ABSTRACT

The dependency of healthcare industry on the information and communication technology (ICT) domain is consistently on the rise in order to conceptualize and provide sophisticated services to various stakeholders including patients, caregivers, support service providers, medical practitioners, and experts. There are a variety of decisive advancements in the diagnosis, medication and surgical processes, medical electronics, instruments and equipment, healthcare-centric robots, a bevy of cloud-based healthcare software solutions, medical data hubs, etc. One direct offshoot of all these developments is that the amount of multi-structured data is exponentially growing. There is a litany of support and expert systems in order to lessen the doctors' workloads. However the brewing challenges and new-generation requirements include the real-time processing of medical data to extract real-time insights and decision-enablement, the substantial enhancements in appropriate and accurate processing and understanding of various and overlapped symptoms towards correct and strategically sound decisions, the real-time analytics of medical data, the empowerment of medical devices to assist surgeons and specialists in performing their tasks in an assured manner, etc.

The Problem Description - Medical imaging is one of the fundamental and most important areas of the healthcare system. This needs accuracy in processing and producing best results for further diagnosis and action. There are various factors impelling medical imaging like patient preparation, different scanning modalities, the scanner used to capture the image and various algorithms adopted for processing the captured images.

The Solution Approaches and Algorithms - One of the influencing techniques in medical image processing is segmentation. But it is a challenging and tricky problem since there are different factors such as complexity of the images and the inadequate efficient procedures to work on the human anatomy that fully capture the potential distortions in each structure. Especially to detect a tumor in the brain, image segmentation plays a vital role predominantly all through analysis of Magnetic Resonance (MR) images. There are a variety of algorithms in the literature and many of them are getting updated every day. The primary aim of these algorithms is to bring forth the mandated accuracy in finding the tumours in an effective and efficient way.

At the same time, the involvement of Digital Imaging and Communications in Medicine (DICOM) which is a standard for handling, storing, printing, and transmitting information in medical imaging also encourages precisely the need for image processing in the networks. The evolution of cloud computing and its rapid growth also transform the healthcare system to be easily accessible and faster in processing different services.

Our Contributions - In our research, we have focussed on two important aspects.

1) Brain MR Image segmentation using Parallel Genetic Algorithm (PGA).
2) Performance Evaluation of the proposed algorithm by implementing using Apache Hadoop and Spark in Cloud Services provided by AWS.

The first part comprises of various activities accomplished in a particular sequence, i) Pre-processing of the brain MR images, ii) Segmentation of the tumour from MR image, iii) Classification of the tumor.

The segmentation process starts with pre-processing of the brain MR image to remove the noise without losing the quality of the images. Therefore, the images become suitable for segmentation with optimum accuracy. Then, the segmentation is carried out using the PGA which has been proved as an efficient methodology in the literature when the search space is huge and complex. The migration operator which has been estimated as one of the important operators is averaged with the proposed methodology to increase the segmentation accuracy. Afterwards, classification of the images is done with Support Vector Machine using the features such as shape and texture to find whether the input image comprises either benign or malignant tumour or not. The Comparative analysis is done on the classified images are examined in terms of sensitivity, specificity, and accuracy. The details are given in the synopsis document.

The second part starts with the implementation of the formerly said algorithm and deploying it on cloud-based Apache Hadoop and Spark and the execution time has gone down significantly. We have chosen Apache Spark for partitioning the huge amount of data and to parallelize the tasks so that we can speed up the process of arriving at the result.

The execution time is calculated for different sizes of data 2GB, 4GB, 6GB, 8GB and 10GB of image data respectively in Hadoop and Spark. The same sets of data are executed with single, two, three, four and five clusters using Hadoop in Amazon Web Services. Execution time is noted down for each category of clusters count. The same procedure is repeated for Spark and the execution time is recorded for the same set of clusters. The comparison of the execution time in Hadoop and Spark using the proposed algorithm is analysed at the end of the research work.