Mutual Informative MapReduce and Minimum Quadrangle Classification for Brain Tumor Big Data

Manikandan Ramachandran[1], Rizwan Patan[2], Ambeshwar Kumar[3], Soheil Hosseini[4], Amir H. Gandomi[5]

ABSTRACT **Machine learning algorithms such as Support Vector Machine (SVM) have been widely used to detect brain tumors in big data environments. However, the SVM classifier is unsuitable for a large dataset as the complexity involved is found to be high. Therefore, in this study, a MapReduce model is introduced with SVM to handle large-scale data and deal with this issue. In this article, a framework called Mutual Informative MapReduce and Minimum Quadrangle Classification (MIMR-MQC) is introduced for brain tumor detection to handle challenges associated with big data classification. Here, preprocessing is performed using MIMR, which removes unwanted and redundant attributes in the brain tumor dataset. This technique reduces the computation complexity and time using a big dataset for detecting the brain tumors. Then, the Minimum Quadrangle Support Vector Machine model is created using Lagrange multipliers and Radial Basis Kernel function for improving the efficiency of the classification process. The MIMR-MQC framework is validated on a standard dataset called Central Brain tumor Registry of the United States (CBTRUS). Results show that the proposed model observed 21% of higher detection accuracy by minimizing the computational complexity and detection time by 37% and 27% respectively in comparison with existing models. A comparison with state-of-the-art machine learning techniques, the MIMR-MQC framework performs better in terms of brain tumor detection time and accuracy due to the better distribution of data.**

Keywords—**Big data, Mutual Information, MapReduce, Quadrangle Classification, Lagrange multipliers, cancerous, non-cancerous.**

1. INTRODUCTION

With the exponential growth of big data in the biomedical and healthcare communities, advanced disease detection has enabled timely help to patient and community service. Various big data-as-a-service frameworks were developed to process the big data in Wang et al. (2018); Wu et al. (2018) to provide proactive services. However, the accuracy was minimized with the insufficient quality of medical data. Divergent regions additionally manifested unique features of certain regional diseases, which may weaken the detection of disease. Machine learning algorithms Amin et al. (2019), streamlined in this context, were used for successful detection of chronic disease breakout in disease-frequent societies, named, Convolution Neural Network-based Multimodal Disease Risk Prediction (CNN-MDRP), Chen et al., (2017). The structured and unstructured data in the healthcare field were combined to evaluate the risk of disease. Initially, a latent factor model was

R. Manikandan is with the School of Computing, SASTRA Deemed University, Tamil Nadu, India.(e-mails: srmanimt75@gmail.com) **P. Rizwan** (corresponding author) is with the Department of computer science and engineering, Velagapudi Ramakrishna Siddhartha engineering college, Vijayawada, Andhra Pradesh, India (email: prizwan5@gmail.com) **Ambeshwar Kumar** is with the School of Computing, SASTRA Deemed University, Tamil Nadu, India. (E-mail: ambeshwar.kumar@gmail.com).**Soheil Hosseini** is with Whiting School ofEngineering, Johns Hopkins University, Baltimore, MD, USA (email: ssadath1@jhu.edu) **A.H. Gandomi** is with the Faculty of Engineering & Information Systems, University of Technology Sydney, Australia (e-mail: gandomi@uts.edu.au).

used to restore the missing data from the medical records. Then, statistical knowledge was utilized to ascertain the major chronic diseases. Finally, a novel CNN-based multimodal disease risk prediction (CNN-MDRP) method was created for both structured and unstructured data. With this, the disease prediction was done efficiently by integrating structured and unstructured features. Conducted experiments on the hospital dataset revealed higher prediction accuracy with minimum running time. Despite the observed higher prediction, computational cost or time is increased through obtaining more essential features from a big dataset. Hence, to address this issue, the Mutual Informative MapReduce technique is developed based on higher correlation essential information that is further utilized for brain tumor detection. In the MIMR-MQC framework, the input is taken as preprocessed information. The Mutual Informative MapReduce method performs minimum preprocessing to find the higher correlation information and remove the redundant attributes. By the process of preprocessing, the computational cost or time for brain tumor detection are decreased. The MapReduce function provides the resultant reduced features, where more suitable features are selected to find the tumor are called high correlation. In addition, the features that are not appropriate for detecting brain tumor are termed lower correlation information. Subsequently, the lower correlation information is discarded to decrease the computational cost or time involved during preprocessing.

Another machine learning technique is the Deep Neural Network (DNN) (Amin et al., (2018) Mohsen, et al. (2018)). A DNN-based architecture has efficiently been segmented and classified the brain tumor with the aid of magnetic resonance images (MRI). The DNN-based architecture employed 07 layers of classification. It includes 03 convolutional, 03 ReLU and a SoftMax layer. The input MR image was initially split into multiple patches. Then, the center pixel value of each patch was provided as input to the DNN. Finally, the DNN-based architecture allocated labels on the basis of center pixels and conducted the task of segmentation Liu et al., (2018). The DNN-based architecture provides better performance in terms of accuracy and average processing time. However, brain tumor detection time was not efficiently minimized in DNN. A framework based on Mutual Informative MapReduce and Minimum Quadrangle Classification (MIMR-MQC) was developed for brain tumor detection in the big dataset to address this issue. The MapReduce function was particularly well suited for large-scale data analysis that divided the dataset into smaller groups. MapReduce works on Map and Reduce functions to handle the big dataset in the MIMR-MQC framework. The map and reduce process were performed to sort the more relevant features from the input dataset. SVM also executes data classification by means of a drawn optimal hyperplane that acts as a separator between the two classes. The combination of MapReduce with the classification model Selvapandian, & Manivannan, K. (2018) in this paper, to improve the tumor detection performance and decreases the computation complexity in a large dataset.

Mutual Information (i.e., informative) Criterion MapReduce and Minimum Quadrangle Support Vector Machine model are used in this article to detect the big brain tumors data. The contribution of this work includes the removal of unnecessary features for detecting tumors using MIC-based MapReduce model. At first, Mutual Information Criterion MapReduce is developed to remove the redundant attributes present in the brain tumor dataset and extract the relevant attributes for brain tumor detection. MapReduce is created with the aid of Mutual Information in preprocessing. It ignores the uncertainty in the input dataset, executes the map, and reduces functions for providing more related features. This, in turn, removes irrelevant features in the proposed framework. Next, the Minimum Quadrangle Support Vector Machine model is designed to efficiently classify cancerous and non-cancerous cells by applying the Support Vector Machine model for tumor detection at an early stage with minimum computation overhead. The MAX-MIN margin values and radial basis kernel function is also compared to discover the brain tumor at an early stage. If the kernel value is larger than the MAXMIN margin value, then the input is cancerous; otherwise, the **input** is noncancerous (used multiple times). This means that the brain tumor detection rate is highly improved.

This paper has the following organization. Section 2 describes the background knowledge, including MapReduce, classification, and brain tumor detection model. Section 3 proposes the Mutual Information Criterion MapReduce model and the Minimum Quadrangle Support Vector Machine model for brain tumor detection. Sections 4 and 5 present the experimental settings and discuss the results, respectively. Finally, Section 6 presents the conclusion.

II. BACKGROUND

Organizing, managing and interpreting big healthcare data is demanding, economically costly, and challenging. A blueprint for sustained supervision and interpretation of such complex data remains difficult without a powerful, fundamental theory for characterization, investigation, and reasoning. Dinov (2016) investigated several big data challenges, opportunities and software mechanisms for combining healthcare data, advanced analytic tools, and distributed scientific computing Xing et al., (2015). Examples were also provided for processing heterogeneous datasets with the help of automated and semi-automated classification techniques. However, the precise detection of brain tumors was not performed. Taheri et al. (2010) introduced a threshold-based scheme that was search-based and adaptive for efficient brain tumor segmentation. The scheme also calculated the global threshold for tumor segmentation through, which the tumor's density estimation was efficiently made by discarding the unnecessary tumors. A method was also presented to devise the value of the threshold iteratively, resulting in improved tumor detection accuracy. However, accuracy was compromised with a large dataset.

Pereira et al. (2016) investigated an automation segmentation method based on the Convolutional Neural Network to address the issue by applying intensity normalization. However, accurate brain image segmentation was difficult and consumed more time. The brain is one of the most complicated organs in the human body with billions of cells. A brain tumor is occurred when there is an unbounded partition of cells, resulting in an abnormal group of cells inside the brain. Deep learning, a machine learning technique is applied to several complex problems, classified the brain tumors

in humans (Mohsen et al., 2017). This was performed utilizing discrete wavelet transform and principal component analysis, resulting in the improvement of tumor cell classification. However, efficient differentiation between the healthy tissues and non-healthy tissues were not performed. Menze et al. (2016) studied a generative probabilistic model for modeling brain tissues to address this issue by applying Gaussian Mixture and Expectation Maximization to delineate lesion areas. However, the more detailed information about brain tumor could not be analyzed.

Automated computer-aided detection is one of the most important topics in medical imaging. Roth et al. (2016) studied a two-tier cascade framework by applying coordinates of the region of interest, performed sampling through scale transformations, random translations, and rotations based on the extracted region of interest to different medical imaging datasets. However, the two-tier cascade framework consumed more time to detect the brain tumor with large-scale data. For large-scale data clustering, Banharnsakun (2017) investigated a MapReduce-based Artificial Bee Colony model to cluster huge volumes of data in a reasonable time. This was performed by applying squared Euclidean distance between each data and centroid. Nevertheless, the relevant features failed to be extracted in a brain tumor image. Tripathi et al. (2018) designed an Enhanced Grey Wolf Optimizer (EGWO) algorithm with binomial crossover and levy flight steps to grow the search capability. Large-scale datasets were also analyzing din EGWO by applying the MapReduce model, and the method was proved better in terms of F-measure. The task of data analytics is growing in the healthcare industry over the last few years with the immense influx of multimodality data. This has also resulted in the growing interest in the generation of analytical, data-driven methods, and mechanisms of machine learning algorithms in health informatics. Using deep learning, Ravı et al. (2017) studied an elaborate review of research in health informatics, including its merits, demerits, applications and so on. Al-Ayyoub et al. (2012) presented yet another method that used a decision tree to classify brain tumor images. However, the accuracy level was compromised with increased noise or unwanted data. Aslam et al. (2015) significantly improved an edge detection algorithm to extract tumor cells using the image independent thresholding method. However, the early stage of brain tumor detection was not performed. Shboul et al. (2017) presented a quantitative analysis for brain tumors; although, its accuracy was not guaranteed while handling noisy images.

Shree & Kumar (2018) studied an algorithm for removing the unwanted noise, feature extraction using Gray Level Co-occurrence Matrix and segmentation of brain tumor based on Discrete Wavelet Transform to minimize the complexity involved in tumor detection. The performance of brain tumor detection was improved using this method. However, the size of the input data was not reduced. Aswathy et al. (2017) investigated a wrapper-based genetic algorithm by combining machine learning with a support vector machine learning technique to detect and segment brain tumor images. However, due to anatomical variability, segmenting complex structures were found to be difficult. However, the accurate segmentation of complex structures is difficult due to anatomical variability. Xue et al. (2017) intended a survey of deep learning for automatic brain tumor detection. Blanc-Durand et al. (2018) proposed another automatic lesion detection and segmentation method by using a convolutional neural network, but the

performance was not improved. Jha et al. (2017) reported a novel method, including Wiener filter to decrease the noise or unwanted information, for feature extraction using 2D-discrete wavelet transform (2D-DWT). With the extracted results, probabilistic principal component analysis (PPCA) reduced the dimensions and, finally, a random subspace ensemble (RSE) classifier efficiently detected brain tumor. However, the dimensionality reduction remained unaddressed.

Paul & Rho (2016) developed a probabilistic model with mobile and dynamic machines to identify the shortest path and the lowest cost among machines using a graph-based model. However, the real-time power management was not performed. Bhattacharjee et al. (2018) presented the Plant Growth Simulation Algorithm (PGSA) to automate and optimize the leukocyte analysis process. Despite optimization, time consumption was not reduced. However, time consumption was no sufficiently decreased. Din & Paul (2018) discussed a healthcare architecture-dependent analysis for energy conservation of health monitoring sensors and recognition of big data analytics. The computations involved in healthcare architecture remained unaddressed. Rathore et al. (2016) discussed the Real-time Medical Emergency Response System, including IoT-based medical sensors, for handling the massive volume of heterogeneous data. However, detection accuracy was not efficiently increased. Paul et al. (2016) created a new concept of Smart Buddy for the analysis of wearable devices and big data to establish human behaviors. The time consumed for analysis was not minimized. Paul (2014) designed an embedded system working model to apply reinforcement learning to recognize and automate directed recognition that supports various devices. However, accuracy in detection remained unconsidered in the embedded system.

Shi et al. (2017) developed an improved rough-fuzzy c-means clustering algorithm to perform the parameter selection strategy for adaptively modifying the weighted parameter based on distributive character. It failed to solve complex problems. Shi et al. (2014) developed an unsupervised change detection technique independent of a fuzzy active contour model and a genetic algorithm. The fuzzy technique examined the image differences. The issues related to computational complexity remained unresolved. Bhattacharjee et al. (2018) created an educational model with virtual reality to apply an evolutionary learning algorithm to produce a personalized learning path. Venkatesan et al. (2018) presented several classifiers for arithmetic beat categorization. The time delay was not reachedthe desired level by using the SVM classifier. Torti et al. (2018) applied the Support Vector Machines (SVMs) algorithm to categorize hyperspectral (HS) images. However, performance efficiency remained inefficient. Amin et al.(2017)presented an automated method to distinguish between cancerous and non-cancerous Magnetic Resonance Imaging (MRI). However, the detection accuracy was reduced.

Ramakrishnan & Sankaragomathi (2017) modeled the Support Vector Machine (SVM) with different kernel functions to categorize tumor and non-tumor images. The accuracy of detecting non-tumor images was not improved. Vallabhaneni & Rajesh (2018) described an automatic brain tumor detection technique to protect the edges in the process of the De-noising image. Computational complexity issues were not resolved by employing automatic brain tumor detection technique. Soltaninejad et al. (2017) planned a fully automated method to identify brain tumors by classifying super pixels into tumorous or healthy brain tissue, classification time was not reduced. However, tumor classification time was not decread. Lakshmi et al. (2017) designed a brain tumor detection algorithm to determine the abnormality of brain images by using SVM and pointing to the kernel classifier. Jayachandran & Dhanasekaran (2014) presented a robust brain tumor classification to classify tumor and non-tumor tissue with the aid of structural analysis. However, the computational complexity remained unaddressed. Arnaud et al. (2018) designed a modular, fully automatic and data-driven procedure to detect and characterize abnormalities in MRI data. Although thedetection accuracy was insufficient, Naik& Patel (2014) planned a tumor detection and classification method to classify tumors using a brain MRI. A machine learning decision treemap algorithmwas designed for successful prediction of various disease occurrences in disease-frequent societies. However, the storage complexity was not decreased. Zhang et al. (2018) designed a novel smart pathological brain identification method with wavelet packet Tsallis entropy, extreme learning machine and Jaya algorithm. But the brain tumor detection time was higher. Zhang, Y et al. (2018) introduced a new brain detection technique depends on pseudo zernike moment and kernel support vector machine.

Certain problems are recognized from the previously described existing methods, such as precise brain tumor detection, the time required for brain tumor detection, relevant feature extraction, the early stage of brain tumor detection, computation complexity, performance efficiency and so on. Motivated by the above observation, a Mutual Informative MapReduce and Minimum Quadrangle Classification (MIMR-MQC) framework is developed with the objective of accurately detecting brain tumors in the big dataset. In this paper, the Mutual Informative MapReduce has been combined with the extraction of the most relevant features to be extracted, followed by Minimum Quadrangle Classification with efficient class labels as a classifier tool. This article studies the extraction of features or attributes from the brain tumor big dataset to detect brain tumors at an early stage. Our outcome leads to the conclusion that with this proposed framework, it helps radiologists make an effective and timely decision toward brain tumor detection.

III. PRELIMINARIES

The Map and Reduce functions of MapReduce are described with respect to data structured in (key, value) pairs. Map considers one pair of data with a form in one data domain and provides a list of pairs in a different domain:

$$Map(k1, v1) \rightarrow list(k2, v2)$$

The Map function is used in every pair ($keyed\ by\ k1$) in the input big dataset. This offers a list of pair ($keyed\ by\ k2$) for each call. The MapReduce framework gathers pairs with the identical key (k2) from the whole lists and combines to make one group for each key. Then, the Reduce function is employed to each group in a similar manner, which, in turn, provides a collection of values in the same domain:

$$Reduce\ (k2,\ list\ (v2)) \rightarrow list(v3)$$

Each Reduce call typically offers either one value or an empty return, though one call is permitted to provide more than one value. The returns of all calls are gathered as the preferred result list. Therefore, the MapReduce framework changes a list or index of

(key, value) pairs into an index of values.

3.1. Proposed framework

The proposed Mutual Informative MapReduce and Minimum Quadrangle Classification (MIMR-MQC) framework is designed to perform brain tumor detection in two distinct phases, namely, the Mutual Information Criterion MapReduce (MIC-MapReduce) phase and the Minimum Quadrangle Support Vector Machine (MQ-SVM) phase. In the MIC-MapReduce phase, pre-processing is carried out for the brain tumor big dataset to remove unwanted and irrelevant attributes or features. Each reduced feature is used for further classification in the MQ-SVM phase to detect whether it is a cancerous or non-cancerous tumor by applying Lagrange Multipliers with MAXMIN margin. These phases are discussed in detail next.

3.1.1 Mutual Information Criterion MapReduce (MIC-MapReduce) phase:

Brain tumor refers to the collection of abnormal cells in the brain. Several types of brain tumors can occur. Some brain tumors are considered to be noncancerous or benign, whereas certain other brain tumors are cancerous or malignant. The swift progress in the field of big healthcare data have resulted in the progress of custom paradigms for distributed processing that can extract significant brain tumor values and insight for brain tumor detection. Amongst several techniques, the MapReduce environment is considered the most customary criterion used in the distributed processing scenario.

Hence, the proposed framework starts with the preprocessing task of applying the Mutual Information Criterion with MapReduce for brain tumor detection. Figure 1 shows the block diagram of Mutual Information Criterion with MapReduce (MIC-MapReduce)

for brain tumor detection.

Figure 1: show that the brain tumor training and testing samples are divided into two sets, and then the training sample is taken as the input to perform MapReduce operation. The input data are divided into logical chunks and partitioned into various separate sets. These sets are then sorted, and each sorted chunk is passed to the reducer. The MapReduce model used Map and Reduce interfaces to implement the map and reduce function. The map function first reads the brain tumor dataset into independent subproblems and transforms records into a key-value 'k_j- value '' format. A map function is employed in the input data consisting of value pair that produces a set of intermediate value pairs. The reduce function associates these intermediate values corresponding to the similar intermediate key. Transformations in this phase applyto the Mutual Information (MI) criterion to obtain 'k_j – Value' on each record. It is mathematically formulated as follows:

$$MAP(k_1, value_1) \rightarrow LIST(k_2, value_2) \qquad (1)$$

From the above equation (1), the map function 'MAP ()' obtains($k_1, value_1$)' as input and obtains a list of transitional '$LIST$ ($k_2, value_2$)' pairs as output.

Output keys are then shuffled and grouped based on the critical value so that synchronous keys that are mutually related with each other (i.e., having similar meanings) are gathered simultaneously to form a listing of the values in a parallel manner. The keys are then separated and forwarded to the Reducers according to the MI criterion key based scheme previously defined. Finally, the Reducers perform fusion on the lists to eventually generate a single value for each pair.

It is mathematically formulated as follow:

$$REDUCE(k_2, LIST(value_2)) \rightarrow (k_2, value_3) \qquad (2)$$
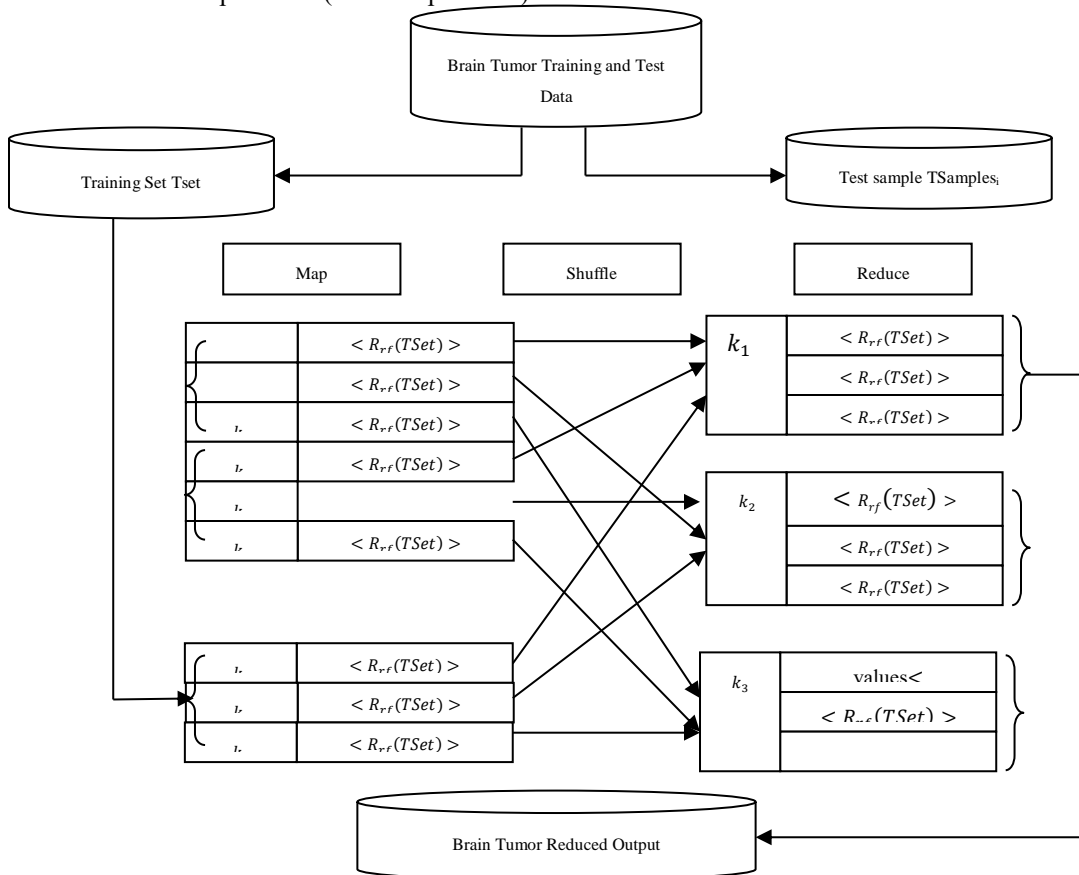
Figure 1: Flowchart of the proposed MIC-MapReduce model

This improvement minimizes the total amount of data present in the big dataset while detecting brain tumors by combining each feature (i.e., attribute) produced in the Map phase into a single pair. The mutual information (MI) in the proposed work obtains the amount of information that one random variable contains about another. In other words, it denotes the reduction of uncertainty of one random variable due to the knowledge of the other variable, therefore addressing the dimensionality reduction. In this manner, the unwanted or repeated attributes present in the input dataset are discarded. It is mathematically expressed as follows:

$$MI(P,Q) = E(P) - E(Q) \qquad (3)$$

$$= \sum_{p \in P} \sum_{q \in Q} prob\ (p,q) \log \frac{prob(p,q)}{prob(p)prob(q)} \qquad (4)$$

From the above equation (4), 'p' and 'q' represents the two random attributes with marginal probability functions '$prob(p)$' and '$prob(q)$', respectively, which '$prob(p,q)$' represents the intersection function, and 'E' is the entropy. With the above resultant values, the redundant attribute values are discarded using the Mutual Information criterion, and mathematically represented as follows:

$$R_{rf}(TSet) = MI(P_i,Q) - \alpha \sum_{A_j \in TSamples}(P_i,P_j) \qquad (5)$$

From the above equation (5), '$R_{rf}(TSet)$' represents the reduced feature 'rf' for training set '$TSet$' with '$TSamples$' representing the current set of test sample features for weight factor 'α', respectively. The algorithm of the Mutual Informative Criterion REDUCE functions as follows:

Algorithm 1 MapReduce

Input: Training Set '$TSet$', Test Samples '$TSamples_i$', Resultant probability feature 'R_{pf}', Instances 'n'
Output: Resultant Reduced Feature 'Red_{rf}'

1: **Begin**
2:     **For** $p = 0$ to $size\ (TSamples_i)$ do
3:         Measure $R_{pf}(TSet) \to$ $Compute\ MI\ (TSet_j, TSamples_i)$
4:         Measure $Result_j \to (< k : p, Value : R_{rf}(TSet) >)$
5:         Compute $REDUCE(k_2, LIST(value_2)) \to (k_2, value_3)$
6:         Remove unwanted repeated features using $R_{rf}(TSet) = MI(P_i,Q) - \alpha\ \sum_{A_j \in TSamples}(P_i,P_j)$
7:     **End for**
8:     **For** $i = 0$ to n do
9:         **If** Resultant Feature 'R_{rf}' is highly correlated to reduced 'Red_{rf}' **then**
10:         Resultant Reduced Value = 'Red_{rf}'
11:         **Else**
12:         Resultant Reduced Value = 'R_{rf}'
13:     Select the next value i=1 to n
14:         Go to step 4
15:         **End if**
16:     **End for**
17:     **End**
18: **End**

As given above, the Mutual Information MapReduce (MIMR) algorithm is performed where every map '$MAP()$' represents a resultant feature set for each test sample 'p' in Test Samples '$TSamples_i$'. The Mutual Informative Criterion MAP algorithm measures for each test sample the probabilistic features using marginal probability function and entropy. The MICMAP makes position invariance over a larger neighborhood, i.e., one per test instance, Then the '$R_{pf}(TSet)$' is outputted as value together with an identifier of test instance as a key. In this manner, the proposed framework allows the use of multiple reducers. Having many reducers is therefore found to be highly useful when dealing with big data. The reduce function comprises the collection of resultant probability features provided by the map function. In reduce phase, more relevant features for detecting brain tumor diseases has the highest probability of being selected in test samples '$TSamples_i$'.

After the map function, all features or attributes possessing a similar key are grouped. The update process has become faster since the vectors coming from the maps are sequenced on the basis of correlation. Therefore, this function Resultant Feature 'R_{rf}' is compared with the reduced 'Red_{rf}' to find high correlation values. If the Resultant Feature 'R_{rf}' is highly correlated, then the reduced 'Red_{rf}' is the resultant reduced feature; otherwise, the process is followed with the next value. With the resultant reduced feature extracted using the MIREDUCE algorithm, the tumor detection time is decreased.

3.1.2 Minimum Quadrangle Support Vector Machine (MQ-SVM) phase
Upon successful accomplishment of this feature, the result features or attributes and their class labels are fed to the classifier to detect brain tumor diseases. The traditional support vector machine learning model does not perform well while validating big data [30]. It results in an increase in computational overhead or increases computational complexity. Figure 2 shows the working diagram of the Minimum Quadrangle Support Vector Machine (MQ-SVM) model, which considers the training set as the input. Here, for each training sample, the weight factor 'w', bias term 'b' and the error values 'e' are calculated. Then, the Lagrange multiplier is employed to identify the objective functions i.e., local maxima and minima of a function (MAXMIN). Then the radial basis kernel function is determined for detecting a brain tumor by means of comparing the radial basis kernel value and MAXMIN margin value. If the radial basis kernel is greater than the MAXMIN margin value, then the patient is detected with a brain tumor. This, in turn, improves brain tumor detection accuracy, using the MQ-SVM model.

Therefore, the Minimum Quadrangle Support Vector Machine (MQ-SVM) model has been studied here to avoid the issues related to computational overhead and improve brain tumor detection accuracy. This additionally solves a set of linear equations due to the presence of fairness constraints in the formulation of the MQ-SVM, while the DNN-based architecture solves a non-linear problem. A brief description of the MQ-SVM follows.
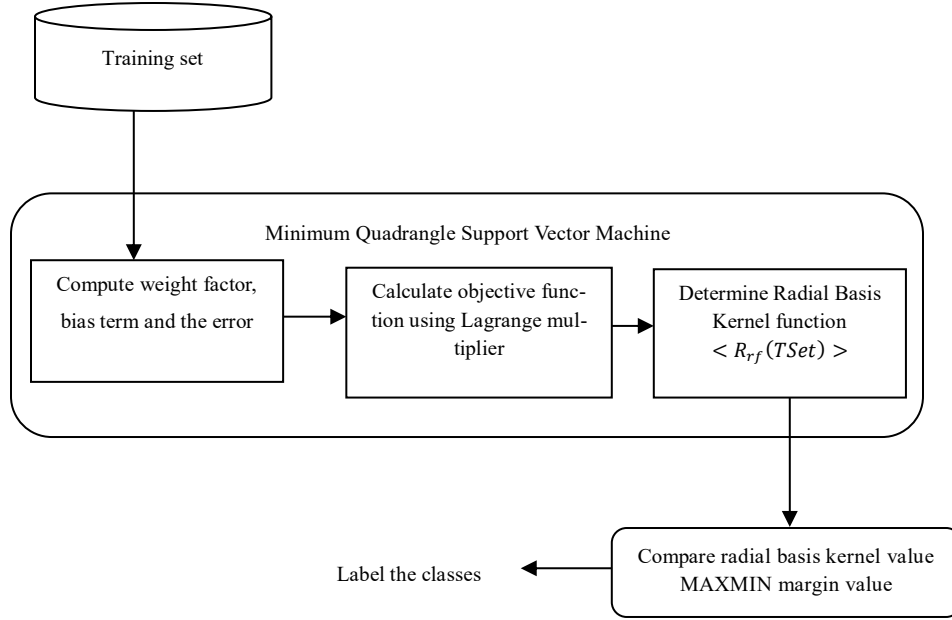
Figure 2: Flow chart of Minimum Quadrangle Support Vector Machine for brain tumor detection

Figure 2 shows the Minimum Quadrangle Support Vector Machine flow chart for brain tumor detection. Each training sample first calculates the weight factor, bias term, and error values. The local maxima and minima of a function (i.e., MAXMIN margin) are then identified by using Lagrange multipliers in MQ-SVM. Lastly, the radial basis kernel function is applied to detect brain tumors at an early stage. The input sample is cancerous if the resultant kernel value is higher than the MAXMIN margin value; otherwise, the input sample is considered as noncancerous.

As shown in figure 3, let '(Red_i, Res_i)' represent a training set of 'n' samples, where '$i = 1, 2, \ldots n$' with input data '$Red_i \in Red_{rf}$' and class labels '$Res_i = \{-1, +1\}$'; then MQ-SVM is mathematically represented as follows:

$$MIN\ J(w^T, b, e) = \frac{1}{2}w^T w + R\ \frac{1}{2}\sum_{i=1}^{n} e_i \qquad (6)$$

Subject to a fairness constraint

$$Res_i = [w^T Red_i + b] \qquad (7)$$

In equations (6) and (7), 'Red_i' represents the resultant reduced features for training samples '$TSamples$', and 'w' is the weight factor, respectively, with 'R' denoting the regularization factor. In addition, 'b' and 'e' represents the bias term, and the error value, respectively. Furthermore, Lagrange multipliers 'γ_i' are used in the design of MQ-SVM to identify the local maxima and minima of a function (i.e., MAXMIN margin) subject to fineness. It is mathematically formulated as follows:

$$L(w, b, e, \gamma) = J(w, b, e) - \sum_{i=1}^{n} \gamma_i([w^T Red_i + b]) \qquad (8)$$

Decision planes that define the boundary between classes are modeled with the calculation's resultant value. IF the training set is distinguishable in a linear fashion, then it results in two parallel hyperplanes, therefore, differentiating the data into two classes. Due to this, the intraclass distance is said to be insignificant and the interclass distance is said to be significant. The region between insignificant and significant data is referred to as the boundary, and the maximum boundary is the hyperplane that lies midway between them, which is mathematically formulated as follows:

$$\alpha * \beta - \delta = 1 \qquad (9)$$
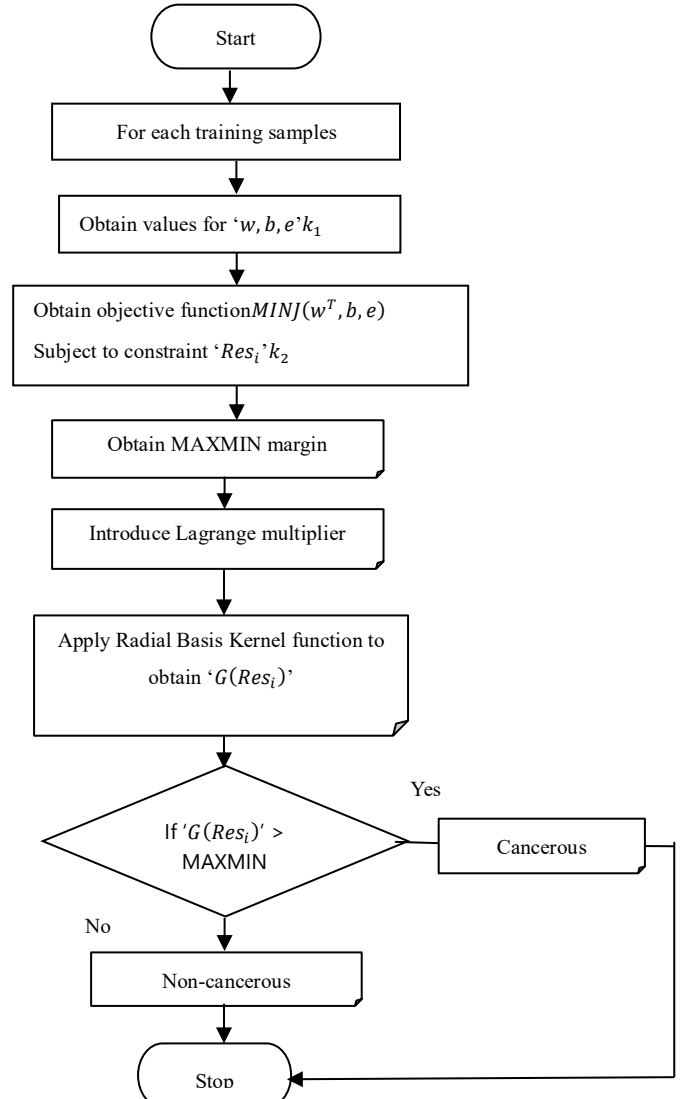$$\alpha * \beta - \delta = -1 \qquad (10)$$

Figure 3: Working Diagram of Minimum Quadrangle Support Vector Machin

A radial basis kernel function is additionally applied to the base hyper-plane and mathematically formulated as follows:

$$G(Res_i) = \alpha_i EXP \left\{ -\delta \frac{(Red - Red_i)^2}{2\sigma^2} \right\} \qquad (11)$$

With the resultant radial basis kernel value, 'G' for the corresponding input data 'Res_i', brain tumor detection is performed by way of comparison with the MAXMIN margin. If the resultant value of the radial basis kernel is greater than the MAXMIN margin value, then the patient is detected with a brain tumor. Conversely, if the resultant value of the radial basis kernel is lesser than the MAXMIN margin value, then the patient is not detected with a brain tumor. The following diagram shows the algorithm of Minimum Quadrangle Support Vector for brain tumor detection.

Algorithm 2 Minimum Quadrangle Support Vector Brain Tumor Detection

Input: Resultant Reduced Feature 'Red_{rf}', samples 'n', class labels '$Res_i = \{-1, +1\}$'
Output: Brain tumor detection

1: **Begin**
2:　　　**For** each Resultant Reduced Feature 'Red_{rf}' with samples 'n'
3:　　　Mathematically express the objective function subject to constraints using equation (6) and (7)
4:　　　　Identify MAXMIN margin using equation (8)
5:　　　　Measure the radial basis kernel function
6:　　　**End for**
7: **End**

As algorithm 2 shows, the Minimum Quadrangle Support Vector Brain Tumor Detection algorithm identifies the MAXMIN margin for each resultantly reduced feature with training samples given as input, and the objective function subject to constraints. This is followed by the radial basis kernel function being applied to the reduced resultant feature. Finally, a comparison is made with the MAXMIN margin values and radial basis kernel function values to detect a brain tumor at an early stage. As a result, the computational overhead involved in brain tumor detection is reducedto a higher detection rate.

IV. EXPERIMENTAL EVALUATION

The performance of Mutual Informative MapReduce and Minimum Quadrangle Classification (MIMR-MQC) is compared with Convolutional Neural Network-based Multimodal Disease Risk Prediction (CNN-MDRP) and Deep Convolutional Neural Networks (DNN) and validates the brain tumor detection results. The proposed MIMR-MQC framework is implemented by using the JAVAplatform with a hardware specification of the Intel® Core™ i3-4130 processor with 3.40GHZ, 4 GB memory, 500 GB hard disk and the Windows 7operating system. The test dataset, CBTRUS [4] consists of Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2009–2013. Here, the number of test samples in the range of 20 to 200 is considered for the experimental evaluation.

This dataset providesa comprehensive summary of the current, descriptive epidemiology of primary brain, and other central nervous system (CNS) tumors in the United States (US) population. The dataset contains the Central Brain Tumor Registry of the United States (CBTRUS), Brain and Central Nervous System Tumor Site

Groupings, Brain and Central Nervous System Tumor Histology Groupings, Brain and Central Nervous System Tumor Malignant Histologist, Brain and Central Nervous System Tumor Non-Malignant Histologist and ICD-O-3 Morphology Codes for all Histologist Included in Glioma Major Histology Groupings. This dataset contains a number of brain tumor tissues identified using the International Classification of Diseases for Oncology (ICD-O-3a) Histology Codes. For example, the ICD-O-3a Histology Codeb9400represents the tumor as Diffuse astrocytoma, ICD-O-3a Histology Codec9540, represents Nerve sheath tumors, ICD-O-3a Histology Codec8324 denotes Mesenchymal tumors, ICD-O-3a Histology Codec8020 denotes Germ cell tumors, and so on. By using the ICD-O site code, the cancerous patient is identified using the proposed MIMR-MQC. Brain tumor detection accuracy and computational complexity are measured to evaluate the performance of the MIMR-MQC framework'stumor detection time.

The metrics applied to measure the performance evaluation of the MIMR-MQC framework are brain tumor detection accuracy, brain tumor detection, and computational complexity or overhead. The brain tumor detection accuracy '$BTDA_s$' of an individual sample 's'depends on the number of samples correctly detected (true positives plus true negatives), and it is evaluated by the following formula:

$$BTDA_s = \frac{CD}{n} * 100 \qquad (12)$$

From (12), 'CD' presents the number of sample cases correctly detected; 'n' is the total number of sample cases. The brain tumor detection time '$BTDT_s$' of an individual sample 's' depends on the number of samples correctly detected (true positives plus true negatives), and the time taken to detect the brain tumor samples. The brain tumor detection time is evaluated by the following formula:

$$BTD_t = TSamples_i * Time(CD) \qquad (13)$$

From equation (13), the brain tumor detection time is measured by the time taken for correct detection '$Time(CD)$' with respect to the test samples '$TSamples_i$' provided as input. Finally, the computational complexity involved in brain tumor detection refers to the number of resources required to run the algorithm. The computational complexity, however, refers to the minimum of the complexities of all possible algorithms for brain tumor detection expressed in terms of milliseconds and formulated as:

$$CC = \sum_{i=1}^{n} TSamples_i * Time[MIMAP + MIREDUE + MQSV] \qquad (14)$$

From equation (14), the computational complexity in the design of the MIMR-MQC is highly influenced by the three different algorithms used, MIMAP, MIREDUCE, and MQSV with respect to the test samples.

V. RESULTS AND DISCUSSION

Firstly, the proposed Mutual Informative MapReduce and Minimum Quadrangle Classification (MIMR-MQC) for brain tumor detection is compared to malicious tumor detection with two other tumor detection methods, the Convolutional Neural Network-based Multimodal Disease Risk Prediction (CNN-MDRP) Chen et al., (2017) and Deep Convolutional Neural Networks (DNN)Amin et al., (2018). Three performance measures were employed to evaluate the tested tumor detection framework. First, brain tumor detection accuracy is observed that refers to the sample cases correctly detected to the input provided for experimentation. The second metric is the brain tumor detection time that refers to the time consumed in detecting the tumor cells.

Finally, the third measure evaluates thecomputational overhead or computational complexity.

5.1 Analysis of brain tumor detection accuracy

The brain tumor detection accuracy is measured based on the number of samples provided fora particular instance.Figure 3 illustrates the extensive experiments conducted to better understand the effectiveness of the proposed MIMR-MQC framework. Figure 3 presents a comparative analysis of the brain tumor detection accuracy for a different number of samples that combine different age groups.

The experiments were conducted using a brain tumor dataset, in the range of 20 and 200, and the brain tumor detection accuracy is measured in terms of percentage (%). The JAVA platform is used to experiment with brain tumor detection accuracy by analyzing the results using graph values. Results are presented for a different number of test samples; the outputs reported here confirm that the outcomes are not linear with the increase in the number of test samples.

Figure 4 shows the performance measure of brain tumor detection accuracy with respect to 200 different samples. The higher the brain tumor detection accuracy, the more efficient the method is (Zhao, et al., 2017). The brain tumor detection accuracy of MIMR-MQC framework is found to be higher than CNN-MDRP [1] and DNN [2]. With a higher amount of scalability achieved in the proposed framework in data-rich environments, the proposed machine learning algorithms become even more accurate and more usable with the increasing data size.

As Figure 4illustrates, the x-axis represents the different number of test samples in the range of 20 to 200 collected between the period 2009 and 2013 in the United States. It also includes samples of different age groups. Conversely, the y-axis represents brain tumor detection accuracy.It is also evident from the figure thatan increase in the number of test samples reduces brain tumor detection accuracy. However, a slight improvement is observed with test samples of 100 using MIMR-MQC and CNN-MDRP. The brain tumor detection accuracy was found to be improved when compared to CNN-MDRP and DNN
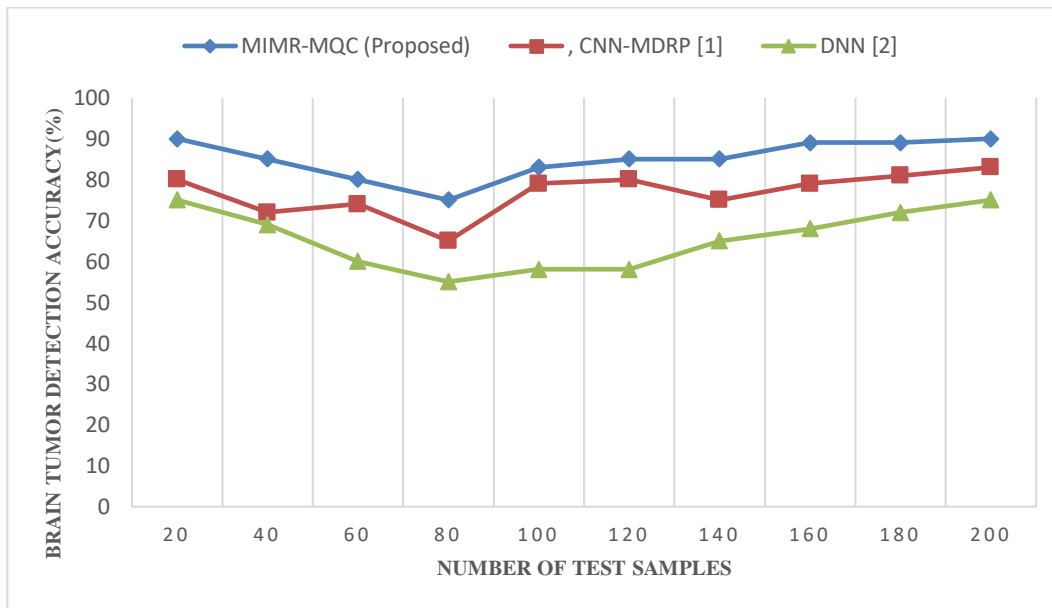

Figure 4: Performance measures of brain tumor detection accuracy using MIMR-MQC, CNN-MDRP and DNN

The brain tumor detection accuracy was improved because of the application of the Minimum Quadrangle Support Vector Machine.Weight factor, bias term,and the error values are calculated for each input sample. By applying the Minimum Quadrangle Support Vector Machine, an objective function subject to the fairness constraint was included with which the Lagrange multipliers were applied toobtain the MAXMIN boundary. Besides, a radial basis kernel function was evaluated to reduce the resultant feature and compared to the MAXMIN boundary to detect brain tumors. If the resultant kernel value is higher than the MAXMIN margin value, then the input sample is cancerous. If the comparison value is lesser than the MAXMIN margin value, then the sample is noncancerous. The framework detects the tumor present in the samples only upon a successful comparison. Therefore, brain tumor detection accuracy was improved by 11% compared to CNN-MDRP.In addition by applying the Minimum Quadrangle Support Vector Machine, brain tumor is said to be

detected at an early stage via regularization factor.With this, the brain tumor detection accuracy was improved by 31%. Precision , sensitivity, specificity are also related to detection accuracy. These are defined follow as,

Precision is defined as a number of patient files are correctly pre-dicted as disease into the total number of input files taken for the experimental evaluation. Sensitivity refers to the test's ability to correctly detect ill patients who do have the condition. Specificity relates to the test's ability to correctly reject healthy patients with-out a condition. Precision, sensitivity, specificity are increased in the proposed model MIMR-MQC compared to state-of-the-art methods.

Sample calculation for brain tumor detection accuracy :

Proposed MIMR-MQC **:** Number of test samples is correctly pre-dicted as disease or not is 18 and the total numbers of test samples are 20. Then the tumor detection accuracy is calculated as fol-lows,

$$Detection\ accuracy = \frac{18}{20} * 100 = 90\%$$

Existing CNN-MDRP **:** Number of test samples is correctly predicted as disease or not is 16 and the total numbers of test samples are 20. Then the tumor accuracy is calculated as follows,

$$Detection\ accuracy = \frac{16}{20} * 100 = 80\%$$

Existing DNN: Number of test samples is correctly predicted as disease or not is 15 and the total numbers of test samples are 20. Then the tumor detection accuracy is calculated as follows,

$$Detection\ accuracy = \frac{15}{20} * 100 = 75\%$$

5.2 Analysis of brain tumor detection time

The Brain tumor detection time with respect to a different number of samples based on the incidence rates by sex and histology for 50 different male and female respectively is measured on the basis of correct brain tumor detection. The targeting results of brain tumor detection time using the MIMR-MQC framework are presented in figure 4 for comparison based on the number of samples. A lower brain tumor detection time results in the improvement of the framework
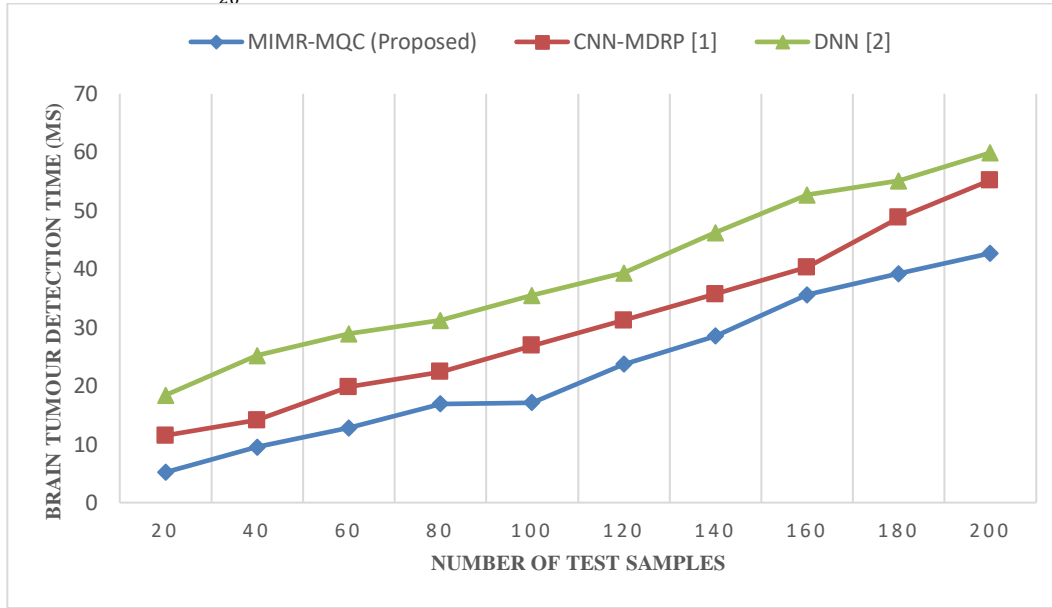

Figure 5 Performance measure of brain tumor detection time using MIMR-MQC, CNN-MDRPand DNN

In Figure 5, the x-axis represents 200 different test samples involving different age groups, $0 - 19$ years and $0 - 14$ years, respectively. Conversely, the y-axis represents the brain tumor detection time. It is evident from the figure that with the increasing number of test samples, the number of images provided as input increases and therefore the brain tumor detection time is also found to be increased. In other words, the size of the test samples increases with ahigher number of test samples considered for experimentation. Therefore, with the increase in the test samplesize, the time taken to detect a brain tumor is found to be higher. However, comparative analysis shows improvement when using MIMR-MQC rather than CNN-MDRP and DNN. This is due to the application of the Mutual Informative Criterion MapReduce algorithm.

Irrelevant and unwanted features or attributes are discarded with the application of the Mutual Information Criterion MapReduce algorithm. Only the features or attributes that contribute to the brain tumor detection are used at the later stage. At first, the map function reads the input samples into a number of independent subproblems and transforms records into a key-value format. The MapReduce function is performed to collect the resultant probability features. The maximum correlated feature is considered as the highest probability of being selected. After the map function; all features compressing a similar key are grouped.Then the function resultant feature is compared with the reduced feature to find high correlation values. If the Resultant Feature includes a higher correlation than the reduced feature, then

the resultant reduced value is the resultantly reduced feature. Otherwise, the process is continued with the next value. Therefore, brain tumor detection time using the MIMR-MQC framework is found to be reduced by 19% when compared to CNN-MDRP. However, only mutually informative or highly correlated attributes of features are extracted during the preprocessing. The proposed framework has the ability to apply the MIMAP and MICREDUCE algorithms while maintaining a reasonable computational timeby using the MIC-based MapReduce algorithm. As a result, the brain tumor detection time is reduced by 34% when compared to DNN.

Sample calculation for brain tumor detection time:

Proposed MIMR-MQC: Number of the test samples is 20 and the time for predicting one test samples is 0.21 ms, then the tumor detection time is calculated as follows,

$$Detection\ time = 20 * 0.21ms = 4.2ms$$

Existing CNN-MDRP**:** Number of the test samples is 20 and the time for predicting one test samples is 0.31ms, then the tumor detection time is calculated as follows,

$$Detection\ time = 20 * 0.31ms = 6.2ms$$

Existing DNN**:** Number of the test samples is 20 and the time for predicting one test samples is 0.359ms, then the tumor detection time is calculated as follows,

$$Detection\ time = 20 * 0.359ms = 7.18ms$$

5.3 Analysis of computational complexity

Finally, the computational complexity involved in brain tumor

detection is evaluated.Lowering thepreprocessing time or discarding the irrelevant attributes or features results in the unique presence of attributes. With the unique attributes acquired, the computational complexity involved in brain tumor detection is reduced; therefore, the brain tumor detection accuracy is improved.
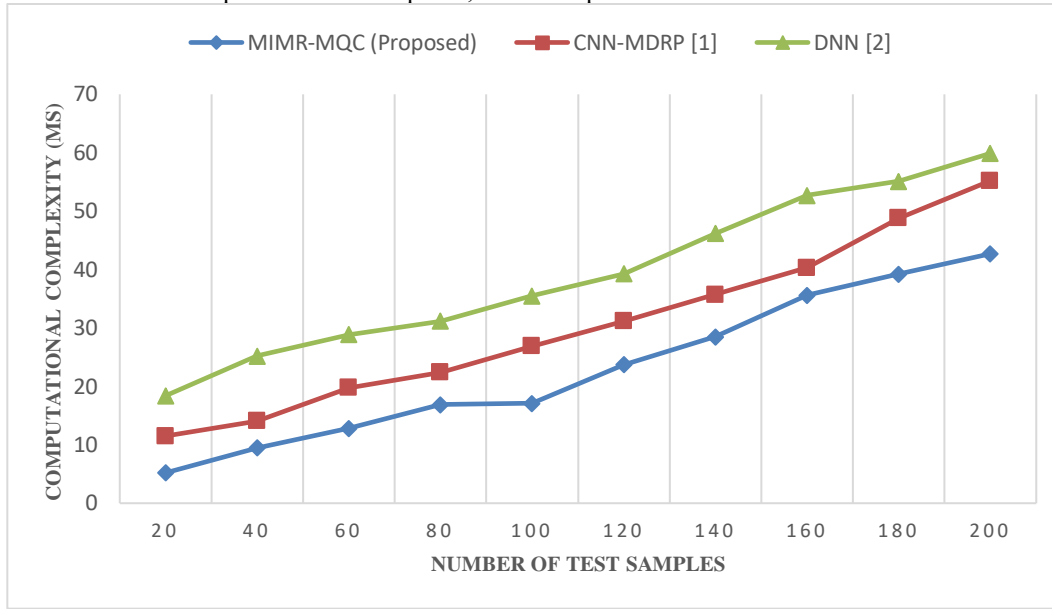

Figure 6:Performance measure of computational complexity using MIMR-MQC, CNN-MDRPand DNN

Figure 6 shows the plot of the computational complexity using MIMR-MQC, CNN-MDRP, and DNN for different test samples in the range of 20 and 200 of children aged 0 – 14 years. The plots indicate that the MIMR-MQC framework achieved significantly lower computational complexity when compared to CNN-MDRP and DNN. However, when the test sample sizes were too small, the computational complexity started to increase by applying all three frameworks. This is because the resultant probability features is found to be higher with the higher amount of test samples and due to this, reduced features also increases, causing an overall rise in the detection time.

Hence, the computational complexity using all the three frameworks is found to be higher with the varying number of samples provided as input. However, improvement is said to be achieved using the MIMR-MQC framework because of the Mutual Information Criterion with MapReduce and Minimum Quadrangle Support Vector Machine model. By applying the Mutual Information Criterion with MapReduce, unwanted features are discarded according to the higher correlation factor, resulting in the reduced feature set. Hence, the computational complexity is reduced by 28% when using the MIMR-MQC framework compared to CNN-MDRP. A higher dimensional feature space also grows the generalization errorfor big data analysis. The MIMR-MQC framework applied the Minimum Quadrangle Support Vector Machine model to address this issue. Here, the generalization error is said to be reduced due to the presence of fairness constrain and linearity factor, hence the computational complexity using the MIMR-MQC framework is reduced by 46% compared to DNN.

Sample calculation for computational complexity:
Proposed MIMR_MQC: Number of the test samples is 20 and the computational complexity for predicting one test samples is 0.081 ms, then the computational complexity is calculated as follows,

Computational complexity = 20 * 0.081ms = 1.62 ms

Existing CNN-MDRP: Number of the test samples is 20 and the computational complexity for predicting one test samples is 0.107 ms, then the computational complexity is calculated as follows,

Computational complexity = 20* 0.107 ms = 2.14 ms

Existing DNN: Number of the test samples is 20 and the computational complexity for predicting one test samples is 0.127ms, then the computational complexity is calculated as follows,

Computational complexity = 20* 0.127ms = 2.54 ms

VI. CONCLUSION AND FUTURE WORK

Deep and Convolutional neural networks have been predominantly proven much faster than other machine learning techniques. However, high computational time and cost are still challenging issues related to big data. MIMR-MQC article has aimed to solve these issues by introducing the Mutual Information with MapReduce and Minimum Quadrangle Support Vector Machine model for big data. The unwanted and redundant attributes in the big dataset are removed by applying Mutual Information with MapReduce functionality. This also helped the proposed MIMR-MQC framework to achieve comparable brain tumor detection time using highly correlative information while having computational gains of a MapReduce function on brain tumor big datasets. Linearity is also addressed through the MAXMIN margin with the application of the Minimum Quadrangle Support Vector Machine. Here, the Lagrange multiplier is employed to classify cancerous and noncancerous cellsto identify the brain tumor at an early stage. A comparison with state-of-the-art machine learning techniques revealed that the MIMR-MQC framework performs better in terms of brain tumor detection time and accuracy due to the better distribution of data. The MIMR-MQC framework additionally achieves a 31% higher brain tumor detection accuracy compared to DNN, a 46% lower computational complexity compared to DNN and a 19% lower brain tumor detection time compared to CNN-MDRP.Future work

will extend the MIMR-MQC framework for other types of tumors, like, breast, lung, abdomen and so on. The proposed MIMR-MQC framework has been tested using CBTRUS dataset for 200 different test samples. In addition, improving the number of test samples also remains an open issue.

REFERENCES

[1]. Chen, M., Hao, Y., Hwang, K., Wang, L., & Wang, L. (2017). Disease prediction by machine learning over big data from healthcare communities. IEEE Access, 5, 8869–8879.

[2]. Amin, J., Sharif, M., Yasmin, M., &Fernandes, S. L. (2018). Big data analysis for brain tumor detection: Deep convolutional neural networks. Future Generation Computer Systems.

[3]. Amin, J., Sharif, M., Raza, M., Saba, T., & Anjum, M. A. (2019). Brain Tumor Detection using Statistical and Machine Learning Method. Computer Methods and Programs in Biomedicine.

[4]. Amin, J., Sharif, M., Yasmin, M., &Fernandes, S. L. (2017). A distinctive approach in brain tumor detection and classification using MRI. Pattern Recognition Letters.

[5]. Al-Ayyoub, M., Husari, G., Darwish, O., &Alabed-alaziz, A. (2012). Machine learning approach for brain tumor detection. In Proceedings of the 3rd International Conference on Information and Communication Systems (p. 23). ACM.

[6]. Arnaud, A., Forbes, F., Coquery, N., Collomb, N., Lemasson, B., &Barbier, E. L. (2018). Fully Automatic Lesion Localization and Characterization: Application to Brain Tumors Using Multiparametric Quantitative MRI Data. IEEE Transactions on Medical Imaging.

[7]. Aslam, A., Khan, E., & Beg, M. M. S. (2015). Improved edge detection algorithm for brain tumor segmentation. Procedia Computer Science, 58, 430–437.

[8]. Aswathy, S. U., Devadhas, G. G., & Kumar, S. S. (n.d.). Brain tumor detection and segmentation using a wrapper based genetic algorithm for optimized feature set. Cluster Computing, 1–12.

[9]. Bahadure, N. B., Ray, A. K., &Thethi, H. P. (2017, June). Feature extraction and selection with optimization technique for brain tumor detection from MR images. In *Computational Intelligence in Data Science (ICCIDS), 2017 International Conference on* (pp. 1-7). IEEE.

[10]. Banharnsakun, A. (2017). A MapReduce-based artificial bee colony for large-scale data clustering. Pattern Recognition Letters, 93, 78–84.

[11]. Bhattacharjee, D., Paul, A., Kim, J. H., &Karthigaikumar, P. (2018). An immersive learning model using evolutionary learning. Computers & Electrical Engineering, 65, 236–249.

[12]. Blanc-Durand, P., Van Der Gucht, A., Schaefer, N., Itti, E., & Prior, J. O. (2018). Automatic lesion detection and segmentation of 18F-FET PET in gliomas: A full 3D U-Net convolutional neural network study. PloS One, 13(4), e0195798.

[13]. Abdel-Maksoud, E., Elmogy, M., & Al-Awadi, R. (2015). Brain tumor segmentation based on a hybrid clustering technique. *Egyptian Informatics Journal*, *16*(1), 71-81.

[14]. Din, S., & Paul, A. (2018). Smart health monitoring and management system: Toward autonomous wearable sensing for internet of things using big data analytics. Future Generation Computer Systems.

[15]. Dinov, I. D. (2016). Methodological challenges and analytic opportunities for modeling and interpreting Big Healthcare Data. Gigascience, 5(1), 12

[16]. Hunnur, S. S., Raut, A., & Kulkarni, S. (2017, June). Implementation of image processing for detection of brain tumors. In *Intelligent Computing and Control Systems (ICICCS), 2017 International Conference on* (pp. 278-283). IEEE.

[17]. Jayachandran, A., &Dhanasekaran, R. (2014). Severity analysis of brain tumor in MRI images using modified multi-texton structure descriptor and kernel-SVM. Arabian Journal for Science and Engineering, 39(10), 7073–7086.

[18]. Jha, D., Kim, J.-I., Choi, M.-R., & Kwon, G.-R. (2017). Pathological brain detection using weiner filtering, 2D-discrete wavelet transforms, probabilistic PCA, and random subspace ensemble classifier. Computational Intelligence and Neuroscience, 2017.

[19]. Kumar, B. H., Malathi, R., & Natarajan, R. A. (2018). Non-invasive grading and prognosis of Brain Tumours using fuzzy inferencing for rural India. *Materials Today: Proceedings*, *5*(1), 1921-1928.

[20]. Lakshmi, A., Arivoli, T., &Rajasekaran, M. P. (2017). A Novel M-ACA-Based Tumor Segmentation and DAPP Feature Extraction with PPCSO-PKC-Based MRI Classification. Arabian Journal for Science and Engineering, 1–17.

[21]. Liu, H., Liu, B., Yang, L. T., Lin, M., Deng, Y., Bilal, K., & Khan, S. U. (2018). Thermal-aware and DVFS-enabled big data task scheduling for data centers. IEEE Transactions on Big Data, 4(2), 177-190.

[22]. Manikandan, R., Monolisa, G. S., &Saranya, K. (2013). A Cluster Based Segmentation of Magnetic Resonance Images for Brain Tumor Detection. *Middle East Journal of Scientific Research*, *14*(5), 669-672.

[23]. Menze, B. H., Van Leemput, K., Lashkari, D., Riklin-Raviv, T., Geremia, E., Alberts, E., …Székely, G. (2016). A generative probabilistic model and discriminative extensions for brain lesion segmentation—with application to tumor and stroke. IEEE Transactions on Medical Imaging, 35(4), 933–946.

[24]. Mohsen, H., El-Dahshan, E. S. A., El-Horbaty, E. S. M., & Salem, A. B. M. (2018). Classification using deep learning neural networks for brain tumors. Future Computing and Informatics Journal, 3(1), 68-71.

[25]. Mohsen, H., El-Dahshan, E.-S. A., El-Horbaty, E.-S. M., & Salem, A.-B. M. (2017). Classification using deep learning neural networks for brain tumors. Future Computing and Informatics Journal.

[26]. Mr.Arun Kumar &Richika (2015, August). A novel Approach for brain tumor detection using support vector machine kmeans and PCA Algorithm. *In Interational journal of computer science and mobile computing* Volume 4 Issue 8 on page (457-474).

[27]. Naik, J., & Patel, S. (2014). Tumor detection and classification using decision tree in brain MRI. International Journal of Computer Science and Network Security (IJCSNS), 14(6), 87.

[28]. Paul, A. (2014). Real-time power management for embedded M2M using intelligent learning methods. ACM Transactions on Embedded Computing Systems (TECS), 13(5s), 148.

[29]. Paul, A., & Rho, S. (2016). Probabilistic model for M2M in IoT networking and communication. Telecommunication Systems, 62(1), 59–66.

[30]. Paul, A., Ahmad, A., Rathore, M. M., &Jabbar, S. (2016). Smartbuddy: defining human behaviors using big data analytics in social internet of things. IEEE Wireless Communications, 23(5), 68–74.

[31]. Pereira, S., Pinto, A., Alves, V., & Silva, C. A. (2016). Brain tumor segmentation using convolutional neural networks in MRI images. IEEE Transactions on Medical Imaging, 35(5), 1240–1251.

[32]. Ramakrishnan, T., &Sankaragomathi, B. (2017). A professional estimate on the computed tomography brain tumor images using SVM-SMO for classification and MRG-GWO for segmentation. Pattern Recognition Letters, 94, 163–171.

[33]. Rathore, M. M., Ahmad, A., Paul, A., Wan, J., & Zhang, D. (2016). Real-time medical emergency response system: exploiting IoT and big data for public health. Journal of Medical Systems, 40(12), 283.

[34]. Ravı, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B., & Yang, G.-Z. (2017). Deep learning for health informatics. IEEE Journal of Biomedical and Health Informatics, 21(1), 4–21.

[35]. Roth, H. R., Lu, L., Liu, J., Yao, J., Seff, A., Cherry, K., … Summers, R. M. (2016). Improving computer-aided detection using convolutional neural networks and random view aggregation. IEEE Transactions on Medical Imaging, 35(5), 1170–1181.

[36]. Selvapandian, A., & Manivannan, K. (2018). Fusion based glioma brain tumor detection and segmentation using ANFIS classification. Computer methods and programs in biomedicine, 166, 33-38.

[37]. Shboul, Z. A., Reza, S. M. S., & Iftekharuddin, K. M. (2017). Quantitative mr image analysis for brain tumor. In European Congress on Computational Methods in Applied Sciences and Engineering (pp. 10–18). Springer.

[38]. Shi, J., Lei, Y., Wu, J., Paul, A., Kim, M., & Jeon, G. (2017). Uncertain clustering algorithms based on rough and fuzzy sets for real-time image segmentation. Journal of Real-Time Image Processing, 13(3), 645–663.

[39]. Shi, J., Wu, J., Paul, A., Jiao, L., & Gong, M. (2014). Change detection in synthetic aperture radar images based on fuzzy active contour models and genetic algorithms. Mathematical Problems in Engineering, 2014.

[40]. Shree, N. V., & Kumar, T. N. R. (2018). Identification and classification of brain tumor MRI images with feature extraction using DWT and probabilistic neural network. Brain Informatics, 5(1), 23–30.

[41]. Soltaninejad, M., Yang, G., Lambrou, T., Allinson, N., Jones, T. L., Barrick, T. R., … Ye, X. (2017). Automated brain tumour detection and segmentation using superpixel-based extremely randomized trees in FLAIR MRI. International Journal of Computer Assisted Radiology and Surgery, 12(2), 183–203.

[42]. Taheri, S., Ong, S. H., & Chong, V. F. H. (2010). Level-set segmentation of brain tumors using a threshold-based speed function. Image and Vision Computing, 28(1), 26–37.

[43]. Torti, E., Fontanella, A., Florimbi, G., Leporati, F., Fabelo, H., Ortega, S., &Callico, G. M. (2018). Acceleration of brain cancer detection algorithms during surgery procedures using GPUs. Microprocessors and Microsystems.

[44]. Tripathi, A. K., Sharma, K., &Bala, M. (2018). A Novel Clustering Method Using Enhanced Grey Wolf Optimizer and MapReduce. Big Data Research.

[45]. Vallabhaneni, R. B., & Rajesh, V. (2018). Brain tumour detection using mean shift clustering and GLCM features with edge adaptive total variation denoising technique. Alexandria Engineering Journal.

[46]. Venkatesan, C., Karthigai kumar, P., Paul, A., Satheeskumaran, S., & Kumar, R. (2018). ECG Signal Preprocessing and SVM Classifier-Based Abnormality Detection in Remote Healthcare Applications. IEEE Access, 6, 9767–9773.

[47]. Wang, X., Yang, L. T., Liu, H., &Deen, M. J. (2018). A big data-as-a-service framework: State-of-the-art and perspectives. IEEE Transactions on Big Data, 4(3), 325-340.

[48]. Wu, D., Zhu, L., Lu, Q., &Sakr, S. (2018). HDM: A composable framework for big data processing. IEEE Transactions on Big Data, 4(2), 150-163.

[49]. Xing, E. P., Ho, Q., Dai, W., Kim, J. K., Wei, J., Lee, S., ...& Yu, Y. (2015). Petuum: A new platform for distributed machine learning on big data. IEEE Transactions on Big Data, 1(2), 49-67.

[50]. Xue, Y., Chen, S., Qin, J., Liu, Y., Huang, B., & Chen, H. (2017). Application of Deep Learning in Automated Analysis of Molecular Images in Cancer: A Survey. Contrast Media & Molecular Imaging, 2017.

[51]. Zhao, D., Qiao, K., Zhou, Z., Li, T., Lu, Z., & Xu, X. (2017). Toward efficient and flexible metadata indexing of big data systems. IEEE Transactions on Big Data, 3(1), 107-117.

[52]. Zhang, Y., Zhao, G., Sun, J., Wu, X., Wang, Z., Liu, H., Govindaraj, VV., Zhan, T., & Li, J.(2018). Smart pathological brain detection by synthetic minority oversampling technique, extreme learning machine, and Jaya algorithm. Multimedia Tools and Applications, 77, 22629–22648.

[53]. Zhang, Y., Jiang, Y., Zhu, W., Lu, S., & Zhao, G. (2018). Exploring a smart pathological brain detection method on pseudo-Zernike moment. Multimedia Tools and Applications, 77(5), 1-15.