Hybrid Group Search Optimizer for Plant Leaf Classification

A.M.Ravishankkar, R.santhosh, P.Amudhavalli.

Abstract: The analysis of Big Data is a data mining discipline in which large quantity of unstructured data is analysed which can be challenging to store and also to retrieve efficiently. The classification of plants is based on the identification of leaf that has a very broad application in both agriculture and medicine. In this work, a method which is computerized is used to recognize a plant leaf based on images is proposed. The proposed method extracts features from the image and these are used for classifying the plant leaf. The process of deciding on the subset for all relevant features to be used in the construction of a system is known as feature selection. The Group Search Optimizer (GSO) is a nature-inspired algorithm that possesses all the qualities used effectively to solve feature selection tasks. In this work, there is a GSO-based algorithm of feature selection along with fuzzy logic and the classifier of a Neural Network (NN) is proposed. The results of the experiment prove the proposed method (GSO-NN) was able to achieve a better level of performance compared to the other methods.

Index Terms: Big Data, Group Search Optimizer (GSO), Feature Selection, Fuzzy Logic and Neural Networks (NN), Plant Leaf Classification..

I. INTRODUCTION

Big Data is a concept that is applied to huge quantity of data which does not ideally conform to the traditional database and its normal structure. For example, the data that is machine-derived will multiply quickly and consist of diverse and rich content which is discovered. The analytics of Big Data will reflect challenges made to data which are vast, unstructured and quite fast moving in order to ensure it is managed by means of employing traditional methods. Considering research institutions, business organizations, governments, and other institutions, data are routinely generated with unprecedented complexity and scope. The gleaning of meaningful information along with competitive advantages from large amounts of such data is now getting increasingly important to global organizations. By attempting to extract meaningful insights from sources of data is very challenging. So, business performance has to increase its market share and the tools have to be made available to be able to handle variety, velocity and volume of Big data. Generally, all these technologies may not always be expensive and most software is the open source [1].

The Hadoop is found to be a new open-source framework that is used to process large data across various clusters of computers that make use of languages of data processing of high levels. The modules will provide languages that are easy to use with graphical interfaces along with tools of administration to handle data of petabytes on many thousands of such computers. The Hadoop and Map Reduce are models that are used widely for the processing of Big Data. The Hadoop is a framework of large-scale processing of data supporting the processing of large data by employing simple models of programming. This Apache Hadoop project contains a Hadoop Distributed File System (HDFS) along with the Hadoop Map Reduce which was in addition to the other modules. This software will be modelled to be able to harvest on the processing for clustered computing at the same time managing node level failures [2].

The plants are the sources that provide fuel, medicine, food, oxygen and so on and thus are an essential aspect of life on earth. There is a need for having a detailed understanding of these plants in order to increase the sustainability and productivity of agriculture. There is an unavoidable growth to the human population with a varying climate that can pose a threat to several ecosystems. So, it becomes vital to be able to identify either a new or a rare species for measuring the scope of geography which is part of a very wide scheme of biodiversity. Thus, there is a need to ensure plant recognition along with its classification. When compared to the other methods like molecule or cell biology, a classification made on the leaf image will be the primary choice for the leaf plant classification. Both sampling and capturing of leaves is very convenient and also inexpensive. This captured leaf image may be easily moved to a computer and all its necessary features can be extracted automatically by means of techniques of image processing [3].

Thus, for the purpose of monitoring plants, the plant database has to be assembled for an effective and speedy grouping or classification which is an important stride. There may be a vast majority of the systems that are dependent on visual components and their extraction such as the hue and also the shape of portrayals for the classification and correlation. Albeit the various parts of a plant such as a blossom, a bud, a root, a seed or any natural produce may be used for making a distinction. There are several visual features and data modelling classifiers or techniques that were proposed for the classification of a plant leaf. There was manifold learning that was based on the features that are found in the leaf images [4].

Feature selection is a major task in the problems of classification. Looking at it in advance, for making discrimination among the classes, the features may not