

Computational Predictions of MGMT Promoter Methylation in Gliomas: A Mathematical Radiogenomics Approach

Ayesha Agrawal¹, V. Maan²

¹Department of Computer Science & Engineering,
Mody University, India

²Department of Computer Science & Engineering,
Mody University, India

Emails: aishaagrawal33@gmail.com¹, vinodmaan.cet@modyuniversity.ac.in²

Article History:

Received: 20-04-2024

Revised: 12-06-2024

Accepted: 24-06-2024

Abstract:

In the treatment of glioblastomas, non-invasive methods for determining MGMT gene promoter methylation status are crucial due to their implications for chemotherapy responsiveness. This study utilizes a mathematical and computational framework to extract and analyze radiogenomic data from MRI images to predict the methylation status. The first step in our framework involves extracting radiogenomic data from MRI images. This process requires sophisticated image processing techniques to convert MRI scans into a format suitable for machine learning analysis. The features extracted include textural patterns, intensity distributions, and other relevant radiomic characteristics. To identify the most significant features, we employ a Random Forest (RF) algorithm. Mathematically, RF is an ensemble learning method that operates by constructing multiple decision trees during training and outputting the mode of the classes for classification tasks. The importance of each feature is evaluated based on its contribution to the accuracy of the model, quantified by metrics such as Gini impurity or information gain. Employing advanced machine learning models like VGG19, ResNet50, and Sequential FCNet Artificial Neural Networks, alongside traditional classifiers such as Naive Bayes and Logistic Regression, we analyze features identified by a Random Forest algorithm. Our mathematical approach ensures rigorous evaluation of model accuracy through sensitivity, specificity, and other performance metrics, presenting the Sequential FCNet ANN combined with ResNet50 as the superior model. This research contributes to the field of precision healthcare by enhancing the mathematical methods used in the non-invasive diagnosis of glioblastomas.

Keywords: Brain tumor ; radiogenomic ; classification ; MGMT promoter methylation ; magnetic resonance imaging (MRI).

1. INTRODUCTION

Glioblastomas (GBMs) are aggressive brain tumors that spread extensively and are resistant to chemotherapy. They often come back after surgery [1]. Studies have shown that methylation of the MGMT gene promoter, found in 30-60% of glioblastomas, can improve the response to TMZ. This methylation status is an important biomarker for predicting outcomes in GBM patients [2]. Currently, the only way to determine MGMT promoter methylation is through invasive brain biopsy or surgical resection [3].

Significant focus has been placed on creating non-invasive diagnostic techniques using images to identify MGMT promoter methylation status for glioblastoma [4]. Radiogenomics combines

radiological imaging with genomic data to potentially characterize gliomas and predict molecular markers without invasive procedures. The RSNA-MICCAI Brain Tumor Radiogenomic Classification challenge is at the forefront of using deep learning and machine learning to predict MGMT promoter methylation status from imaging data. The challenge lies in predicting MGMT promoter methylation status in gliomas due to the intricate relationship between molecular characteristics and radiological features.

Addressing multiple challenges is necessary to integrate radiomics and genomes for an in-depth knowledge of tumour biology and individualised treatment plans in precision healthcare. Utilising non-invasive imaging information for radiogenomic classification expedites the procedure, lowering operational concerns and suffering among patients while allowing for real-time tumour progression and treatment outcome tracking. The multidimensional feature spaces present in the genetic information, however, make feature selection and reduction of dimensionality difficult, especially when trying to find meaningful features that accurately reflect the methylation status of the MGMT promoter without overfitting. To avoid bias towards the majority class, it is essential to deal with inconsistencies in the distribution of methylation versus unmethylated MGMT promoter labelling within datasets.

Employing both machine learning and techniques from deep learning, the goal is to create a reliable model for prediction that can be used to identify the MGMT promoter methylation status of gliomas based on radiological imaging characteristics taken from MRI images. By assuring its capacity to generalise across a variety of patient populations, this approach will deal with the innate biological diversity and heterogeneity of brain tumours, including MGMT promoter methylation, and hence improve the accuracy of predictions. Harmonised preprocessing approaches will enable successful fusion by integrating MRI and genomic data from various sources. In addition, the goal encompasses finding and selecting significant radiomic characteristics linked to glioma MGMT promoter methylation status. In order to determine the highest-performing strategy, the performance of several algorithms based on deep learning and machine learning to predict the MGMT promoter methylation status in gliomas will be contrasted at the final stage. Convolution is a fundamental operation in the field of image processing, particularly within CNN architectures like VGG19 and ResNet50 used in your study. The operation involves sliding a kernel K over the input image I , calculating the dot product at each position. This operation extracts local feature representations which are essential for learning image characteristics relevant to identifying radiological signatures associated with MGMT methylation status.

Convolution Operation:

$$f(x, y) = (I * K)(x, y) = \sum_{i=-a}^a \sum_{j=-b}^b I(x-i, y-j)K(i, j) \quad (1)$$

The ReLU function introduces non-linearity to the network, which is crucial for learning complex patterns in the data. In the context of neural networks processing radiogenomic data, ReLU helps in stabilizing the learning and overcoming the vanishing gradient problem, which is critical when training deep neural networks on complex medical imaging data.

ReLU Activation Function:

$$g(z) = \max(0, z) \quad (2)$$

This loss function measures the discrepancy between the predicted probabilities and the actual labels. It is particularly effective for classification problems, providing a robust metric for optimizing the neural network in predicting the binary methylation status of the MGMT promoter.

Cross-Entropy Loss for Binary Classification:

$$L(y, \hat{y}) = -[y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})] \quad (3)$$

Gradient descent is a cornerstone optimization method used to update the weights θ in neural networks, aiming to minimize the loss function $J(\theta)$. The learning rate η controls the size of the steps taken towards the minimum of J , crucial for effective training of models on radiogenomic data.

Gradient Descent:

$$\theta_{\text{new}} = \theta_{\text{old}} - \eta \nabla_{\theta} J(\theta) \quad (4)$$

SGD optimizes the learning process by updating the model parameters using a randomly selected subset of the data at each iteration. This approach reduces the computational burden and can lead to faster convergence, especially beneficial in large datasets like those used in radiogenomic studies.

Stochastic Gradient Descent (SGD):

$$\theta = \theta - \eta \nabla_{\theta} J(\theta; x^{(i)}, y^{(i)}) \quad (5)$$

Batch normalization standardizes the inputs of each layer within the network to have zero mean and unit variance. This normalization helps in speeding up the training process and reducing the internal covariate shift, making the network more stable and efficient.

Batch Normalization:

$$\hat{x}^{(k)} = \frac{x^{(k)} - \mu_{\mu}}{\sqrt{\sigma_n^2 + \epsilon}} \quad (6)$$

the exponential of each output and then normalizing these values by the sum of all exponentials. This provides a probabilistic interpretation of the model's outputs, useful for the final classification in multilabel tasks, such as differentiating between various radiogenomic features

Softmax Function:

$$\sigma(z)_j = \frac{e^z}{\sum_{k=1}^K e^z} \text{ for } j = 1, \dots, K \quad (7)$$

the exponential of each output and then normalizing these values by the sum of all exponentials. This provides a probabilistic interpretation of the model's outputs, useful for the final classification in multilabel tasks, such as differentiating between various radiogenomic features

Logistic Regression Model:

$$p(y = 1 | x) = \frac{1}{1 + e^{-\sigma^T x}} \quad (8)$$

K-Nearest Neighbours (KNN) Distance Metric (Euclidean): KNN uses this distance metric to find the k closest training examples to a new data point x' classifying the point based on the majority label of these neighbours. This non-parametric method is invaluable for capturing nonlinear relationships in the data without assuming a specific form for the underlying model.

$$d(x, x') = \sqrt{\sum_{i=1}^n (x_i - x'_i)^2} \quad (9)$$

Naive Bayes Classifier:

$$P(y | x_1, \dots, x_n) = \frac{P(y) \prod_{i=1}^n P(x_i | y)}{P(x_1, \dots, x_n)} \quad (10)$$

Naive Bayes classifiers assume that the features are conditionally independent given the target class y and calculate the posterior probability of y given the features. This simplicity and efficiency make it particularly useful for initial explorations and baselines in complex predictive tasks involving high-dimensional data like radiogenomics.

Random Forest Gini Impurity:

$$I_G(p) = 1 - \sum_{i=1}^k p_i^2 \quad (11)$$

Gini Impurity is used within the Random Forest algorithm to decide the best split at each node of the decision trees. It measures the impurity of a node, with 0 representing a perfectly pure node. Minimizing the Gini Impurity helps in creating more accurate and robust decision trees by ensuring that each node splits the data as cleanly as possible, which is crucial for dealing with the heterogeneity of radiogenomic data.

Support Vector Machine (SVM) Optimization:

$$\min_{\theta} \frac{1}{2} \|\theta\|^2 + C \sum_{i=1}^m \max(0, 1 - y^{(i)}(\theta^T x^{(i)} + b)) \quad (12)$$

This formulation represents the optimization problem for training an SVM with a linear kernel. The first term regularizes the model by controlling the magnitude of the weight vector θ preventing overfitting. The second term, known as the hinge loss, penalizes misclassifications. C is a hyperparameter that balances these two aspects. This is especially valuable in radiogenomic data, where clear margins between classes can significantly enhance predictive performance.

Decision Tree Splitting Criterion (Information Gain):

$$IG(T, a) = H(T) - \sum_{v \in \text{values}(a)} \frac{|T_v|}{|T|} H(T_v) \quad (13)$$

Information Gain is a measure used in decision trees to choose the attribute that best separates the classes. It is calculated as the difference between the entropy of the parent set $H(T)$ and the weighted sum of the entropies of each subset $H(T_v)$. This measure helps in maximizing the effectiveness of each split in the tree, which is crucial for handling complex feature interactions in radiogenomic data.

Mean Squared Error (MSE):

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (14)$$

MSE is a common loss function used in regression problems and measures the average of the squares of the errors—that is, the average squared difference between the estimated values and the actual value (y_i). In the context of radiogenomics, although primarily a classification setting, regression frameworks might be employed in scenarios involving quantitative trait predictions.

Regularization (L2 Norm):

$$\Omega(\theta) = \|\theta\|^2 \quad (15)$$

L2 regularization adds a penalty equivalent to the square of the magnitude of coefficients to the loss function. This discourages large weights in the model, leading to simpler models that generalize better from training data to unseen data, thus preventing overfitting—a significant concern in high-dimensional data settings like radiogenomics.

Principal Component Analysis (PCA) Transformation:

$$Y = XW \quad (16)$$

PCA is used for dimensionality reduction by transforming the original variables into a new set of variables (principal components), which are uncorrelated and ordered so that the first few retain most of the variation present in all of the original variables. This transformation Y of the data matrix X using the weight matrix W is crucial for reducing the computational complexity and enhancing the interpretability of radiogenomic data analysis.

Feature Scaling (Standardization):

$$z = \frac{(x - \mu)}{\sigma} \quad (17)$$

Standardization involves rescaling the features so that they have the properties of a standard normal distribution. This is crucial in data preprocessing for machine learning algorithms to perform optimally, particularly in radiogenomics where features might have different units and scales.

AdaBoost Algorithm Update Weights:

$$w^{(i)} \leftarrow w^{(i)} \cdot e^{-\alpha y^{(1)} f(z^{(1)})} \quad (18)$$

AdaBoost is an ensemble technique that combines multiple weak classifiers to form a strong classifier. After each classifier is trained, the weights of incorrectly classified instances are increased so that subsequent classifiers focus more on difficult cases. This method is beneficial in radiogenomics, where the integration of diverse data types and complex patterns makes robust classification challenging.

Deep Learning Weight Initialization (He Initialization):

$$W \sim \mathcal{N}\left(0, \sqrt{\frac{2}{\text{fan_in}}}\right) \quad (19)$$

Proper initialization of neural network weights is critical for ensuring that the network converges quickly and reliably to a good solution. He initialization specifically helps in maintaining the variance

of activations across layers, which is crucial when training deep networks with ReLU activations, commonly used in analyzing medical images. In conclusion, our all-encompassing strategy that makes use of cutting-edge machine learning and deep learning methodologies, as well as meticulous feature selection and data integration procedures, promises to yield precise estimations of the MGMT promoter methylation status in gliomas. This will provide insightful information for customised treatment plans and enhanced outcomes for patients in the field of neuro-oncology.

2. LITERATURE REVIEW

In this work, glioblastoma patients' preoperative brain MRIs are used to identify genetic indicators using artificial intelligence. Models to predict nine genetic indicators were constructed using information from 400 patients. For a number of biomarkers, the combination of radiomics and CNN features produced good results, with ROC AUC >0.85 for some and >0.7 for others. For particular indicators, combining characteristics performed better than separate models, suggesting potential for noninvasive genetic prediction in gliomas [5].

This work investigates the use of MRI radiomics in deep learning techniques to predict MGMT promoter methylation in diffuse gliomas. Utilizing information from 111 patients, radiomic characteristics were taken from four different MRI sequences' discrete regions of interest and supplied into a residual network model. Models were trained and evaluated using five-fold cross-validation; area under the curve and average accuracy metrics were used to evaluate performance. With an AUC of 0.90 and an average accuracy of 0.91, the best-performing model, which was based on the ROI of the tumor core and included T1 contrast-enhanced and apparent diffusion coefficient data, showed superior diagnostic accuracy [6].

In this work, the authors evaluated the viability of using MRI to predict MGMT methylation in gliomas by extensively testing a number of new CNN designs. The thorough trials showed that, in terms of external validation accuracy, almost 80% of the constructed models did not significantly outperform chance levels. The RSNA-MICCAI challenge's top solution performed only moderately well, even after external validation. These results imply that, even with deep learning techniques, predicting the MGMT methylation status in gliomas using preoperative MRI may prove difficult, despite initial promise [7].

An attention-based deep learning network for MGMT methylation status prediction in GBM is presented in this paper. Existing techniques, which rely on radiomic studies, might be deficient in crucial elements needed for precise prediction. Although deep learning techniques have become popular, they frequently necessitate a thorough analysis of the data. The authors suggested a squeeze-and-sequence attention mechanism to rank pertinent slices and areas in order to remedy this. The best AUC of 70.59 was obtained after evaluation using different binary classification measures [8].

This study suggests a novel deep-learning-based classification approach for glioblastoma genetic subtyping that focuses on MGMT promoter methylation detection in order to speed up tumor identification and reduce patient anxiety. The project intends to investigate the viability of employing deep learning algorithms for predicting MGMT promoter methylation status, presenting possibilities for less intrusive diagnosis and treatment solutions. It does this by leveraging the RSNA-MICCAI dataset, which comprises structural multi-parametric MRI scans [9].

In order to improve prognosis, treatment response, and recurrence monitoring, this review emphasizes current developments in the use of magnetic resonance imaging (MRI) radiogenomics to evaluate genetic markers in GBM. Research shows that radiomic characteristics have high sensitivity and specificity; however, there are still issues with accuracy, reproducibility, and therapeutic value. Standardization and thorough data analysis should be given top priority in future initiatives to realize the full potential of this intriguing strategy [10].

This work investigates three deep learning-based methods for utilizing T2WI MRI scans to predict the methylation status of the O-6-methylguanine-DNA methyltransferase (MGMT) promoter in glioblastoma. For individuals with glioblastoma, MGMT status is a critical prognostic factor that influences therapy choices. Conventional techniques for MGMT status determination entail invasive tissue examination. The study employs the Brain Tumor Segmentation (BraTS) 2021 dataset to compare classification methods based on voxels, slices, and the entire brain. The accuracy of the voxel-wise method was 56.84% (SD 4.38%), but the accuracy of the slice-wise and whole-brain approaches was 61.37% (SD 1.48%) and 65.42% (SD 3.97%), respectively [11].

Accurate predictive models are essential for clinical decision-making, as evidenced by the discovery of MGMT promoter methylation as a critical prognostic marker. AI has the potential to improve medical decision support by utilizing technological breakthroughs, but integrating AI with electronic health records (EHRs) is still challenging work. The abundance of healthcare data available now facilitates individualized treatment techniques through precision medicine. In order to overcome these difficulties, this work uses transfer learning to train models on magnetic resonance imaging (MRI) in order to identify MGMT promoter methylation in glioblastoma [12].

The authors of this study [13] developed a brand-new two-stage MGMT Promoter Methylation Prediction (MGMT-PMP) method with the goal of precisely identifying glioblastoma genetic subgroups. In order to extract latent features fused with radiomic information and improve prediction capabilities, this method integrates a deep learning radiomic feature extraction module. To improve model performance, the authors successfully isolated negative training cases using a novice rejection strategy. Evaluations on the 2021 RSNA Brain Tumor Challenge dataset show exceptional performance in terms of classification, with 96.84% accuracy, 96.08% sensitivity, and 97.44% specificity in MGMT methylation status detection accomplished.

With the help of T2-w magnetic resonance imaging (MRI), a deep learning network called MGMT-net was created specifically for this study in order to assess the methylation status of the O6-Methylguanine-DNA Methyltransferase (MGMT) promoter in gliomas. The network was able to predict the MGMT methylation status of 247 participants with a mean cross-validation accuracy of 94.73%, sensitivity of 96.31%, and specificity of 91.66% by using brain MRI and genetic data. Furthermore, the network attained a mean Dice-score of 0.82 and a mean AUC of 0.93 for whole tumor segmentation [14].

3. METHODOLOGY

This section introduces the suggested approach for classifying brain tumors using deep learning across multiple modes. The method proposed includes five main steps- data pre-processing, feature

extraction, feature selection, training and testing the selected features on machine learning techniques and saving the model for future predictions.

The procedure starts with a thorough preliminary processing of the data to adjust image sizes, normalise characteristics, and solve the issue of unbalanced classes. Subsequently, detailed characteristics are extracted from the processed information using pre-trained VGG19 and ResNet50 models, which capture complex patterns. Then, to help with dimensionality reduction, the Random Forest Feature Selector identifies the most meaningful features. Ultimately, the characteristics that were selected are classified by supervised algorithms for machine learning such as KNN, Naive Bayes, Random Forest, Logistic Regression, and Sequential FCnet ANN, which provide accurate estimations of the target labels. Fig 1 illustrates the architecture for brain tumor radiogenomic classification to predict the status of genetic biomarker.

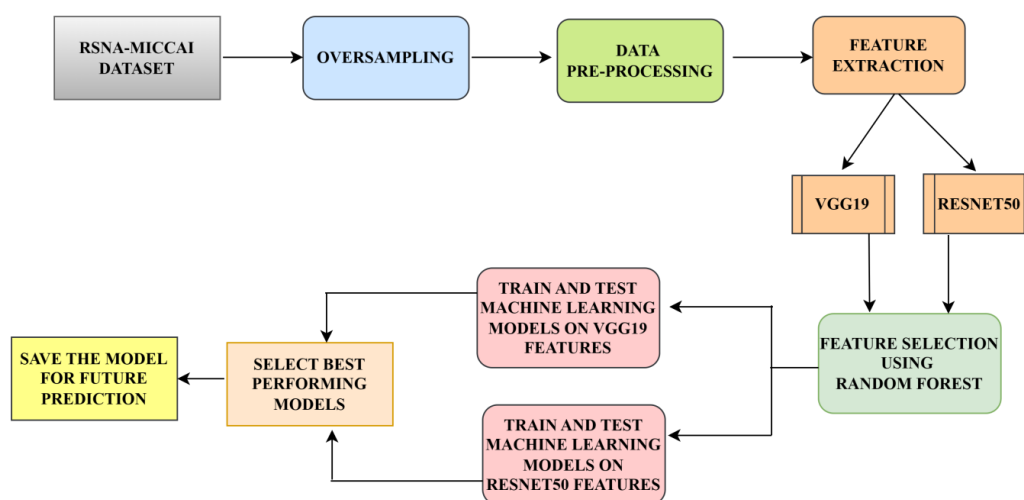


Fig 1: Proposed architecture for brain tumor radiogenomic classification

3.1 Dataset description

Using magnetic resonance imaging (MRI) scans, the RSNA MICCAI Brain Tumour Segmentation (BraTS) 2021 challenge aims to predict one of the common genetic features of glioblastoma (MGMT promoter methylation status) from pre-operative scans using MRI as well as automate tumour sub-region segmentation. The training cohort's organisational structure is as follows: Every single case has its own folder, identified by a five-digit number.

Each of the aforementioned "case" folders contains four sub-files, one containing each of the anatomical multi-parametric MRI scans in DICOM standard format. The following specific scans are included: Pre-contrast T1-weighted (T1w), post-contrast T1-weighted (T1Gd), FLAIR stands for Fluid Attenuated Inversion Recovery, and T2-weighted. There are 585 entries in the training data, each of which represents a particular individual or subject. The distribution of class labels for training data is shown in fig.2.

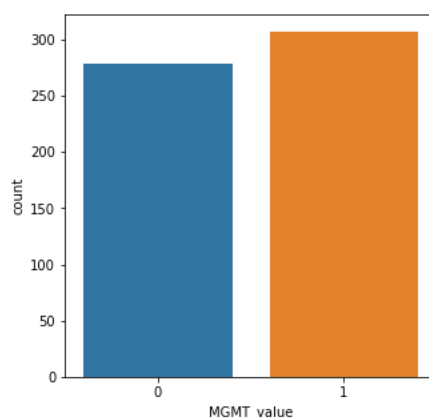


Fig 2: Distribution of class labels for training data

3.2 Data Pre-processing

Data pre-processing is a crucial initial step in preparing brain tumour datasets for efficient machine learning algorithm training. In order to guarantee dataset integrity and resilience, this work proposes a methodical strategy that includes image loading, scaling, normalisation, augmentation, and the remedy of data imbalances.

In the field of machine learning, random oversampling is a strategy that addresses imbalances in class by randomly replicating instances of minority classes until an even distribution has been obtained. In order to increase the minority class's presence in the dataset, it replicates instances that are chosen at random from the minority class. Because this strategy simply replicates data points that already exist, it prevents the development of bias. Random oversampling has the benefit of being straightforward and simple to apply, requiring little modification to current methods. Furthermore, by offering more equitable training data, it can enhance the efficiency of classifiers by improving generalisation and lowering bias towards the majority class [15].

3.3 Feature extraction

The features from the MRI images were extracted using two deep learning methodologies VGG19 and Resnet50.

3.3.1 VGG19

The VGG19 model is composed of 16 convolution layers, 19 ReLu activation layers, four pooling layers, three fully connected layers, and one Softmax layer for classification. By utilizing 3x3 convolutional filters, VGG19 can detect complex patterns and features within image data. The max-pooling layers help decrease input dimensions, reducing computational load. The fully connected layers at the end enable making predictions based on the high-level features identified by the convolutional layers. VGG19 employs the ReLU activation function for introducing non-linearity. Fig 3 illustrates the architecture of VGG19.

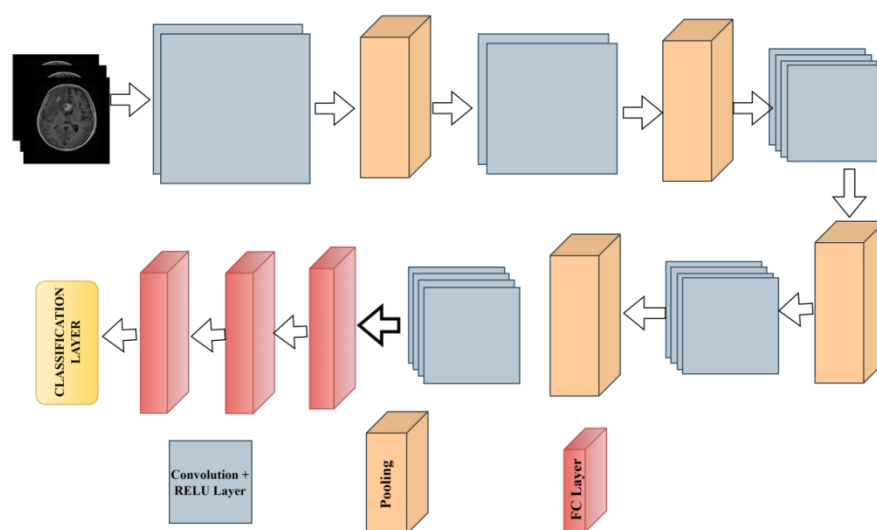


Fig 3: Architecture of VGG19

The VGG19 model is recognized for its simple and consistent design, which makes it straightforward to comprehend and adjust. It is commonly employed as a feature extractor and in transfer learning assignments for different computer vision task.

3.3.2 RESNET50

When facing challenges like disappearing or exploding gradients during training, He et al. introduced ResNet, a new residual network [18]. Fig 4 illustrates the architecture of RESNET50.

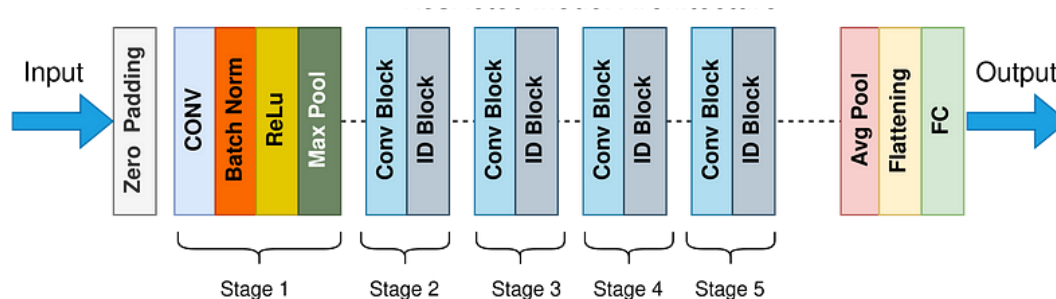


Fig 4: Architecture of RESNET50

ResNet's core idea is to incorporate an identity shortcut connection that skips certain layers, using skip connections to link specific layers and facilitate learning of residual mappings. It makes use of a method called skip connections, which connects skip connections between the architectures on up to three layers that incorporate batch normalisation and ReLU [19]. This approach allows the network to focus on matching the residual mapping instead of learning the underlying one. In simpler terms, by considering $H(x)$ as the target mapping for stacked layers starting with input " x ," ResNet aims to approximate the residual function $F(x) = H(x) - x$, leading to the final function $F(x) + x$.

ResNet-50 is a deep neural network consisting of 50 layers, allowing it to learn intricate features from input data for improved performance in computer vision tasks. The inclusion of residual connections in ResNet-50 addresses the vanishing gradient issue, resulting in higher accuracy for image classification tasks compared to earlier CNN models on standard datasets.

3.4 Feature Selection

Feature selection is a process of identifying and selecting the most important features in image datasets that helps to enhance the performance of machine learning models.

Utilizing Random Forest for feature selection falls within the embedded methods category, which merges the characteristics of filter and wrapper methods. A Random Forest comprises 4 to 1200 decision trees, with each tree constructed using a random selection of observations and features from the dataset. Not all trees see the same features or observations, ensuring that the trees are not correlated and less likely to overfit. Each tree consists of a series of yes-or-no questions based on one or more features. At each node (question), the tree splits the dataset into two groups, each containing observations that are more similar to each other and different from those in the other group. Therefore, the significance of each feature is determined by the level of "purity" in each group [20]. Random Forests typically rely on Gini Index to assess the importance of features which is defined in equation 1.

$$\text{Gini}(t) = 1 - \sum_{i=1}^k p(i/t)^2 \quad (1)$$

This metric is determined by how effectively each feature reduces impurity throughout all decision trees within the forest. Gini Importance values range from 0 to 0.5, with 0 indicating ideal purity (all elements assigned to the same class) and 0.5 indicating maximum impurity (labels evenly distributed across all classes). Now on the basis of feature importance score, select the features that are present at top of the tree as it provides more information. This technique is highly accurate and faster as compared to other techniques.

3.5 Feature fusion and Classification

Finally the features that are selected using feature selection technique from the MRI scans are fused with genomic feature i.e. MGMT promoter methylation into a single file. The fused features are passed to machine learning algorithms for the final classification.

3.5.1 K-nearest neighbour

The popular KNN algorithm is a supervised machine learning method mainly applied for classification tasks [21]. The algorithm involves a variable called k, which indicates the number of 'nearest neighbors'. In operation, KNN locates the closest data point(s) or neighbor(s) from a training set for a given query. It utilizes Euclidean distance to compute the distance between data points, as shown in the formula below in equation 2 [22]:

$$D(y, z) = \sqrt{\sum_i^n (y_i - z_i)^2} \quad (2)$$

Nearest data points are identified based on their proximity to the query point. Once the k nearest data points are determined, a majority voting rule is applied to determine the most frequently occurring class. The class with the highest frequency is designated as the final classification for the query.

K-NN is simple to grasp and use, making it suitable even for those unfamiliar with machine learning. It can manage data with irregular boundaries or non-linear connections between features and labels.

3.5.2 Naive Bayes

The Naive Bayes method simplifies the calculation of conditional probabilities by assuming that the features are independent given the class. Naive Bayes classifiers are created using Bayes' Theorem, consisting of algorithms that operate on the principle that each pair of features to be categorized is independent [23]. The Naive Bayes model is suitable for handling very large datasets and conducting additional analysis. It is a straightforward yet effective classification approach that delivers strong performance, even in complex situations [24]. Utilize Bayes' theorem to compute the posterior probability with the provided equation:

$$P(Y/X) = \frac{P(X/Y) P(Y)}{P(X)} \quad (3)$$

$P(X/Y)$ represents the posterior probability of a class, while $P(X)$ indicates the class's prior probability. $P(Y/X)$ shows the likelihood, which is the probability of a predictor given a class, and $P(Y)$ represents the prior probability. It is suitable for both binary as well as multi-class classification.

3.5.3 Random forest

Random Forest (RF) is a machine-learning technique that constructs multiple decision trees using training data sets to create a classification model. This method selects trees based on popular choices, resulting in high accuracy when handling large data sets [25]. The random forest (RF) is a collection of tree-based classifiers arranged hierarchically. Text data often have numerous dimensions, with many irrelevant attributes in the dataset. Only a few important attributes provide information for the classifier model. The RF algorithm utilizes a predetermined probability to select the most important relevant attribute. Breiman developed the RF algorithm by using sample data subsets to create multiple decision trees through random feature subspace sampling. The RF algorithm, linked to a set of training documents D and N features, can be outlined as follows:

(1) Initialization.

(2) A decision tree model is built for every document D_k . Training documents are chosen at random from the available features using a subspace of m -try dimension. Probabilities are calculated based on all m -try features. The best data split is determined by the leaf node. This process is repeated until the saturation criterion is met.

(3) Merge the K unpruned trees $h_1(X_1)$, $h_2(X_2)$,.....into a random forest group and utilize the highest probability value for making classification decisions.

By combining bagging and random selection, RF utilizes two feature selection methods to generate a more efficient ensemble model. Employing numerous trees within the RF approach reduces the risk of overfitting and decreases training time. Additionally, RF provides estimates for crucial classification variables and missing data, ultimately enhancing accuracy.

3.5.4 Logistic regression

Logistic regression is employed to determine the connection between one or more predictor variables and a binary outcome variable. A binary variable is a type of categorical variable that has only two possible values, such as 0 or 1. Logistic regression extends linear regression by assuming a linear relationship between the probability of the outcome and the independent variable (X), avoiding

predicted probabilities beyond 0-1 range. The intercept (b_0) and slope (b_1) of this line are represented by the regression coefficient which is shown in equation 4.

$$\ln(p/(1-p)) = b_0 + b_1X \quad (4)$$

Solving for the probability in this equation reveals a sigmoidal relationship with the independent variable, restricting the estimated probabilities to a range of 0 to 1 [26].

Logistic regression is effective even with small datasets, making it ideal for situations with limited or costly data collection. It supports regularization methods that prevent overfitting and enhance generalization by penalizing large coefficient values. Logistic regression is capable of managing datasets with numerous irrelevant or weakly correlated features.

3.5.5 Sequential FCNet Artificial neural network

Artificial neural networks, often referred to as neural networks, are mathematical or computational models inspired by the structure and function of biological neural networks. A neural network consists of a connected group of artificial neurons, which work together to solve particular problems [27]. The neuron operates in two modes: the training/learning mode and the using/testing mode. In most cases, an ANN is an adaptive system that adjusts its structure based on information from the network during the learning phase. Modern neural networks are tools for modelling non-linear statistical data, commonly used to represent intricate relationships between inputs and outputs or to identify patterns in data. The structure of feedforward neural network is shown in Fig 5.

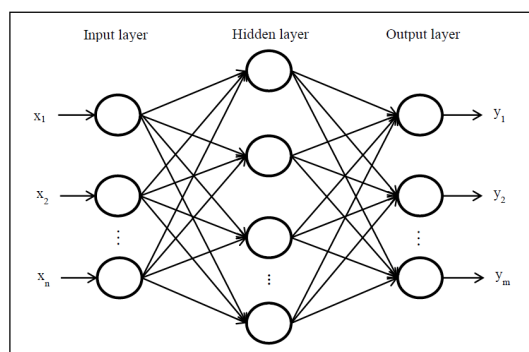


Fig 5: Structure of feedforward neural network

The back propagation learning algorithm is a key advancement in neural networks. It is a supervised learning method commonly used in multilayer feed-forward networks with continuous differentiable activation functions such as Tan-sigmoid and log-sigmoid. These networks, known as back-propagation learning networks (BPNs), adjust weights to categorize inputs correctly based on a provided set of training input-output pairs. This process involves a gradient-descent method similar to simple perceptron networks. By sending errors back to hidden units, the algorithm aims to find a balance between memorization and generalization. This balance enables the network to respond accurately to both familiar inputs and those that are similar yet distinct.

4. Performance evaluation metrics

When doing binary classification, data instances are usually predicted as either positive or negative. A positive label indicates the presence of illness, abnormality, or deviation, while a negative label means

there is no difference from the baseline. As a result, each binary label prediction can have four potential outcomes- a true positive (TP) occurs when a positive outcome is predicted correctly. On the other hand, a false positive (FP) happens when a negative instance is incorrectly predicted as positive. A true negative (TN) is when a negative outcome is predicted accurately; while a false negative (FN) is when a positive instance is wrongly predicted as negative.

The evaluation metrics most frequently used for binary classification are sensitivity, accuracy, precision, and specificity. These metrics indicate the percentage of correctly classified instances in the total set of instances, the actual positive instances, the actual negative instances, or the instances identified as positive, respectively [28]. Their formulas are mentioned in equation (4-7) as given below:

$$\text{Acc} = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

$$\text{Sen} = \frac{TP}{TP+FN} \quad (5)$$

$$\text{Spec} = \frac{TN}{TN+FP} \quad (6)$$

$$\text{Pre} = \frac{TP}{TP+FP} \quad (7)$$

In addition to accuracy, the mentioned metrics are commonly paired together, like precision and recall or sensitivity and specificity. Sensitivity and specificity can provide more insights into the model compared to accuracy, especially when there is a significant imbalance in the number of true positive and negative instances. Additionally, there are various evaluation metrics, such as the F1-score, that rely on the combined values of precision and recall. The formula for F1-score is mentioned in equation 8 [29].

$$\text{F1-score} = \frac{2*Pre*Sen}{Pre+Sen} \quad (8)$$

These measurements offer valuable information on various aspects of model performance, aiding in informed decision-making throughout the model evaluation process.

5. Result Analysis

The purpose of this analysis is to evaluate how well different predictive models perform by using various performance evaluation metrics.

5.1 VGG19 Image Classification with Various Classifiers

The analysis shows how different algorithms perform when using VGG19 features. Sequential Ann for VGG19 achieved the highest accuracy at 96.30%, followed closely by Random forest classifier at 95.58%. Naïve Bayes and KNN algorithms also did well, with accuracies of 93.93% and 92.80% respectively. Logistic regression had a slightly lower accuracy of 90.41% but still demonstrated good predictive ability.

Fig 6 displays the graphical representation of the accuracy of each model utilizing VGG19. The results highlight how different algorithms can accurately predict results using VGG19 features, providing useful information for improving models in the future.

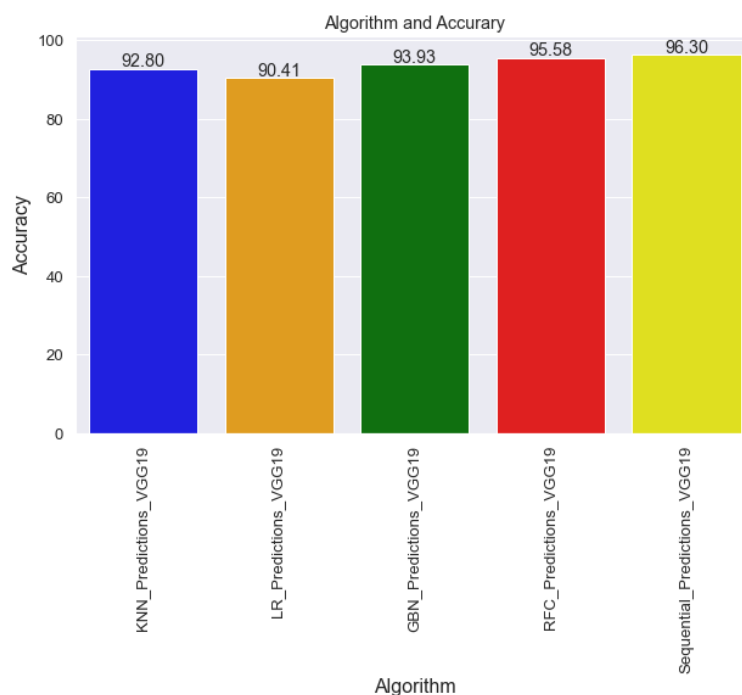


Fig 6: Graphical representation of the accuracy of each model utilizing VGG19

Analysis of algorithmic performance based on error rates shows different results for each model using VGG19 features. Similarly, Fig 7 displays the graphical representation of the error rate of each model utilizing VGG19.

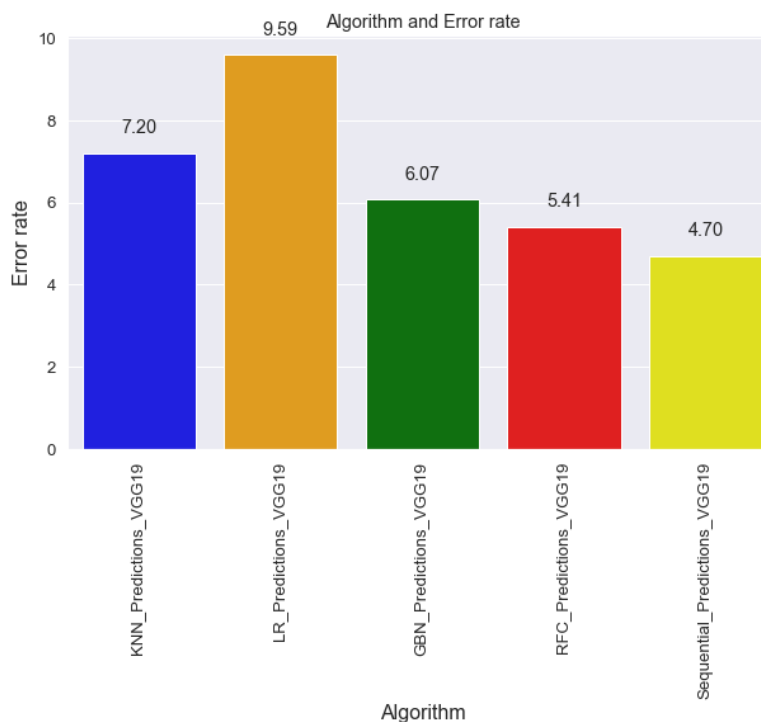


Fig 7: Graphical representation of the error rate of each model utilizing VGG19

Among them, Sequential_Predictions_VGG19 stands out as the most effective algorithm with the lowest error rate of 4.70%, showcasing its superior predictive accuracy. Following closely is

RFC_Predictions_VGG19, which demonstrates a competitive error rate of 5.41%, indicating strong performance in predictive modelling. GBN_Predictions_VGG19 displays a moderate error rate of 6.07%, showing consistent performance. However, LR_Predictions_VGG19 and KNN_Predictions_VGG19 have slightly higher error rates at 9.59% and 7.20% respectively, suggesting less precision in their predictions compared to the other models. Similarly, Fig 8 displays the graphical representation of the precision of each model utilizing VGG19.

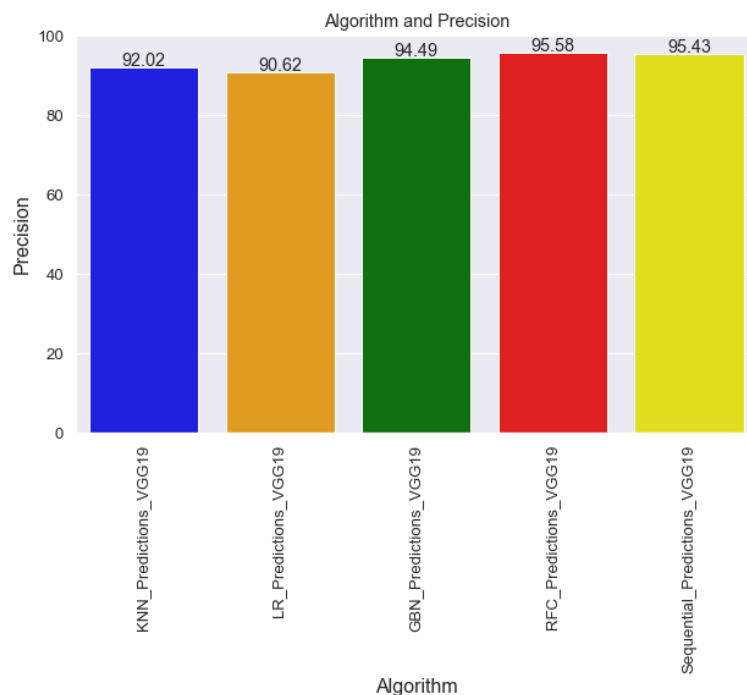


Fig 8: Graphical representation of the precision of each model utilizing VGG19

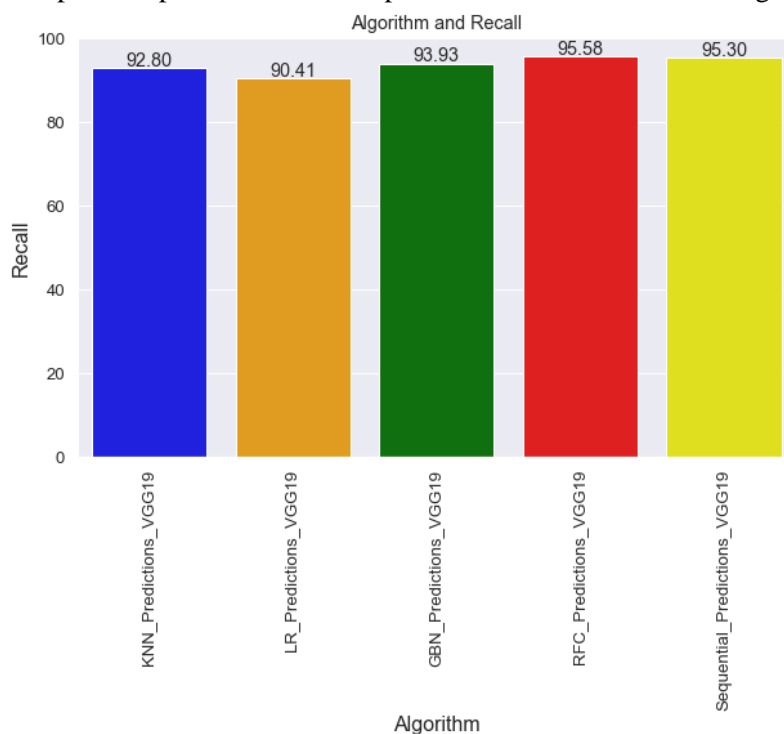


Fig 9: Graphical representation of the recall of each model utilizing VGG19

Similarly, Fig 9 displays the graphical representation of the recall of each model utilizing VGG19. The Random Forest Classifier (RFC) performs the best in terms of precision and recall among all models, showing its ability to accurately identify positive samples while reducing false positives. The Sequential ANN model comes next after RFC with slightly lower precision and recall values, indicating similar performance in classification tasks. The Gaussian Naive Bayes (GBN) also displays high precision and recall rates, showing its effectiveness in correctly classifying positive samples. The K-Nearest Neighbours (KNN) model exhibits decent precision and recall values, although slightly lower than RFC, ANN model, and GBN. Lastly, the Logistic Regression (LR) model shows slightly lower precision and recall rates compared to other models, suggesting relatively lower effectiveness in classification tasks.

The Random Forest Classifier (RFC) shows the highest F1-score compared to all models, demonstrating its effectiveness in balancing precision and recall. The Sequential ANN Model closely trails behind RFC with a slightly lower F1-score, indicating similar performance in terms of classification accuracy. Gaussian Naive Bayes (GBN) performs well, displaying a high F1-score and showcasing its ability in classification tasks. The K-Nearest Neighbours (KNN) model performs decently with a respectable F1-score, albeit ranking lower than RFC, Sequential Model, and GBN. The Logistic Regression (LR) model exhibits the lowest F1-score among the analysed models, showing comparatively lower effectiveness in classification tasks. The performance gap between RFC, Sequential Model, and GBN is minimal, implying that any of these models could be selected based on other factors such as computational complexity, interpretability, or specific task requirements. Fig 10 displays the graphical representation of the F1-Score of each model utilizing VGG19.

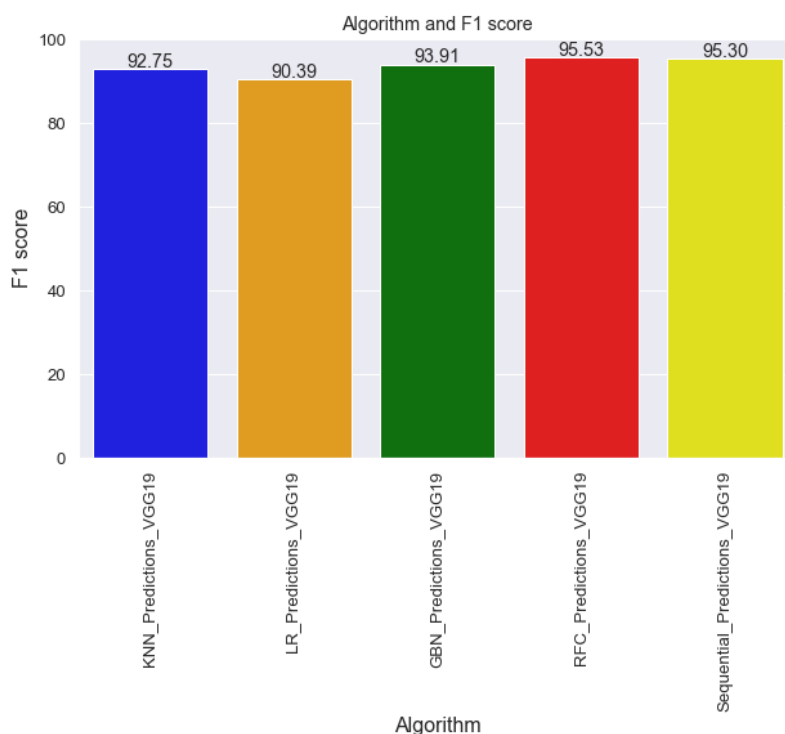


Fig 10: Graphical representation of the F1-score of each model utilizing VGG19

The examination of results shows significant differences in training times among various machine learning algorithms. Fig 11 displays the graphical representation of the training time of each model in utilizing VGG19.

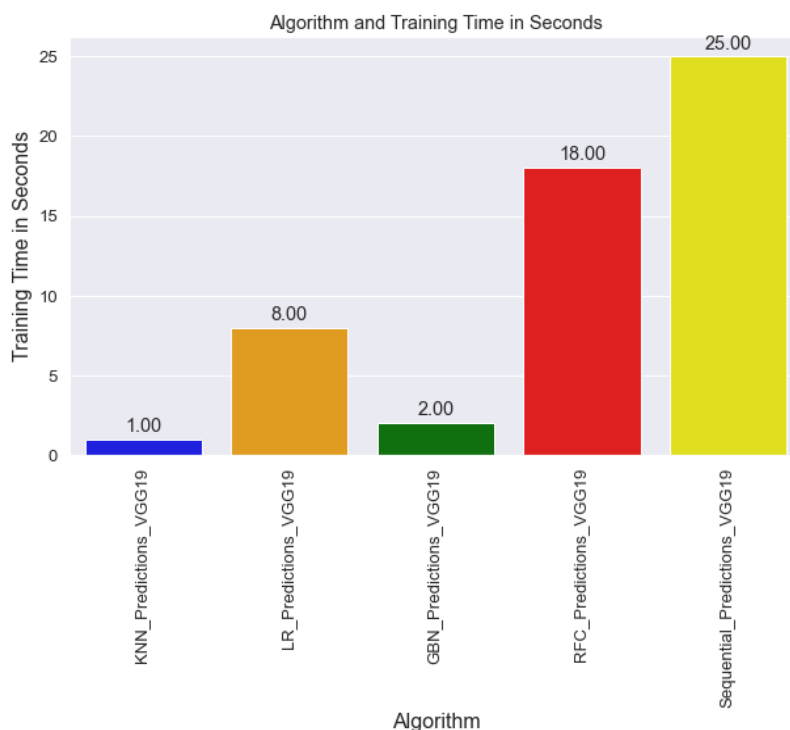


Fig 11: Graphical representation of the training time in seconds for each model utilizing VGG19

K-Nearest Neighbours (KNN) had the shortest training time at 1 second, highlighting its computational efficiency. Logistic Regression (LR) followed with a training time of 8 seconds, indicating its suitability for moderate-scale datasets. Gaussian Naïve Bayes (GBN) had a slightly longer training time of 2 seconds, showing its ability to handle more complex models than LR. Random Forest Classifier (RFC) had a notably longer training time of 18 seconds, possibly due to its ensemble nature and complexity. The Sequential model had the longest training time at 25 seconds, likely because of its deep learning architecture, which typically requires more computational resources and time to converge.

5.2 RESNET50 Image Classification with Various Classifiers

The results from the experiment on image classification using the RESNET50 model and various classifiers indicate significant differences in accuracy and error rates. K-Nearest Neighbours (KNN) achieved 93.9984% accuracy with a 7.001975% error rate, showing good performance with a slightly higher error margin. Logistic Regression (LR) had lower accuracy at 91.2526% and a higher error rate of 8.78475%. Gaussian Naïve Bayes (GBN) performed better with an accuracy of 94.5636% and a lower error rate of 5.436399%, suggesting its effectiveness in improving predictions.

The RFC achieved an accuracy of 96.78082% and an error rate of 4.219178%, demonstrating its effectiveness in image classification. On the other hand, the Sequential model surpassed all other classifiers by achieving the highest accuracy of 98.84765% and a minimal error rate of 2.152641879%.

This highlights its superior performance in image classification tasks when used with the RESNET50 model. Fig 12 displays the graphical representation of the accuracy of each model utilizing RESNET50.

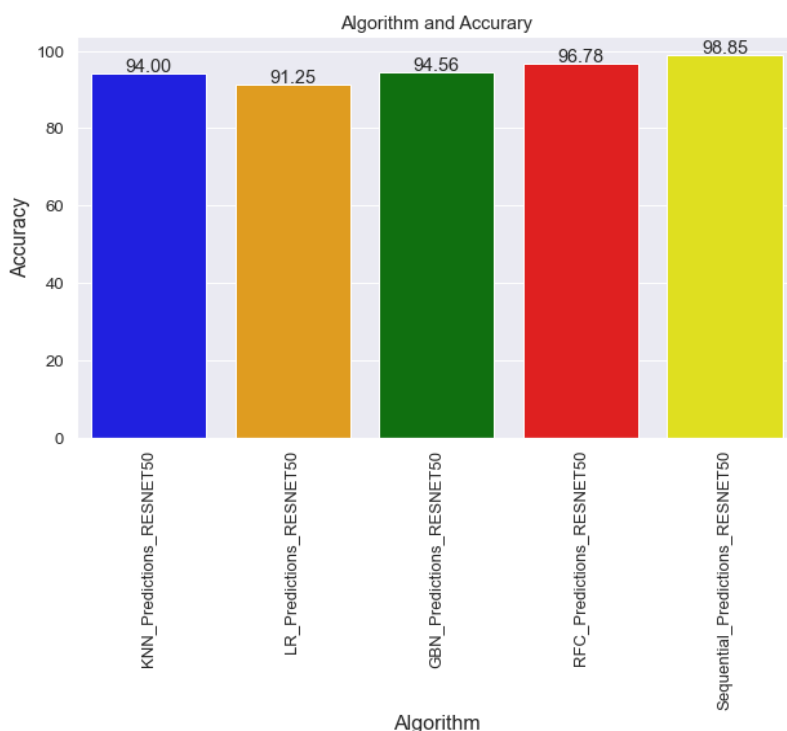


Fig 12: Graphical representation of the accuracy of each model utilizing RESNET50

Fig 13 displays the graphical representation of the error rate of each model utilizing RESNET50. These findings emphasize the importance of choosing the right classifiers to improve accuracy and minimize errors in image classification applications.

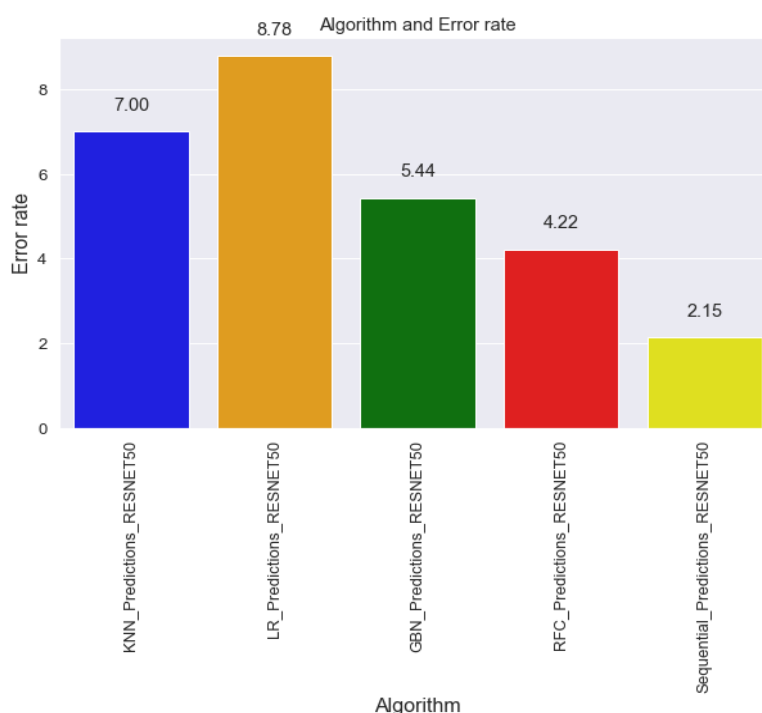


Fig 13: Graphical representation of the error rate of each model utilizing RESNET50

The examination of outcomes from the RESNET50 model using different classifiers, with a focus on precision and recall metrics, provides valuable insights into how well each classifier performs in image classification tasks. Fig 14 displays the graphical representation of the precision of each model utilizing RESNET50.

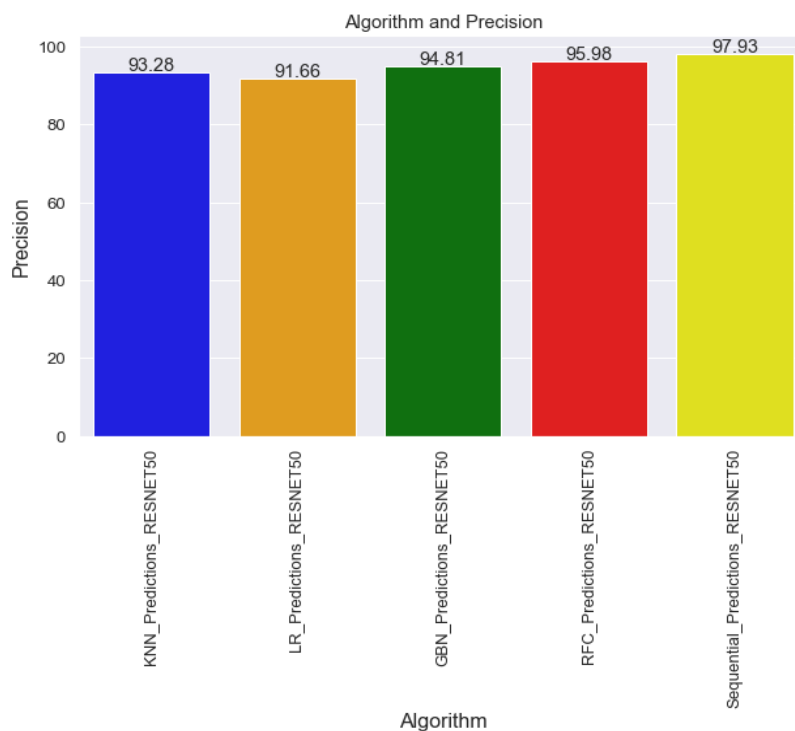


Fig 14: Graphical representation of the precision of each model utilizing RESNET50

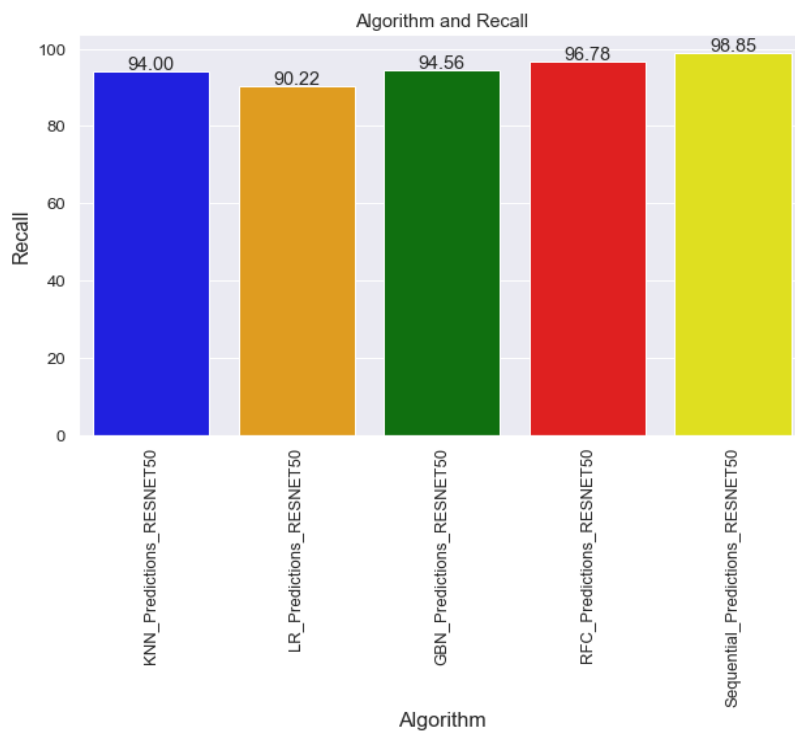


Fig 15: Graphical representation of the recall of each model utilizing RESNET50

Fig 15 displays the graphical representation of the recall of each model utilizing RESNET50. K-Nearest Neighbors (KNN) demonstrated a precision rate of 93.27631% and a recall rate of 93.99804%, indicating its capacity to accurately classify images while upholding a high precision in recognizing pertinent instances. Conversely, Logistic Regression (LR) exhibited a slightly lower precision rate of 91.65832226% and a recall rate of 90.21526419%, implying a moderate performance in accurately identifying relevant instances and reducing false positives.

Naive Bayes demonstrated enhanced recall (94.5636%) and precision (94.81437%), demonstrating its efficacy in correctly identifying cases while reducing false negatives. With greater precision and recall values of 95.97739% and 96.78082%, respectively, the Random Forest Classifier (RFC) proved its capacity to precisely group instances and retrieve pertinent examples from the dataset. With a maximum precision of 97.93441349% and a recall of 98.84765%, the sequential model surpassed the rest of the classifiers, demonstrating its greater proficiency in correctly detecting meaningful instances and minimising false positives and negatives. The significance of recall and precision metrics in assessing the effectiveness of classifiers for tasks such as image classification is highlighted by these findings, which provide well-informed selection according to particular requirements and limitations.

When analysing the results, the F1 score is a crucial metric for evaluating how well different models perform in image classification using the RESNET50 architecture. Fig 16 displays the graphical representation F1-score of each model utilizing RESNET50.

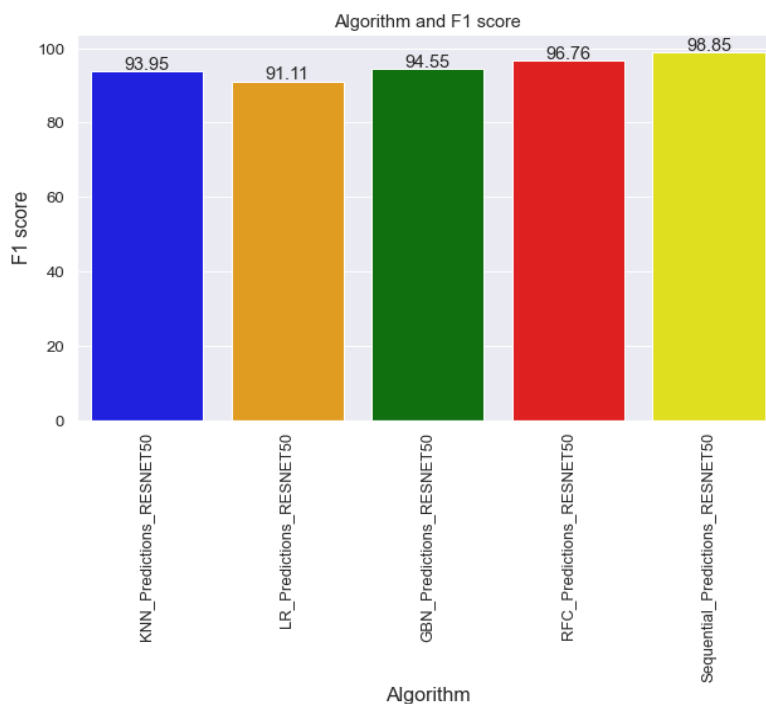


Fig 16: Graphical representation of the F1-score of each model utilizing RESNET50

The Sequential ANN model stands out with the highest F1 score of 98.84536% among the models assessed, showing a strong balance between precision and recall in image classification. The Random Forest Classifier (RFC) follows closely behind with an F1 score of 96.76431%, indicating reliable performance in identifying true positives while minimizing false positives and false negatives. The Gaussian Naive Bayes (GBN) model performs well with an F1 score of 94.54889%, proving its

effectiveness in classification tasks. In contrast, Logistic Regression (LR) and K-Nearest Neighbors (KNN) show slightly lower F1 scores of 91.10771% and 93.94778%, respectively, suggesting a less balanced precision-recall trade-off. These F1 scores reveal insights into the strengths and weaknesses of each model, helping with potential optimizations to improve overall classification performance.

The time it takes to train in seconds is a key measure for assessing how efficient different models are when using the RESNET50 design for image classification jobs. Fig 17 displays the graphical representation training time in seconds for each model utilizing RESNET50.

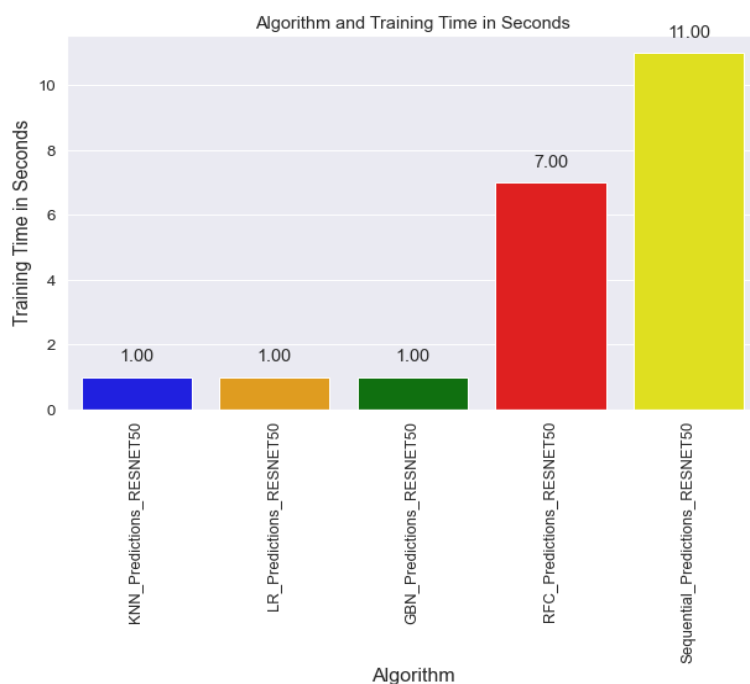


Fig 17: Graphical representation of the training time in seconds for each model utilizing RESNET50

Notably, the K-Nearest Neighbors (KNN), Logistic Regression (LR), and Gaussian Naive Bayes (GBN) models all have very short training times of just 1 second, highlighting their computational efficiency in training the model. Even though the Random Forest Classifier (RFC) is efficient, it takes about 7 seconds longer to train, possibly because of its ensemble learning approach and complexity. The Sequential model, which achieves the best classification performance, requires a longer training time of 11 seconds, likely due to its deep learning architecture and extensive parameter tuning. These insights on training times are important when choosing a model, especially in situations with limited computational resources or time constraints, helping researchers pick the most suitable model for their needs.

In conclusion, the analysis of the results reveals important findings about how well various models perform and how efficient they are when used for image classification tasks with VGG19 and RESNET50 architecture.

6. Conclusion and future scope

In conclusion, combining radiological imaging with genomic data via radiogenomics shows promise in enhancing the diagnosis and treatment of aggressive brain tumors such as glioblastomas. Recent

research has highlighted the potential of using deep learning and machine learning methods to non-invasively predict MGMT promoter methylation status, addressing a significant challenge in managing glioblastomas. By utilizing advanced techniques like VGG19 and ResNet50 for feature extraction, Random Forest for feature selection, and various classification algorithms for prediction, this study has successfully provided accurate estimations of MGMT promoter methylation status. The results emphasize that using Sequential FCnet ANN with ResNet50 is the most effective algorithm in this situation. Additionally, the model's robustness is enhanced by thorough preprocessing steps such as resizing images, normalizing features, and handling class imbalances.

Blending radiological and genomic characteristics in a cohesive framework is a crucial advancement for precision healthcare in neuro-oncology. This progress enables customized and efficient treatments for individuals with glioblastomas and other aggressive brain tumors. In the future, this research aims to improve and confirm the created model using bigger and more varied datasets to make it more widely applicable in clinical settings.

References

- [1] N. Q. K. Le, D. Thi, F. Y. Chiu, E. K. Y. Yapp, H. Yeh, and C.-Y. Chen, "XGBOOST improves classification of MGMT promoter methylation status in IDH1 wildtype glioblastoma," *Journal of Personalized Medicine*, vol. 10, no. 3, p. 128, Sep. 2020, doi: 10.3390/jpm10030128.
- [2] X. Chen et al., "Automatic prediction of MGMT status in glioblastoma via Deep Learning-Based MR Image analysis," *BioMed Research International*, vol. 2020, pp. 1–9, Sep. 2020, doi: 10.1155/2020/9258649.
- [3] C. G. B. Yogananda et al., "MRI-Based Deep-Learning Method for Determining Glioma MGMT promoter methylation status," *American Journal of Neuroradiology*, vol. 42, no. 5, pp. 845–852, Mar. 2021, doi: 10.3174/ajnr.a7029.
- [4] M. V. A. Rao, "Novel Local Radiomic Bayesian Classifiers for Non-Invasive Prediction of MGMT methylation status in glioblastoma," *arXiv (Cornell University)*, Nov. 2021, [Online]. Available: <http://arxiv.org/abs/2112.03259>.
- [5] E. Calabrese et al., "Combining radiomics and deep convolutional neural network features from preoperative MRI for predicting clinically relevant genetic biomarkers in glioblastoma," *Neuro-oncology Advances*, vol. 4, no. 1, Jan. 2022, doi: 10.1093/noajnl/vdac060.
- [6] S. Chen et al., "Predicting MGMT Promoter Methylation in Diffuse Gliomas Using Deep Learning with Radiomics," *Journal of Clinical Medicine*, vol. 11, no. 12, p. 3445, Jun. 2022, doi: 10.3390/jcm11123445.
- [7] B. Kim et al., "Validation of MRI-Based models to predict MGMT promoter methylation in Gliomas: BRATS 2021 Radiogenomics Challenge," *Cancers*, vol. 14, no. 19, p. 4827, Oct. 2022, doi: 10.3390/cancers14194827.
- [8] A. Iranmehr, S. Gutta, and A. Hadaegh, "Improved Prediction of MGMT Methylation Status in Glioblastoma using a Deep Attention Network," *2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, Nov. 2022, doi: 10.1109/iceccme55909.2022.9988694.
- [9] K. R. Spoorthy, A. R. Mahdev, B. Vaishnav, and S. M. L. J., "Deep learning approach for radiogenomic classification of brain tumor," *2022 IEEE 19th India Council International Conference (INDICON)*, Nov. 2022, doi: 10.1109/indicon56171.2022.10039760.
- [10] L. Gatto et al., "Beyond imaging and genetic signature in glioblastoma: Radiogenomic Holistic Approach in Neuro-Oncology," *Biomedicines*, vol. 10, no. 12, p. 3205, Dec. 2022, doi: 10.3390/biomedicines10123205.
- [11] S. Faghani, B. Khosravi, M. Moassefi, G. M. Conte, and B. J. Erickson, "A comparison of three different Deep Learning-Based models to predict the MGMT promoter methylation status in glioblastoma using brain MRI," *Journal of Digital Imaging*, vol. 36, no. 3, pp. 837–846, Jan. 2023, doi: 10.1007/s10278-022-00757-x.
- [12] Sakly H, Said M, Seekins J, Guetari R, Kraiem N, Marzougui M. Brain Tumor Radiogenomic Classification of O⁶-Methylguanine-DNA Methyltransferase Promoter Methylation in Malignant Gliomas-Based Transfer Learning. *Cancer Control*. 2023 Jan-Dec;30:10732748231169149. doi: 10.1177/10732748231169149. PMID: 37078100; PMCID: PMC10126792.

- [13] S. A. Qureshi *et al.*, “Radiogenomic classification for MGMT promoter methylation status using multi-omics fused feature space for least invasive diagnosis through mpMRI scans,” *Scientific Reports*, vol. 13, no. 1, Feb. 2023, doi: 10.1038/s41598-023-30309-4.
- [14] L. Robinet, A. Siegfried, M. Roques, A. Berjaoui, and E. C.-J. Moyal, “MRI-Based Deep Learning Tools for MGMT promoter methylation detection: a thorough evaluation,” *Cancers*, vol. 15, no. 8, p. 2253, Apr. 2023, doi: 10.3390/cancers15082253.
- [15] F. Kamalov, H. -H. Leung and A. K. Cherukuri, "Keep it simple: random oversampling for imbalanced data," *2023 Advances in Science and Engineering Technology International Conferences (ASET)*, Dubai, United Arab Emirates, 2023, pp. 1-4, doi: 10.1109/ASET56582.2023.10180891.
- [16] M. A. Khan *et al.*, “Multimodal brain tumor classification using deep learning and robust feature selection: a machine learning application for radiologists,” *Diagnostics*, vol. 10, no. 8, p. 565, Aug. 2020, doi: 10.3390/diagnostics10080565.
- [17] Mathivanan, S.K., Sonaimuthu, S., Murugesan, S. *et al.* Employing deep learning and transfer learning for accurate brain tumor detection. *Sci Rep* **14**, 7232 (2024). <https://doi.org/10.1038/s41598-024-57970-7>
- [18] He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
- [19] W. Islam, M. Jones, R. Faiz, N. Sadeghipour, Y. Qiu, and B. Zheng, “Improving Performance of Breast Lesion Classification Using a ResNet50 Model Optimized with a Novel Attention Mechanism,” *Tomography*, vol. 8, no. 5, pp. 2411–2425, Sep. 2022, doi: 10.3390/tomography8050200.
- [20] Iranzad, R., Liu, X. A review of random forest-based feature selection methods for data science education and applications. *Int J Data Sci Anal* (2024). <https://doi.org/10.1007/s41060-024-00509-w>.
- [21] S. Uddin, I. Haque, H. Lu, M. A. Moni, and E. Gide, “Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction,” *Scientific Reports*, vol. 12, no. 1, Apr. 2022, doi: 10.1038/s41598-022-10358-x.
- [22] Venkatesh and M. Leo, “MRI brain image segmentation and detection using K-NN classification,” *Journal of Physics: Conference Series*, vol. 1362, no. 1, p. 012073, Nov. 2019, doi: 10.1088/1742-6596/1362/1/01207.
- [23] V. Sheth, U. Tripathi, and A. Sharma, “A Comparative analysis of Machine learning Algorithms for classification purpose,” *Procedia Computer Science*, vol. 215, pp. 422–431, Jan. 2022, doi: 10.1016/j.procs.2022.12.044.
- [24] V. Jackins, S. Vimal, M. Kaliappan, and M. Y. Lee, “AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes,” *The Journal of Supercomputing*, vol. 77, no. 5, pp. 5198–5219, Nov. 2020, doi: 10.1007/s11227-020-03481-x.
- [25] K. Sumwiza, C. Twizere, G. Rushingabigwi, P. Bakunzibake, and P. Bamurigire, “Enhanced cardiovascular disease prediction model using random forest algorithm,” *Informatics in Medicine Unlocked*, vol. 41, p. 101316, Jan. 2023, doi: 10.1016/j.imu.2023.101316.
- [26] P. Schober and T. R. Vetter, “Logistic regression in medical research,” *Anesthesia & Analgesia*, vol. 132, no. 2, pp. 365–366, Jan. 2021, doi: 10.1213/ane.0000000000005247.
- [27] S. Jain, “Brain cancer classification using GLCM based feature extraction in artificial neural network,” *International Journal of Computer Science & Engineering Technology*, Jan. 2013, [Online]. Available: <http://www.ijcset.com/docs/IJCSET13-04-07-077.pdf>
- [28] M. Dehmer and S. C. Basak, *Statistical and machine learning approaches for network analysis*. 2012. doi: 10.1002/9781118346990.
- [29] Emmert-Streib, F., Moutari, S. & Dehmer, M. *Elements of Data Science, Machine Learning, and Artificial Intelligence Using R* (Springer, 2023).
- [30] S. K. A. Secure Algorithm for File Sharing Using Clustering Technique of K-Means Clustering. *IJRITCC* 2016, 4, 35-39.
- [31] Satish Kumar Alaria and Abha Jadaun. “Design and Performance Assessment of Light Weight Data Security System for Secure Data Transmission in IoT”, *Journal of Network Security*, 2021, Vol-9, Issue-1, PP: 29-41.
- [32] Pratiksha Mishra Satish Kumar Alaria. “Design & Performance Assessment of Energy Efficient Routing Protocol Using Improved LEACH”, *International Journal of Wireless Network Security*, 2021, Vol-7, Issue-1, PP: 17-33.