International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-6, April 2020

Development of Easyget Algorithm for Deep Learning Sculptor Deepcnet Model using Hadoop Architecture

Jyoti S.Patil, G.Pradeepini

Abstract*: Manual segmentation in the brain tumors analyses for malignancy prognosis, via massive amount MRI images produced through medical routine, frustrating task and is a hard. There is a dependence on automated brain tumor graphic segmentation. The amount of precision necessary for scientific purposes is normally as yet not known, and so can't be conveniently quantified actually by means of professional physicians. That is a fascinating point, which includes just sparsely been resolved in the literature, but is nonetheless truly relevant up to now. Additionally, storage space automatization for medical images is essential need nowadays. To carry out very quickly analysis as well as, prognosis there's an imperative want of automated photo storage. Hence, this paper focused on development of new algorithm called "EasyGet" for automatic data storage and retrieval using Hadoop architecture.

Keywords: Deep learning, Big data, Hadoop storage, Machine learning

I. INTRODUCTION

Artificial intelligence (AI) [1, 2, 3] is an essential concept which usually facilitates regularly interpersonal existence and financial actions. It provides significantly to the advantageous progress of Japan's overall economy and resolves numerous communal situations. Recently, AI offers captivated interest as an integral for expansion through developing countries and so many parts of the world such as China and then India. Even though just lately created AI technological know-how undoubtedly exceeds expectation in extracting particular patterns, there are numerous boundaries. Many ICT versions are extremely reliant on big data, absence a self-thought function, and so are difficult [4].

Brain tumor is among the best extreme cancers on the planet.Gliomas will be the more prevalent brain tumors which usually occur right from growths [5]. Based on the cancerous amount of gliomas, they could be grouped into 2 grades: low-grade gliomas (LGG) and high-grade gliomas (HGG),

Revised Manuscript Received on April 30, 2020.
***** Correspondence Author
Ms.Jyoti Patil*, Ph. D. Research Scholar, Department of CSE, Koneru Lakshmaiah Education Foundation (KLEF), Guntur, A.P.India. Email: jyotipatilnba@gmail.com
Dr.G.Pradeepini*, Professor, Department of CSE, Koneru Lakshmaiah Education Foundation (KLEF), Guntur, A.P.India. Email: pradeepini_cse@kluniversity.in

the original one have a tendency to be harmless, develop further gradually because of reduce examples of cells infiltration as well as , growth, the second one are actually lignant, need instant treatment and more aggressive. Magnetic resonance imaging (MRI) [6, 7] is a trusted visualizes strategy to evaluate these types of cancer, since it offers an excellent soft tissue comparison with no radiation. An automatic segmentation of brain growth in magnetic resonance images (MRI) is essential meant for diagnosis, treatment and monitoring.

Additionally, probably the most important and hard difficulties in software program units study is to build up software as well as , equipment pertaining to safe-keeping, manage, and managing details with huge amounts of data. Presently, a lot of the data is kept in a nonstructural manner, formats and using different languages [8]. New alternatives possess come forth and Hadoop is among the most effective suggestions, it could be utilized to actually course of action these kinds of channels of therapeutic data. Even so, without a productive strategy style and architecture, these activities will never be critical as well as, useful for medical managers [9]. Info Ingestion [10] may be the procedure for adding and so gaining info for prompt makes use of or perhaps storage area in a data source or HDFS. Therefore, Data could be transmitted in real time or sometimes taken in batches. When it's assimilated instantly, it is imported as it is usually transferred by means of the foundation. Then again, in the event that data is usually taken in batches, it really is brought in discrete blocks at periodic period intervals.

Hence, in this paper to support brain tumor segmentation of images processed by use of 2Cnet sculptor model is focused for batch processing with development of new EasyGet algorithm as an extension to Hadoop HDFS.

II. LITERATURE REVIEW

Achievement in Deep Learning software was conceivable due mainly to present developments in the advancement of components systems, like images control models. Certainly, any lot of nodes needed to discover complex associations as well as, habits within just data can bring about vast amounts of variables that require to become boosted through the training phase. Because of this, DL networks require a large amount of training info, which raises the computing performance, had a need to evaluate them [11].

Cancer can be explained as a disease which in turn causes body cells to grow uncontrollably because of a plausible exploit or perhaps distributed to various other section of the body.

The tumors happen to be engorged mass partly of the body due to an abnormal development among cells. The happening of tumor in brain resulted in brain tumors.

The cancer could be frequently of cancerous or simply benign. Cancerous tumors are malignant plus they maximize without the influence and so impacts additional body cells. The harmless tumor is usually a non-cancerous tuberosidad this means they'll not pass on to the close by cells and therefore significantly less unsafe [12].

Lately, deep learning applying convolutional neural networks have proven they are in a position to control difficult signals, consisting of noise, even though preserving a higher standard of precision [13]. Applying this system, it might subsequently be feasible to max capacity the cells program noise devoid of commercial masks, as a consequence preventing the dependence on virtually any registration, and also benefitting from the increased inference swiftness generally connected with deep learning. Positron emission tomography (Family pet) resolution is a good device intended for facilitating through right difference of tumor advancement via reactive adjustments. PET/MRI is usually questioned by means of insufficient an immediate way of measuring photon attenuation. Approved alternatives for damping correction are probably not relevant to pediatrics [14].

The quantity of medical image info is speedily developing with the maintaining motion of digital therapeutic information and utilization of significant medical images pertaining to decision assist. Because of the breakthrough among Wellness Data Exchange [15] the levels of graphic info to become included and so supervised turn into gigantic. There exists a major demand to supply remarkably scalable data administration as well as, posting facilities to control, predicament, and then discuss any statistics over the network. The inbuilt features of DICOM images [16] and the raising scales offer several main concerns for the purpose of info administration: scalability meant for large scales as well as, provider of extensive concerns [17].

```
<dicom>
  <ImageType>
    <value>ORIGINAL</value>
    <value>PRIMARY</value>
    <value>AXIAL</value>
  </ImageType>
  <SOPClassUID> <value>1.2.840.10008.5.1.4.1.1.2</value> </SOPClassUID>
  <SOPInstanceUID>
    <value>1.3.6.1.4.1.9328.50.1.2843</value> </SOPInstanceUID>
  <PatientName> <value>1.3.6.1.4.1.9328.50.1.0004</value> </PatientName>
  <PatientID> <value>1.3.6.1.4.1.9328.50.1.0004</value> </PatientID>
  <PatientSex> <value>M</value> </PatientSex>
  <PatientAge> <value>057Y</value> </PatientAge>
  <StudyInstanceUID> <value>1.3.6.1.4.1.9328.50.1.2655</value> </StudyInstanceUID>
  <SeriesInstanceUID> <value>1.3.6.1.4.1.9328.50.1.2841</value> </SeriesInstanceUID>
</dicom>
```

Figure 1: Representation of DICOM header [18]

As shown in figure 1 above, the XML illustration likewise facilitates multi-value tags, by means of reproducing the worthiness component in the tag aspect, for example, “ImageType” is usually a multi-valued tag as well as , three child value elements are given. XML centered DICOM structure has also were recommended in [18] for enhancing the legibility of DICOM documents, nonetheless it relies with a little group of tags and will not supply the extensibility. XML data administration devices present vital positive aspects because they services typical data classification dialects predicated on XML requirements. One particular important problem meant for visualize big data is scalable data administration. There will be two main structure concepts for scalability: scaling up as well as, scaling out. Scaling up is founded on raising the energy among pc servers, for instance, conjoining innovative costly big machines with an increase of processing electric power or perhaps memory space.

The Gigantic facts qualities are actually quantity, assortment, as well as, velocity. Considerable information refers to large quantities of info. The quantity of information is established by means of numerous advantage frameworks just like web based network the quantity concerning details to become examined is tremendous. Plenty of data identifies different resources of data [19,20].

III. PROPOSED RESEARCH

The proposed “EasyGet” architecture as a service algorithm is developed within this research using Apache Hadoop framework. Hadoop is open up resource structure, conforms because of different graphic platforms and may become founded amongst several hospitals to shop, retrieve images and share. Many overall performance metrics such as for example stability, precision, interoperability, privacy as well as, security will be developed much like the usage of Hadoop. Modified Hadoop image storage architecture is shown in figure 2 below

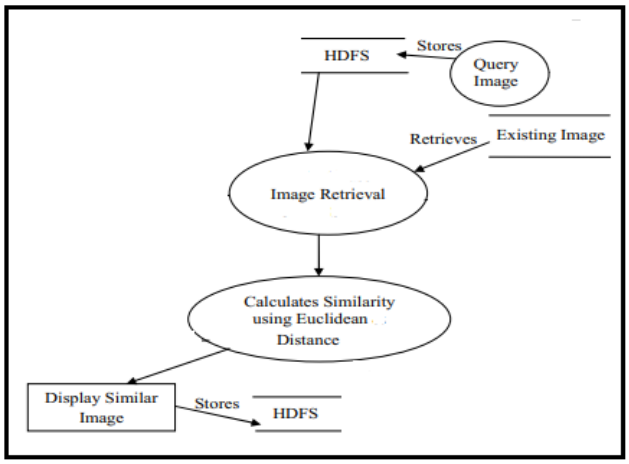

Figure 2: Modified Hadoop image storage Architecture

In one node cluster, both expert and the slave reside within the comparable node. Although, in multiple node cluster, one particular node within just the grid functions just as the master as well as, all the nodes become slaves. Processing is performed by means of any master and so jobs will be designated to slaves. The picture query is usually sent in to the get better at exactly who employs the Hadoop system to course of action the data kept in Hadoop Distributed Document Program (HDFS) and then shows outcomes recovered via the neighborhood file system.HDFS [21] offers learn/slave structure (Refer figure 3 below). An HDFS cluster consists of a solitary NameNode, a grasp machine which usually copes with any record program namespace as well as, controls usage of data files by means of consumers. Moreover, there are a variety of DataNodes, generally one particular per node in the group, which control storage space mounted on the nodes that may individuals operate on.

HDFS reveals a document system namespace and enables consumer data to be kept in data files. Internally, a document is put into a number of blocks and these blocks happen to be placed in a couple of DataNodes. The NameNode completes file program namespace procedures like starting, terminating, as well as , renaming data and so directories. In addition, it pinpoints the mapping of chunks to DataNodes. The DataNodes are in charge of serving read and create requests via the data file system's client. The DataNode likewise carry out block establishment, deleting, and then reproduction with training from the NameNode.

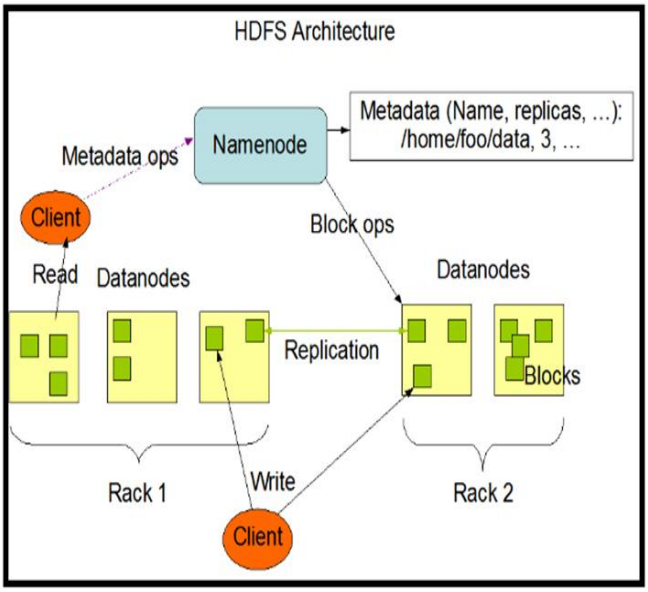

Figure 3: Hadoop- HDFS Architecture

An average deployment has committed equipment that may operate just the NameNode application. Each one of the various other devices in the group operates one particular example of the DataNode computer software [22]. The architecture will not preclude operating diverse DataNodes on a single machine however in a genuine application that's infrequently the circumstance.

A. Development of Algorithm

To make use of Hadoop pertaining to safe-keeping of CNN prepared images dataset which can be known by doctors for pre as well as, post-surgical overview of the subject. To supply closure to get rid of software level availability for doctors, radiologist and surgeons, we offered extra service intended for storage area and so recuperation of images which may be convenient to examine through doctors. We developed new algorithm for Hadoop called "EasyGet".

Algorithm: EasyGet

1. Check flag "0" or "1" for new entry of images processed by Sculptor DeepCnet
2. If flag != null, then
3. Activate executable batch file to create cluster
4. Load image cluster for sequential buffer entry
5. Automatically run node manager to set permissions to upload to Hadoop Node manager
6. Set flag "0" after cluster uploaded to Hadoop cluster and Repeat step 1 to 7
7. If action is "retrieve"
8. Execute interface launch
9. Select image names
10. Set down-flag = 1
11. Send request to Node manager
12. Get image handle
13. Retrieve images
14. End

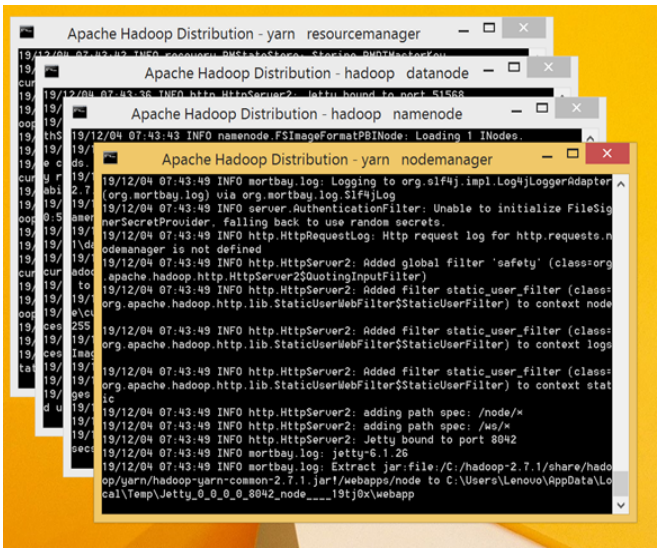

Figure 4: Hadoop Resource Manager Initiation

The presence of an individual NameNode in a group significantly simplifies the architecture of the machine. The NameNode may be the arbitrator as well as, databases meant for all HDFS metadata. The machine is engineered so the fact that user data by no means passes by using the NameNode (refer figure 4 above). HDFS accessed from applications in lots of distinct approaches. Furthermore, an HTTP browser can be used to see the documents of an HDFS example. Work is happening to expose HDFS with the aid of the WebDAV process. The processed images of BRATS dataset will be uploaded applying python. That is batch upload of prepared images so; doctors can upload images easily.

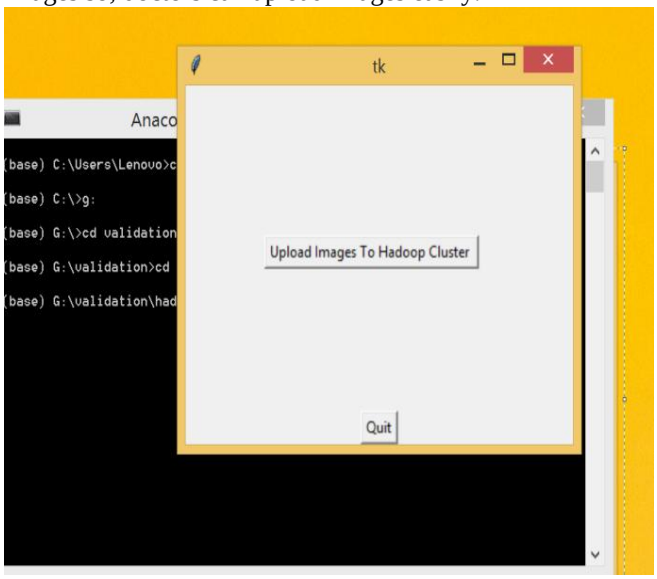

Figure 5: Hadoop Data Cluster CNN Processed Images Upload

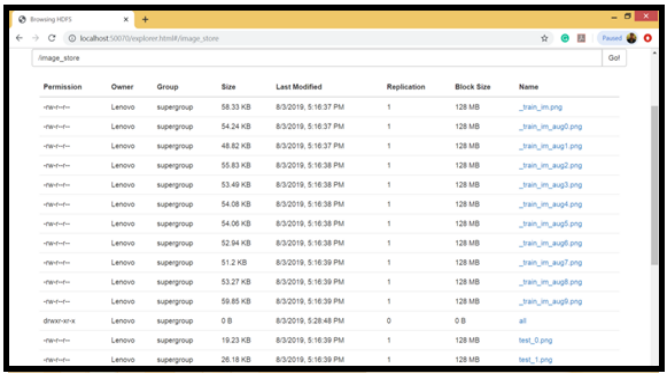

Figure 6: Hadoop Data Cluster CNN Processed Images

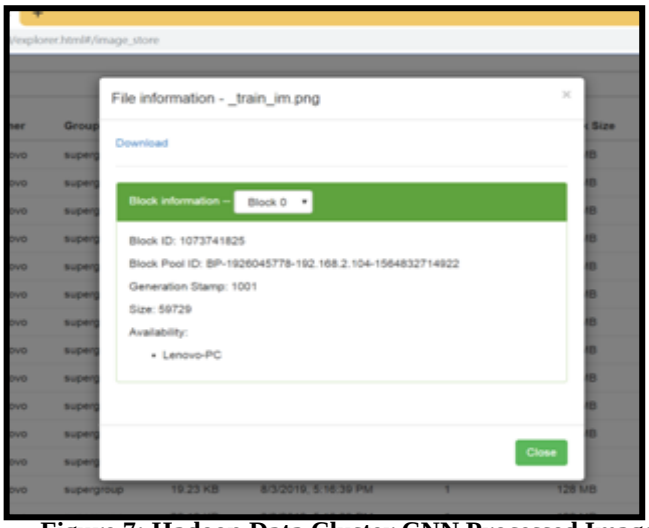

Figure 7: Hadoop Data Cluster CNN Processed Image Download

Similarly, downloading images can be done using Hadoop interface via cluster formed by processing of data nodes. This is the solution to manage and keep track of brain images for patient's pre-surgical and post-surgical analysis (Refer figure 5, 6 and 7).Thus, the overall flow of proposed work is to train BRATS dataset images using proposed algorithm, followed by validation and comparison of proposed and existing parameters as Dice Similarity Coefficient, Positive Predictive Value, Tolerance and Sensitivity. Where, Dice similarity coefficient is a spatial overlap index and a reproducibility of validation metric. Positive predictive value is the probability that subjects with a positive screening test truly have the disease. Tolerance interval limit is statistical concepts must be similar to sensitivity.

IV. RESULTS AND ANALYSIS

Based on validation tests, it is concluded that proposed system results are improved than existing system. Refer results shown in the following table 1.

Table 1: Sculptor DeepCNet Comparison Analysis

Parameters Tested	**Proposed System Results** (%)	**Existing Results** (%)
DSC	0.91	0.80
PPV	0.86	0.23
Tolerance	0.76	0.74
Sensitivity	0.77	0.67

Following graphs (figure 8) shows performance evaluation of proposed system with existing research evaluation and existing paper implementation evaluation.

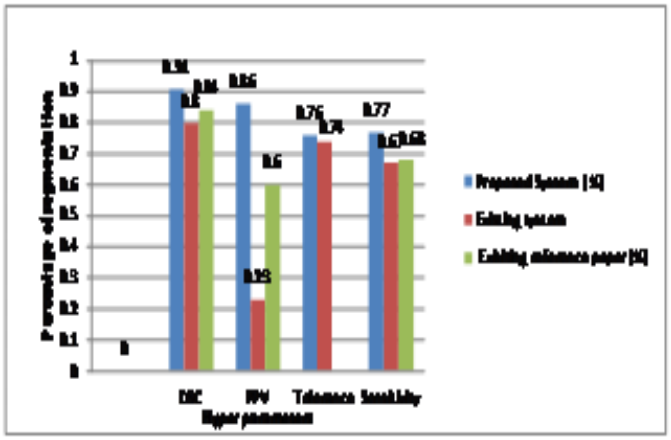

Figure 8: Performance evaluation of Sculptor DeepCnet

The processed images are stored in output folder by Sculptor DeepCnet [23] which further used to automatically upload and retrieve as a batch processing unit.

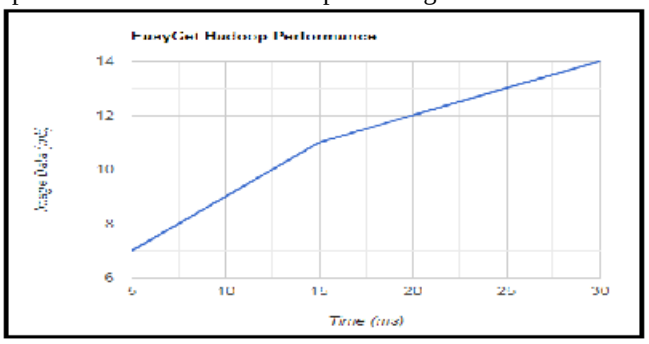

Figure 9: Performance evaluation of Hadoop EasyGet

The automated admin scripts are part of EasyGet algorithmic model. Above figure 9 shows the batch processing for one set with round trip as a upload and retrival of one batch with particular image batch size (in MB) and duration taken by system to process the same (in ms).

V. CONCLUSION

In this paper the two applications are presented to increase the Hadoop data storage utilization. The efforts taken to upload and retrieve the batch of processed medical images. The deep learning model Sculptor Deep Cnet provided very efficient processing of images. The core challenge for medical professionals is storage of report images so; EasyGet provides batch processing of any medical image data. The performance shows that this is quicker image uploading and retrieval system using automatic indexing. This definitely will curtail down the need of actual paper report storage and manual indexing. As a future development one can convert manual reports to provide facility for hospitals. Also, similar data can be stored using "nii" compressed image file format.

REFERENCES

1. Topol, Eric J. "High-performance medicine: the convergence of human and artificial intelligence." Nature medicine 25.1 (2019): 44-56.

2. Kim, Dong Wook, et al. "Design characteristics of studies reporting the performance of artificial intelligence algorithms for diagnostic analysis of medical images: results from recently published papers." Korean journal of radiology 20.3 (2019): 405-410.
3. Vasant, Pandian. "A general medical diagnosis system formed by artificial neural networks and swarm intelligence techniques." Biotechnology: Concepts, Methodologies, Tools, and Applications. IGI Global, 2019. 788-803.
4. Adir, Omer, et al. "Integrating Artificial Intelligence and Nanotechnology for Precision Cancer Medicine." Advanced Materials (2019): 1901989.
5. Shandong, Wu, et al. "Pittsburgh Center for Artificial Intelligence Innovation in Medical Imaging." (2020).
6. Visvikis, Dimitris, and Catherine Cheze Le Rest. "Functional Imaging for Customization of Oncology Treatment." Healthcare and Artificial Intelligence. Springer, Cham, 2020. 145-151.
7. Lin, Dana J., et al. "Artificial Intelligence for MR Image Reconstruction: An Overview for Clinicians." Journal of Magnetic Resonance Imaging (2020).
8. Sebaa, A., et al. "Research in big data warehousing using Hadoop." Journal of Information Systems Engineering & Management 2.2 (2017): 10.
9. Sebaa, Abderrazak, et al. "Medical big data warehouse: Architecture and system design, a case study: Improving healthcare resources distribution." Journal of medical systems 42.4 (2018): 59.
10. Erraissi, Allae, Abdessamad Belangour, and Abderrahim Tragha. "Digging into hadoop-based big data architectures." International Journal of Computer Science Issues (IJCSI) 14.6 (2017): 52-59.
11. Sun, Jingchao, et al. "A Deep Learning Method for MRI Brain Tumor Segmentation." International Conference on Frontier Computing. Springer, Singapore, 2019.
12. Anil, Abhishek, et al. "Brain Tumor detection from brain MRI using Deep Learning." Brain 3.2 (2019): 458-465.
13. Iqbal, Sajid, et al. "Deep learning model integrating features and novel classifiers fusion for brain tumor segmentation." Microscopy research and technique 82.8 (2019): 1302-1315.
14. Ladefoged, Claes Nøhr, et al. "Deep learning based attenuation correction of PET/MRI in pediatric brain tumor patients: evaluation in a clinical setting." Frontiers in neuroscience 12 (2019): 1005.
15. Lydia, E. Laxmi, et al. "Apache Hadoop for processing image files using Sequence file." TEST Engineering & Management 82 (2020): 1232-1244.
16. Jyoti Patil,Dr.G.Pradeepini, “Brain Tumor Levels Detection in Three Dimensional MRI using Machine Learning and MapReduce”,Indian Journal of Public Health Research and Development volume 10,number 6,June 2019.921-927.
17. Nazari, Elham, Mohammad Hasan Shahriari, and Hamed Tabesh. "BigData Analysis in Healthcare: Apache Hadoop, Apache spark and Apache Flink." Frontiers in Health Informatics 8.1 (2019): 14.
18. Teng, Dejun, Jun Kong, and Fusheng Wang. "Scalable and flexible management of medical image big data." Distributed and Parallel Databases 37.2 (2019): 235-250.
19. MAMDOUH, RAFEEK, et al. "Converting 2D-Medical Image Files “DICOM” into 3D-Models, based on Image Processing, and Analysing Their Results with Python Programming."
20. Li, Wei, et al. "MISS-D: A fast and scalable framework of medical image storage service based on distributed file system." Computer methods and programs in biomedicine 186 (2020): 105189.
21. Manike, Chiranjeevi, Ashok Kumar Nanda, and Tejashwini Gajulagudem. "Hadoop Scalability and Performance Testing in Homogeneous Clusters." Proceedings of ICETIT 2019. Springer, Cham, 2020. 907-917.
22. Sharma, Anil, and Gurwinder Singh. "A Review of Scheduling Algorithms in Hadoop." Proceedings of ICRIC 2019. Springer, Cham, 2020. 125-135.
23. Jyoti Patil,Dr.G.Pradeepini, “Development of deep learning Algorithm for Brain Tumor segmentation.”,International Journal of Engineering and Advanced Technology,vol 9,issue -1,October 2019.2800-2803.

AUTHORS PROFILE

MS. Jyoti Patil,Currently working as Head of Department and Associate professor Department of Information Technology JSPM‟s Jayawantrao Sawant college of Engineering Hadapsar ,Pune,India. Her Major Area of Research are Deep learning,Data Analytics, Hadoop MapReduce, Image processing. She is Pursuing Ph.D. in deep learning bigdata from KLEF deemed tobeUniversity, Guntur,A.P,India.She has Published Patent On “Detection Of Brain Tumor Levels In MapReduce 3D MRI Images Using Hadoop”

Dr G Pradeepini, Working as Professor in department of CSE at Koneru Lakshmaiah Education Foundation, Guntur,A.P.she has Completed Ph.D from PADMAVATHI MAHILA University. She has vast experience in datamining,machine learning, and image processing. She wrote almost 19 articles in international journals and 9 papers in conferences.