Spark Machine Learning Pipelines to Predict Brain Tumor using Deep Learning

Abhishek Sawant, Naveen Kumar N

Abstract: ***One of the challenging points in the field of Healthcare and Medicine is to detect cancer cells from a brain image generated in terms of MRI, CT-Scan etc. For efficiently determining the cancer cells in an image first the background of the image must be eliminated, Background check verification of patients is necessary. As multiple CT-Scan images show multiple cancer cells, sometimes though there are multiple cancer cells found in an image, yet we forget the root cancer cell and focus upon every cancer cells visualizable. The main target must be to visualize the number of cancer cells which are the root cause for child cancer cells and a re-verification on images which has multiple cancer cells detected. During such observation we ought to lack the cancer cell differentiation such kind of cancer and its type. In this paper we present the solution for the stated problem by processing an Image Dataset of Brain Tumor of various Patients using Tensor flow. The Image are contrasted based on features analyzed from the Dataset. Modelling is carried out using Deep Neural Featurizer which utilizes Inception V3 and Image Classification based on Logistic Regression. HealthCare's, Govt Agencies and Corporate companies working for Healthcare are looking for ways to assist patients, customers, staff and assets to tackle the cancer during the initial stage, provide solutions to help detect cancer at early stages. In this paper we introduce to deep learning models, Large-scale computing platform and combined altogether to learn powerful feature representations within image classification and retrieval.***

Index Terms: Brain Tumor, Deep Learning, DCNN, Image **Classification, Inception, Logistic Regression.**

I. INTRODUCTION

Mage Retrieval, classification of Images are the hottest Tre nds in the field of Computer vision and have pulled in extra ordinary heed these days with the rapid emergence of Large -scale information. The main aimis to reduce code complexi ty, no of lines. Yet Deep learning models have been utilized forhuge scale calculations. TensorFlow is an AI framework that works everywhere on large scale images and in hetero geneous conditions. Tensor Flowis nothing, but a flow that utilizes dataflow diagrams to show calculations, shared stat e, and the processes that transforms that state. Mapping of dataflow graphs is criss-crossed over some of the machines in a group and within machines extending different comput ational Devices, broadly useful GPU's and customized ASI Cs known as Tensor Processing Unit.

Revised Manuscript Received on May 10, 2019

Abhishek Sawant, Pursuing MTech School of Computer Science and Engineering, VIT, Vellore.

Naveen Kumar N, Professor, VIT, Vellore.

Deep Learning Use Tensor Flow to change pictures on num eric highlights. Tensor Flow is basically a Non-Licensable programming library for numerical calculation utilizing Da ta flow charts. Hubs in the flow diagrams point to nothing b ut numerical processes, while the Data flowchart edges poin t to the Multidimensional information exhibiting impart c arried out between them. The adaptable workflow makes su re that you' re able to send a calculation to at least one CPU /GPU in a work area, servers, or Smartphones with an exist ing API. nitially Tensor Flow was created by "Google Brai n Team" to lead AI and Deep neural systems research. Spar k-deep-learning library originates from Data blocks and use s Spark for its two most grounded aspects: In the soul of Sp ark and Spark ML lib, it gives easy to-utilize APIs that emp ower deep learning in not many lines of code. It utilizes Spa rk's amazing dedicated Engine to scale out deep learning on enormous datasets. The initial step to applying deep learni ng on pictures is the capacity to stack the pictures. Deep Le arning Pipelines incorporates utility capacities that can stac k a large number of pictures into a Spark Data Frame and d isentangle them consequently in a disseminated design, per mitting control at scale. For pictures characterization, we pr oduce a model utilizing learnings which is to be transferred . Inductive exchange is an examination issue in Artificial In telligence/machine learning that revolves around distributin g data grabbed while handling one issue and applying it to an alternate one. For instance, information picked up while figuring out how to perceive vehicles could apply when atte mpting to perceive trucks. Brain Tumour is nothing but a lu mp which is caused due to un-controlled growth of brain cel ls and cell division. These Tumours can be malignant or be nign. Brain Tumours are like wildfire they spread anywhere within the body to the brain. MRI is one of the best techniq ues that produces High quality images of various parts of th e body. These kind of imaging techniques are usually requir ed to detect tumours in foot, leg and ankle. A high-quality i mage like this gives a clear view of Abnormalities and suspi cious information related to Tumour and Anatomical Data. Some of the methodologies used for image categorization ar e segmentation variation, atlas methods, knowledge based a nd shape-based methods. The next section of the paper desc ribes the survey conducted on different techniques used for Brain Tumour detection in Section II, followed by the propo sed approach which shows the approach for detecting Brain Tumour in Section III, Section IV consists of results obtain ed and Section V concludes the paper.

II. LITERATURE SURVEY

V.P. Gladis Pushpa Rathi and S. Palani [1]. They exhibited a system which comprises of three modules, specifically Segmentation, Classification, and Feature extraction modules. Image pre-processed is sectioned utilizing Probabilistic Clustering. Along these lines, key attributes are segmented for each part dependent on the shape, surface and power. Aftermath feature extraction, key attributes will be chosen to utilize Linear Discriminant Analysis for sequentialising order reason. Finally, Deep learning classifier is utilized for grouping into Tumour/non-Tumour. T Chithambaram and K Perumal [2]. They proposed Genetic Algorithm (GA) which chooses the arrangement of ideal features from this input set. Two Hybrid Machine Learning models are actualized utilizing SVM and ANN: (Artificial Neural Networks) and are tried on two diverse datasets. SVM with Genetic Algorithm is proposed for discovering fundamental possibilities in distinguishing class comprising of Tumour's and ANN with Genetic Algorithm is utilized for affirming number for precision and Accuracy. Vinay Rao et al [3]. They proposed an answer of applying DNNs to cerebrum Tumour division for the BRATS 2015 test. Our way to deal with finding Tumour's in brain images is to carry out a pixel-wise classification and deep description for every pixel dependent on its neighbourhood in every methodology (T1, T1c, T2 and Flair) and consolidate these to shape a multimodal portrayal for every pixel. A. Anbarasa Pandian et al [4]. They exhibited Image Retrieval based on structural features extracted from MRI brain Tumour dataset. We utilized dataset of type: T1-weighted Images of MRI cerebrum Tumour pictures. Two modules comprise of (i) Deep extraction procedure (ii) classification. The regulated learning calculations like Deep neural networks and machine learning techniques are utilized to group the Images of brain Tumour. Ankit Vidyarthi et all [5]. They introduced the constraints and main points of every single such methodologies in machine learning based analysis. The relative division results are talked about with certain grouping execution measures to examine the adequacy of every calculation. Segmentation is done and cluster assessments is done dependent on two variables for example inner and outside assessment. Calculations like Morphological activities, K-Means, FCM grouping and Hierarchical are executed. Each calculation is investigated in regard of clustering most extreme irregularity locale from a MR picture. To check the proposed calculations, a dataset of pivotal post differentiate MR pictures is utilized having threatening Tumour's. K-Means and FCM grouping are dependably has superior outcome when contrast to K-Means. The hereditary methodology of the quantity of groups for K-Means, makes it identical to fuzzy calculation as ACLIME and Fuzzy calculations get same DB list and practically equal Dunn list. Amruta Hebli, Dr Sudha Guptha [6]. They introduced a framework which would identify the Tumour and will characterize them as benign and threatening. Using Image Processing in combination with machine learning will aid to identify the Tumor segmentation through morphological activities. Tumour feature extraction utilizing DWT (Discrete Wavelet Transform) and characterization is done with the chemical information. Lina Chato Et al [7]. They proposed a DWT (DWT: denoising wavelet) change to enhance the precision of a model being used for prediction for large life time of cerebrum Tumour patients utilizing Magnetic resonance imaging pictures dependent on characterization approach. Dataset used from Brats-2017. MRI pictures are used to extricate Histogram highlights to prepare a forecast model utilizing ML techniques. Different ML strategies have been utilized to build up an exact expectation display. Larger the MRI imaging framework, larger will be the MRI data corruption. The outcomes demonstrate that the two-measurement denoising wavelet change technique quite marginally enhanced the exactness of a model dependent on histogram features for prediction. When a 10% specific approval technique is utilized, the tree achieves 66.7% precision. K.S. Deepak et al [8]. They displayed a framework which discloses ways to know the unknowns regarding the Brain Tumour utilizing a Machine Learning Techniques. For ex: Segmentation and classifying Images are considered as a huge standout amongst the most critical modes of passing on Information. Understanding Images and extricating the data from them are a major part of Machine Learning. One of the initial phases in assessing the Images and then to disintegrate them and discover diverse items in them. K-Means Clustering calculations are used for Image Segmentation. In the second step we perform disintegration of MRI Images utilizing choice tree and SVM calculations and then anticipate which is better. The proposed method is utilized to discover infected mind picture of patient to coordinate with the database Tumour picture. R. Geetha Ramani, K. Sivaselvi [9]. They proposed a framework to examine the ability of different regulated machine learning techniques for classifying the Brain MRI Images. Initially, images are pre-prepared, and the feature extrication is carried-out. At that point, these are used for the supervised Data Mining Techniques Viz. Relief include determination to decide important highlights. SVM has accomplished the most noteworthy Precision of 71.33% with the highlights separated through Relief include determination with Odd one out cross-validation. Random Tree accomplished a Precision of 82%. The characterization will help the division of mind Tumour from huge arrangement of MRI cuts by wiping out the typical cuts. This significantly decreases the time for computation and memory utilization. Felix Fernando Gonzalez-Navarro et al [10]. They proposed filter include Feature selection techniques. Techniques used are: long reverberation time, short reverberation time and a specially appointed blend of both. The outcomes demonstrate that feature selection licenses to radically diminish the measurement, offering in the meantime appealing arrangements both regarding forecast exactness and the capacity to translate the included unearthly frequencies.

A straight dimensionality decrease strategy that protects the class separation capacities is furthermore utilized for perception of the chose frequencies. Hai Su et al [11]. They proposed an Automated cell detection. The fundamental commitments of the strategy are: i) Approach of re-construction to deal with split contacting cells. ii)An adaptive machine learning technique utilised to handle cell varieties. The proposed technique has been widely incorporated on a dataset with more than 1900-2000 cells. The proposed strategy accomplishes the best cell location precision with a F1 score equating to 0.96.

III. METHODOLOGY

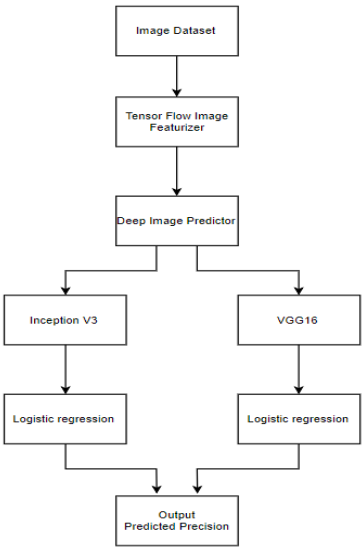

Figure 1: Flow Diagram

Deep Learning Pipelines on Apache Spark empowers quick exchange learning with a Featurizer (that change images to numerical values). We join it with *Spark InceptionV3* (a convolutional neural system prepared for images characterization) and *Logistic Regression* (a factual strategy utilized on machine learning to examine autonomous features (factors) that decides a result (for our situation where two sorts of images are there)). The Deep Image Featurizer consequently strips off the last layer of a pre-prepared neural system and utilizes the yield from all the previous layers as features for the logistic regression algorithm.

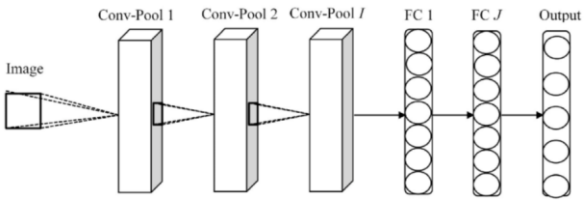

Fig 2**:** DCNN architecture

Deep Convolutional Neural Networks (DCNNs) are comprised of neurons, Neurons optimize themselves by understanding. Any single neuron would still receive an input and perform based on countless ANNs. To the final yield from the input raw picture vectors, yet the whole system will still generate keen perceptive score function. The final layer comprises of classes associated with loss functions and tricks developed for traditional ANNs would still apply.

Using the power of DCNN, Object Identification in images, Classification in images has been consistently achieved with high accuracy across various classification of images. In recent years, Deep learning methodologies specifically DCNN, has achieved quite a remarkable success in various fields. Starting from Image classification to Object detection, etc.

Convolutional layer

The convolutional layer plays a vital role in how CNNs perform. The layer's parameters revolve around the utilization of kernels. These are typically meagre in spatial dimensionality yet has wide spread along the sum of the depth of the information. When the information hits a convolutional layer, the layer convolves each filter over the spatial dimensionality of the contribution to create a 2Dimensional initiation guide.

Pooling layer

To steadily diminish the dimensionality and accordingly further decreasing the number of attributes and the computational unpredictability of the model is what pooling layer does. From the information, the layer works over every activation map and scales its dimensions utilizing the "Maximum" work.

Fully-connected layer

This layer comprises of neurons that are specifically associated with the two nearby layers, without being again associated with any layers inside them. It closely resembles the way the orchestration of neurons in custom types of ANN's are carried-out.

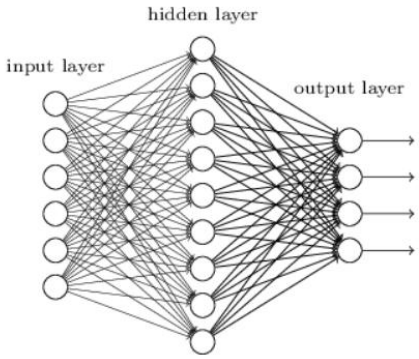

Fig 3: Layered Overview of Neural Networks

1.1 **Experimental Details**:

1. **Dataset:** "Perelman School of Medicine"

1.1 **Type of Dataset:** MRI, (Greyscale | DCM)

1.2 **Section:** T1 Type: Weighted.

1.3 **Name:** Brats-2019 Dataset

2. **Techniques**

Using Deep Featurizer following techniques are used:

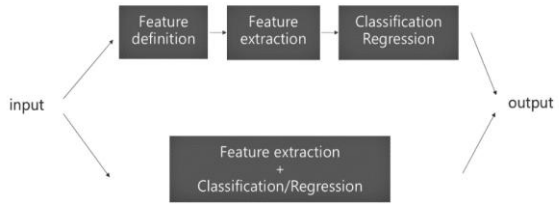

Fig 4: Deep Featurizer

2.1 **Inception V3**: An object recognition Modelling which uses pre-trained neural networks. It comprises of two stages: Stage-I: Feature Extraction with CNN and Stage -II: Classification (soft max layers).
Stage-I: Feature Extraction with CNN comprises of taking input images for the feature extraction where the image information is converted to floating values (*feature vectors*).
Stage-II: Classification consists of using pre-trained image data.
Training for the specific part to be classified and deploying them. The whole steps incorporated till now refers to Transfer Learning.

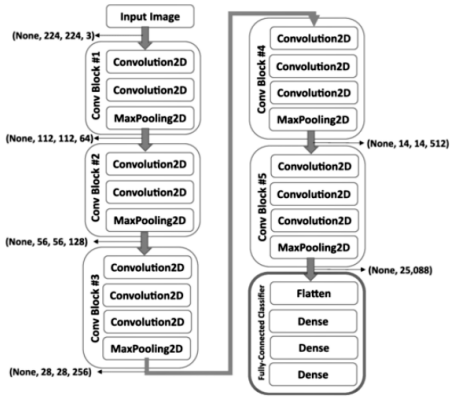

Fig 5: VGG16 Architecture

2.2 **VGG Net:** It is a Network Model where there are two types VGG16 and VGG19. Basically 16 and 19 shows that there are particularly that much number of layers in it. We are using VGG16 here in our proposed system where 16 layers is being used. There are like multiple layers which consumes majority of the parameters. As the Layers keep on increasing the parameters utilization increases too.

2.3 **Logistic Regression**: Logistic regression is a factual strategy for dataset segmentation in which there are at least one free factor decides the result. Result estimation is done with a possibility resulting variable (just two possible results). The reliant variable is paired, for example: it just contains information coded as 1 (TRUE) or 0 (FALSE). The Main objective of is to locate the best fit model to portray the connection between the dichotomous normal for intrigue and a lot of free factors.

A. Accuracy

Based on Dataset taken, obtained accuracy is 89.75% with Image Modelling as Inception V3 and 84.33% accuracy with Image Modelling as VGG16.

MODELLING NAME	ACCURACY
Inception V3	89.75%
VGG 16	84.33%

Table 1.1

B. Visualization

As shown in Figure 4 it is observed that after 539-552 stages the features extracted and Probability produced influences prediction rate-based image modes say RGB, JPEG, PNG and Greyscale.

2.4 **Deep Image Classifier**: It is a characterization administration that will recognize among 1000 distinctive image labels. It is utilized to run a Deep Detect server with an image grouping administration dependent on deep neural system pre-prepared on a subset of Image net (ILSVRC12).

2.5 **Deep Image Predictor:** It is another type of Modelling which consists of transforming images combined with classes of Image-Net by Modelling Name say 'Inception V3' for recognition of Objects to improve efficiency.

IV. RESULTS AND DISCUSSION

The implementation has been finished utilizing python, Spark and Hadoop and is appropriate for running on top of Tensor Flow. An API that consolidates Apache Spark and Tensor stream to prepare and send an image classifier. Actualizing deep learning models are made to work rapidly and be as feasible as possible for creativity. It gives different calculations that can be utilized to identify and perceive malignant growth cells in an image and group. Additionally, used to separate 3D models of articles, differentiating gaps in a picture.

V. CONCLUSION

Deep Convolutional Neural network is necessary for classifying the objects or entities. In our paper, we propose a solution to detect Brain Tumour and classify them based on features. In this research, we considered 620 images for the classification using Logistic Regression. By applying DCNN, Modelling by Deep Image Featurizer and Deep Image Predictor with Inception V3, VGG Net technique to classify, we incurred a fine accuracy for the DCNN models on image dataset. In the part of model, design of the model is carried out using Deep Learning, the process of model training takes place based on images provided with Deep Featurizer and Deep Predictor.

At the end, this work can be expanded by integrating with other Languages such which will hopefully be done very soon especially with Scala. It is reliable and efficient in terms of computational complexity.

REFERENCES

1. V.P. Gladis Pushpa Rathi and S. Palani, "Brain Tumor Detection and Classification Using Deep Learning Classifier on MRI Images" , Research Journal of Applied Sciences, Engineering and Technology, 2015.
2. T Chithambaram and K Perumal. "Brain Tumor Segmentation using Genetic Algorithm and ANN Techniques", IEEE 2017.
3. Vinay Rao et al "Brain Tumor Segmentation with Deep Learning", IEEE 2015.
4. A. Anbarasa Pandian and Dr. R. Balasubramanian, "Analysis on Shape Image Retrieval Using DNN and ELM Classifiers for MRI Brain Tumor Images", IJIEEB 2016.
5. Ankit Vidyarthi, Namita Mittal, "Brain tumor Segmentation Approaches: Review, Analysis and Anticipated Solutions in Machine Learning", IEEE 2015.
6. Amruta Hebli, Dr. Sudha Gupta, "Brain Tumor Prediction and Classification using Support Vector Machine", IEEE 2017.
7. Lina Chato et al, "Wavelet Transform to Improve Accuracy of a Prediction Model for Overall Survival Time of Brain Tumor Patients Based on MRI Image", IEEE 2018.
8. K.S. Deepak et al "An efficient approach to predict tumor in 2D brain image classification techniques", IEEE2013.
9. R. Geetha Ramani, K. Sivaselvi "Classification of Pathological Magnetic Resonance Images of Brain using Data Mining Techniques", IEEE 2017.
10. Felix Fernando Gonzalez-Navarro et al "Using Machine Learning Techniques to Explore 1H-MRS data of Brain Tumors", IEEE 2009.
11. Hai Su et al "Robust Cell Detection of Histopathological Brain Tumor Images Using Sparse Reconstruction and Adaptive Dictionary Selection", IEEE 2016

AUTHORS PROFILE

Author-1 Photo

Abhishek Sawant, Pursuing MTech in the School of Computer Science and Engineering, VIT, Vellore.

Author-2 Photo

Naveen Kumar N, Professor
VIT, Vellore.