Plant Disease Detection Using Data Mining

B. Neelima

Department of Computer Science and Engineering NS Raju Institute of Technology, Visakhapatnam, Andhra Pradesh-531173, India.

R. Roshini

Department of Computer Science and Engineering NS Raju Institute of Technology, Visakhapatnam, Andhra Pradesh-531173, India.

P. Niteesh Kumar

Department of Computer Science and Engineering NS Raju Institute of Technology, Visakhapatnam, Andhra Pradesh-531173, India.

M. Anil Kumar

Department of Computer Science and Engineering NS Raju Institute of Technology, Visakhapatnam, Andhra Pradesh-531173, India.

Abstract

Identification of the plant diseases is the key to preventing the losses in the yield and quantity of the agricultural product. The Machine Learning (ML) field has gained its momentum in almost any domain of research and just recently has become a reliable tool in the medical domain. The studies of the plant diseases mean the study of visually observed patterns seen on the plant. Health monitoring and disease detection on the plant is very critical for the substantial growth. It is very difficult to identify the diseases on the plant manually and provide the treatment for that appropriate disease. It requires a tremendous amount of work experience and should be expertise in the plant diseases and also requires excessive time for processing. Here in this proposed application we try to find out the disease of the plant based on the inputs which we observe physically on any plant. For any plant there are 5 levels of disease occurrences. We try to design a medical dictionary in which all the physical inputs are substituted according to any of the level and then try to detect which disease plant is suffered with and it will try to provide cure for that appropriate disease. Our evaluation results on the proposed method using ML approach for identifying diseases on plant which is able to identify the diseases accurately and try to provide a solution for the end users.

1. Introduction

People care deeply about their health and want to be, now more than ever, in charge of their health and healthcare [1-2]. Life is more hectic than has ever been, the medicine that is practiced today is an Evidence-Based Medicine (hereafter, EBM) in which medical expertise is not only based on years of practice but on the latest discoveries as well. Tools that can help us manage and better keep track of our health such as Google Health and Microsoft Health Vault are reasons and facts that make people more powerful when it comes to healthcare knowledge and management [3-5]. The traditional healthcare system is also becoming one that embraces the Internet and the electronic world. Electronic Health Records (hereafter, EHR) are becoming the standard in the healthcare domain [4-6]. Researches and studies show that the potential benefits of having an EHR system are

- Health information recording and clinical data repositories— immediate access to patient diagnoses, allergies, and lab test results that enable better and time-efficient medical decisions.

- Medication management—rapid access to information regarding potential adverse drug reactions, immunizations, supplies, etc [7-8].

- Decision support—the ability to capture and

use quality medical data for decisions in the workflow of healthcare

- Obtain treatments that are tailored to specific health needs—rapid access to information that is focused on certain topics [8-10].

- First the symptoms provided by the user are processed by a expert system for identifying the diseases. If the rules required for processing the data by the above are not present in the database, then the system automatically calls the machine learning algorithm technique.

2. Objective of the Project

In order to embrace the views that the EHR system has, we need better, faster, and more reliable access to information. In the medical domain, the richest and most used source of information is Medline, a database of extensive life science published articles. All research discoveries come and enter the repository at high rate, making the process of identifying and disseminating reliable information a very difficult task. The work that we present in this paper is focused on two tasks: automatically identifying sentences published in medical abstracts (Medline) as containing or not information about diseases and treatments, and automatically identifying semantic relations that exist between diseases and treatments, as expressed in these texts. The second task is focused on three semantic relations: Cure, Prevent, and Side Effect

The tasks that are addressed here are the foundation of an information technology framework that identifies and disseminates healthcare information. People want fast access to reliable information and in a manner that is suitable to their habits and workflow. Medical care related Information is a source of power for both healthcare providers and laypeople. Studies reveal that people are searching the web and read medical related information in order to be informed about their health. Show how a new outbreak of the influenza virus can be detected from search engine query data. Our objective for this work is to show what Natural Language Processing (NLP) and Machine Learning (ML) techniques—what representation of information and what classification algorithms—are suitable to use for identifying and classifying relevant medical information in short texts.

We acknowledge the fact that tools capable of identifying reliable information in the medical domain stand as building blocks for a healthcare system that is up-to-date with the latest discoveries. In this research, we focus on diseases and treatment information, and the relation that exists between these two entities. Our interests are in line with the tendency of having a personalized medicine, one in which each patient has its medical care tailored to its needs. It is not enough to read and know only about one study that states that a treatment is beneficial for a certain disease. Healthcare providers need to be up-to-date with all new discoveries about a certain treatment, in order to identify if it might have side effects for certain types of patients.

We envision the potential and value of the findings of our work as guidelines for the performance of a framework that is capable to find relevant information about diseases and treatments in a medical domain repository. The results that we obtained show that it is a realistic scenario to use NLP and ML techniques to build a tool, similar to an RSS feed, capable to identify and disseminate textual information related to diseases and treatments. Therefore, this study is aimed at designing and examining various representation techniques in combination with various learning methods to identify and extract biomedical relations from literature.

3. Scope of the Project

The present work deals with the concepts of Machine learning algorithms and the development of Web based online expert systems. An expert system follows the methodology of task-based specification and independent problem solving techniques. Here Machine learning technique is considered to find a good matching for the symptoms in the database. The new algorithm mainly focuses on the determination of the diseases

affected to the Natural plants. First the symptoms provided by the user are processed by a rule based expert system for identifying the diseases. If the rules required for processing the data by the above are not present in the database, then the system automatically calls the machine learning algorithm technique. As a whole, the system results global solution for recognizing the diseases in Natural plants. And corresponding treatments to the diseases may also be suggested to the users.

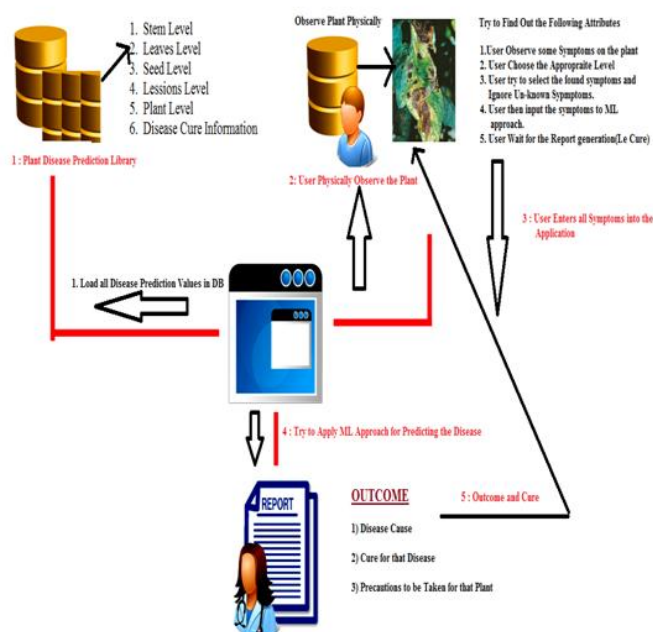## 4. SYSTEM ARCHITECTURE
4.1 System Architecture Diagram

Fig. 1 System Architecture Diagram

4.2 Process Specification
4.2.1 Input Design
The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding

extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things

- What data should be given as input?
- How the data should be arranged or coded?
- The dialog to guide the operating personnel in providing input.
- Methods for preparing input validations and steps to follow when error occur.

4.2.2 Objectives
1. Input Design is the process of converting a user-oriented description of the input into a computer-based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system.

2. It is achieved by creating user-friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way that all the data manipulates can be performed. It also provides record viewing facilities.

3. When the data is entered it will check for its validity. Data can be entered with the help of screens. Appropriate messages are provided as when needed so that the user will not be in maize of instant. Thus the objective of input design is to create an input layout that is easy to follow.

4.2.3 Output Design
A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to help user decision-making.

1. Designing computer output should proceed in an organized, well thought out manner; the right output must be developed while ensuring that each output element is designed so that people will find the system can use easily and effectively. When analysis design computer output, they should Identify the specific output that is needed to meet the requirements.

2. Select methods for presenting information.

3. Create document, report, or other formats that contain information produced by the system.

The output form of an information system should accomplish one or more of the following objectives.

- Convey information about past activities, current status or projections of the Future.
- Signal important events, opportunities, problems, or warnings.
- Trigger an action.
- Confirm an action.

4.3 Modules
1. Registration
2. Upload Data
3. Machine Learning Technique
4. Search Diseases
5. Garlic Expert System
6. Natural Expert System

4.3.1 Modules Description

4.3.1.1 Registration:
In this module, 4 users Doctors, Patients, Hospital Staff's, Medical experts want to upload the diseases details and treatment details. For that purpose they first registered his/her personal details and then login into page to upload medical files.

4.3.1.2 Upload Data:
In this module, we describe about that the users are upload the medical details like disease details, with the treatment details, etc., it is very useful for the many people. They can easily retrieve the data's from the upload database.

Each disease have separate id to generate. It stores the data in the database.

4.3.1.3 Machine Learning Technique:
In this module, the upload files are to be learned by the machine (computer). Machine learning is the study of how to make computers learn; the goal is to make computers improve their performance through experience.

Machine learning approaches in Information Retrieval is the ability of a machine to improve its performance based on previous results. Machine learning is an area of artificial intelligence concerned with the development of techniques, which allow computers to "learn". More specifically, machine learning is a method for creating computer programs by the analysis of data sets.

4.3.1.4 Search Diseases:
In this module, the search engine is designed to search for information on the databases like World Wide Web. The search results are generally presented in a list of results often referred to as SERPS, or "search engine results pages".

The information may consist of web pages, information. Some search engines also mine data available in databases or open directories. Unlike web directories, which are maintained only by human editors, search engines also maintain real-time information by running an algorithm.

In this module, finally admin verify the upload data details. To check who uploaded the data details. What is the purpose of the details. And give authority to those details. And save the data details in database.

4.3.1.5 Garlic Expert System:
In the implementation of the present system, rules are important for diagnosing a disease in the Garlic plants using machine learning expert system. Here, the rules and rule combinations are prepared according to the data given by the subject experts and stored in the database in a table format. Here machine learning algorithm is applied to get better optimization results in the present Garlic expert system.

4.3.1.6 Natural Expert System:

In the implementation of the present system, rules are important for diagnosing a disease in the Natural plants using machine learning expert system. Here, the rules and rule combinations are prepared according to the data given by the subject experts and stored in the database in a table format. Here machine learning algorithm is applied to get better optimization results in the present Natural expert system.

5. OUTPUT SCREENS
Home Page

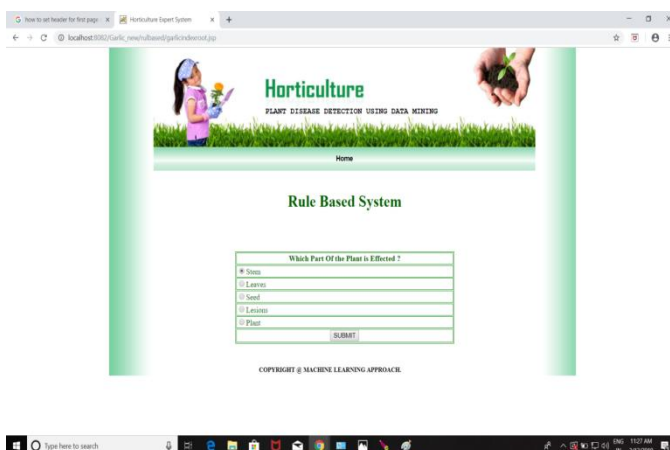Fig. 2 Rule Based Expert System to identify the plant diseases

Fig. 3 Here we need to choose which part of plant is effected and based on the selection this expert system will tell the type of disease For example we choose stem level
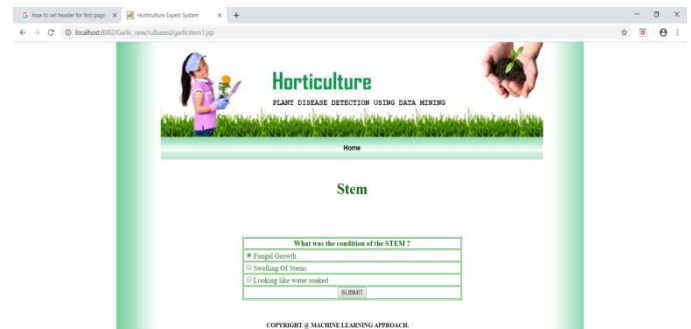

Fig. 4 Here for example we choose fungal growth on the stem, Then it will ask us what is the condition about that stem
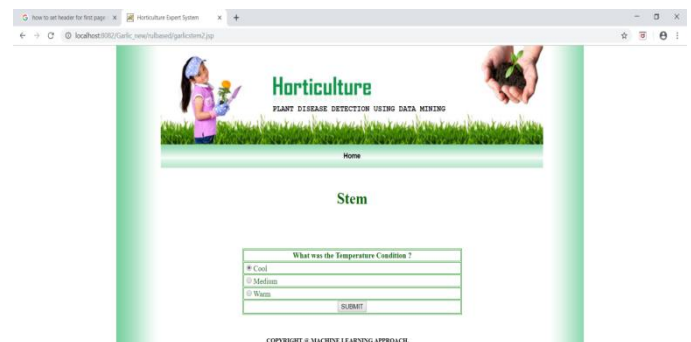

Fig. 5 Then from the above window we can choose what is the temperature condition in that region, We have choosed cool temperature
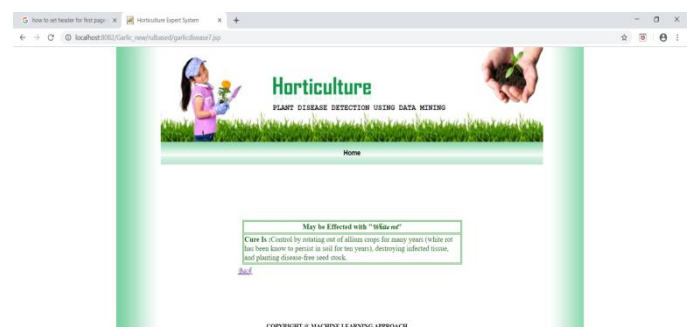

Fig. 6 It will give the disease name and cure.

6. CONCLUSION & FUTURE ENHANCEMENT

The conclusions of our study suggest that domain-specific knowledge improves the results. Probabilistic models are stable and reliable for tasks performed on short texts in the medical domain. The representation techniques influence the results of the ML algorithms, but more informative representations are the ones that consistently obtain the best results.

The first task that we tackle in this paper is a task that has applications in information retrieval, information extraction, and text summarization. We identify potential improvements in results when more information is brought in the representation technique for the task of classifying short medical texts. We show that the simple BOW approach, well known to give reliable results on text classification tasks, can be significantly outperformed when adding more complex and structured information from various ontologism.

7. REFERENCES

[1]   R. Bunescu and R. Mooney, "A Shortest Path Dependency Kernel for Relation Extraction," Proc. Conf. Human Language Technology and Empirical Methods in Natural Language Processing (HLT/ EMNLP), pp. 724-731, 2005.

[2] R. Bunescu, R. Mooney, Y. Weiss, B. Scho¨ lkopf, and J. Platt, "Subsequence Kernels for Relation Extraction," Advances in Neural Information Processing Systems, vol. 18, pp. 171-178, 2006.

[3] A.M. Cohen and W.R. Hersh, and R.T. Bhupatiraju, "Feature Generation, Feature Selection, Classifiers, and Conceptual Drift for Biomedical Document Triage," Proc. 13th Text Retrieval Conf.(TREC), 2004.s

[4] M. Craven, "Learning to Extract Relations from Medline," Proc. Assoc. for the Advancement of Artificial Intelligence, 1999.

[5] I. Donaldson et al., "PreBIND and Textomy: Mining the Biomedical Literature for Protein-Protein Interactions Using a Support Vector Machine," BMC Bioinformatics, vol. 4, 2003.

[6] C. Friedman, P. Kra, H. Yu, M. Krauthammer, and A. Rzhetsky,"GENIES: A Natural Language Processing System for the Extraction of Molecular Pathways from Journal Articles," Bioinformatics, vol. 17, pp. S74-S82, 2001.

[7] O. Frunza and D. Inkpen, "Textual Information in Predicting Functional Properties of the Genes," Proc. Workshop Current Trends in Biomedical Natural Language Processing (BioNLP) in conjunction with Assoc. for Computational Linguistics (ACL '08), 2008.

[8] R. Gaizauskas, G. Demetriou, P.J. Artymiuk, and P. Willett, "Protein Structures and Information Extraction from Biological Texts: The PASTA System," Bioinformatics, vol. 19, no. 1, pp. 135-143, 2003.

[9] C. Giuliano, L. Alberto, and R. Lorenza, "Exploiting Shallow Linguistic Information for Relation Extraction from Biomedical Literature," Proc. 11th Conf. European Chapter of the Assoc. for Computational Linguistics, 2006.

[10] J. Ginsberg, H. Mohebbi Matthew, S.P. Rajan, B. Lynnette, S.S. Mark, and L. Brilliant, "Detecting Influenza Epidemics Using Search Engine Query Data," Nature, vol. 457, pp. 1012-1014, Feb. 2009.