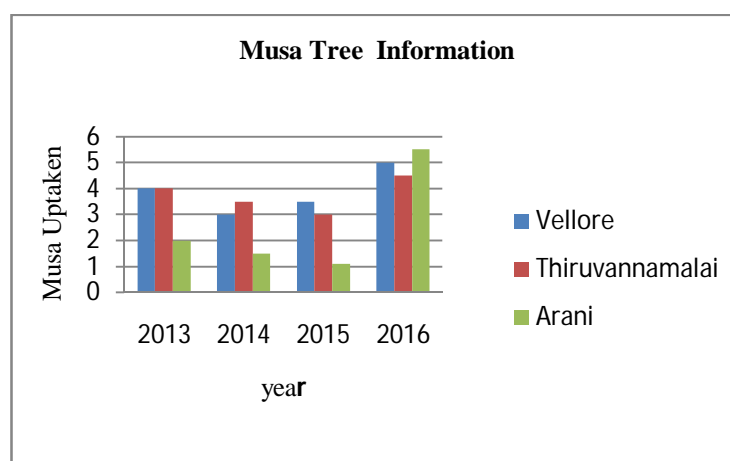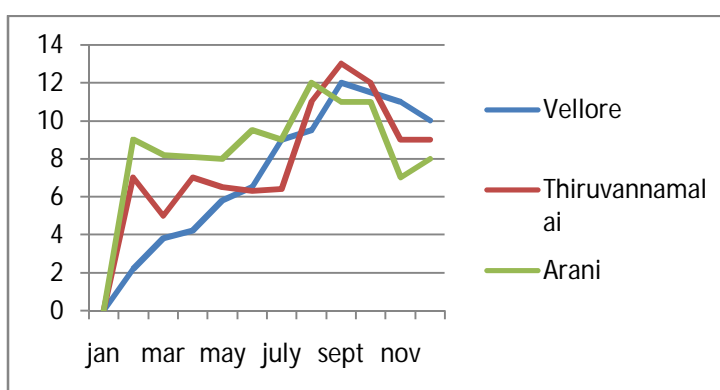

Fig.2.Musa Tree Information

Fig.3.Monthly improvement in crops

IV. CONCLUSION AND FUTURE WORK

In our project to overcome the disadvantages of existing system & to improve the quality and quantity of the crop. Cloud was proposed with various servers inorder to overcome the disadvantage of existing system. The accuracy of soil moisture retrieval over agricultural fields was examined using soil tests. The identified diseases helps to keep track of disordered of diseases in the fruit crop. The task of diseases symptoms classification using image processing technique & DNA sequence testing was successfully implemented. The proposed work aims to development of methodology for identification and classification of diseases symptoms affected on fruits(banana) crop. We wish to proposed

architecture to remotely monitor the crop for possible diseases and detect as early as possible to aid further loss of crop. The work is complex and challenging in terms of high variability in outdoor conditions and general symptoms.

REFERENCES

1. Setubal, J.C., Meidanis, J.: Introduction to Computational Molecular Biology. Brooks/Cole Publishing Company (1997).
2. Smith, T.F., Waterman, M.S.: Identification of common molecular sub-sequences. Journal of Molecular Biology (1981) 195–197.
3. et al., S.F.A.: Gapped blast and psi-blast: a new generation of rotein database search programs. Nucleic Acids Research **25** (1997) 3389–3402.
4. Pearson, W.R., Lipman, D.L.: Improved tools for biological sequence comparison. In: Proc.National Academy Of Science, NAS (1988) 2444–2448
5. Martins, W.S., Cuvillo, J.B.D., Useche, F.J., Theobald, K.B., Gao, G.R.: A multithreadparallel implementation of a dynamic programming algorithm for sequence comparison. In:Symp. on Computer Architecture and HPC (SBAC-PAD). (2001) 1–8.
6. S. Breß. The Design and Implementation of CoGaDB:A Column-oriented GPU-accelerated DBMS.Datenbank-Spektrum, pages 1–11, 2014.
7. S. Dorok, S. Breß, H. L¨apple, and G. Saake. Toward
8. efficient and reliable genome analysis usingmain-memory database systems. In SSDBM, pages 34:1–34:4, 2014.
9. S. Dorok, S. Breß, and G. Saake. Toward efficient variant calling inside main-memory database systems. In BIOKDD-DEXA, 2014.
10. Wimmer MA, Eichert T (2013) Review: mechanisms for boron deficiency-mediated changes in plant water relations. Plant Sci 203:25–32.