Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

Neuronet: A Multi-Symptom Analysis For Early Detection Of Gliomas Using Principal Component Analysis (PCA) And Artificial Neural Networks (ANN)

Nidhi Shakhapur[1], Sakshi Lele[2], Somya Mishra[3] and Dr. Sathya Karunanithi[4]

[1] Department of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India
[2] Department of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India
[3] Department of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India
[4]Department of Database Systems, Vellore Institute of Technology, Vellore, Tamil Nadu, India
Email: [1]nidhi.shakhapur2021@vitstudent. ac.in, [2]sakshi.lele2021@vitstudent.ac.in, [3]somya.mishra2021@vitstudent.ac.in, [2]sathya.k@vit.ac.in
Corresponding Author: Dr. Sathya Karunanithi*

Brain tumors pose a significant challenge to patients and healthcare providers, often complicating normal life due to their complex symptoms' identification. To enhance diagnostic accuracy and timely detection, we propose NeuroNet, an intelligent model, that utilizes the dimensionality reduction feature of Principal Component Analysis (PCA), integrated with Artificial Neural Networks (ANN) for predicting the presence of Gliomas by learning through vast clinical data of the patients. This model not only improves early detection and prognosis assessment but also optimizes decision-making for clinicians, ultimately enhancing patient outcomes. With 81.24% accuracy, the model is successfully able to predict the presence glioma and its type, thereby making it much more easier and time conserving for the clinicians to make quick decisions. The model's application can be expanded to other domains of medicine and healthcare, where it can be used to diagnose various other medical disorders or conditions. The advantage of AI-based systems is that the results produced are quick and within seconds. NeuroNet analyses every single symptom of the patient, which otherwise has a chance of being overlooked and missing a diagnosis. We leverage these assets of NeuroNet and Artificial Intelligence for medical diagnostics. The advancements of Artificial Intelligence highlight the critical need for interdisciplinary collaboration among Healthcare Industry and Technology Development..

KEYWORDS: Artificial Neural Networks (ANN), Clinical Symptoms, Gliomas, High Dimensional Data, Principal Component Analysis (PCA)

INTRODUCTION- Brain tumors represent a heterogeneous group of neoplasms originating within the central nervous system (CNS), posing significant challenges to diagnosis, treatment, and prognosis. Among these, gliomas are one of the most prevalent and clinically challenging subtypes, characterized by aggressive behavior and infiltrative growth patterns [39]. Understanding whether these symptoms are associated with gliomas or not, or what kind of glioma is causing the trigger to these symptoms, become very important questions. Research on early diagnosis of gliomas based on minor symptoms has been widely supported worldwide, so that the patient is given early care and their chances of success increase [35]. Key questions that remain in the context of AI application include how AI can enhance early detection and accurate prognostication of gliomas, best practices for integrating AI-driven models with existing diagnostic protocols, and how AI can be utilized to predict treatment responses and tailor personalized therapeutic strategies [37].

A. About Gliomas

Brain tumors are generally classified as primary and secondary tumors, with primary tumors originating within the brain and secondary tumors arising from metastatic spread. Primary brain tumors can be classified based on their cell of origin and histopathological features, with gliomas representing the most prevalent subtype from glial cells within the CNS. Gliomas are graded based on their histological characteristics and malignancy potential, ranging from World Health Organization (WHO) grade I to grade IV. Glioblastoma multiforme (GBM), a grade IV glioma, stands out as the most aggressive and lethal form, presenting significant therapeutic challenges. In this paper, we address four gliomas – Astrocytoma [26], Anaplastic Astrocytoma [27], Mixed Glioma (blend of Astrocytes and Oligodendrocytes) [28] and Oligodendroglioma [29].

Astrocytoma [26] is a malignant brain tumor that arises from astrocytes, which are the most common type of glioma, which usually develops in the brain and sometimes in the spinal cord. 60% of the tumors in the brain are glial tumors and they are responsible for a significant number of deaths and morbidities in both the young and the elderly. Astrocytoma(s) is the most common brain tumor. To eliminate or reduce the high morbidity and mortality rates associated with this condition, quick diagnosis and proper treatment are essential.

Mixed gliomas [28], also known as oligoastrocytomas, are the tumors of the brain that contain two different glial cells - oligodendrocytes and astrocytes in the mix. These disorders come with combined symptoms of oligodendrogliomas and astrocytoma, which leads to more complicated diagnosis and treatment. One of the techniques of treatment of mixed gliomas is combining surgery, radiation therapy, and chemotherapy, which are all personalized according to the genetic makeup of the patients involved and their overall health. Despite the progress, the management of mixed gliomas is still a major challenge due to their

heterogeneous nature, and hence there is a need for ongoing research to improve the health of the patients.

Oligodendrogliomas [29] (OD) are very rare and diffusely spread in the white matter of the brain. They tend to have better sensitivity and prognosis than other gliomas. Among the factors that can be connected with a less unfavorable outcome for the patient are low-grade, loss of both 1p/19q in the chromosome, and the patient has to have been younger or had a good performance status. Oligodendroglioma of lower grade usually introduces as seizures, and fast-growing tumors often represent themselves as internal brain pressure, or cognitive ducking. Afterward, it's recommended to apply radiation therapy for huge or embryonal tumors; anaplastic tumors; or areas that show signs of getting worse. The addition of the adjuvant PCV chemotherapy, which leads to better progression-free survival, does not increase overall survival compared to PCV administered at recurrence. Chemotherapy using either PCV or temozolomide is usually prescribed for the treatment of recurrent or progressive disease.

Anaplastic astrocytoma [27] (AA) is a malignant, diffusely infiltrating primary brain tumor originating from astrocytes. Currently, AA is classified based on histological features, though future classifications will incorporate molecular alterations. AA can be categorized into subgroups with similar molecular profiles, age at diagnosis, and median survival, determined by the status of 1p/19q co-deletion and IDH mutation. Chemotherapy is the current treatment for this type of tumor.

B. Applications of Artificial Intelligence in Medical Technology

Recent times have seen a drastic increase in research projects that delve into the root causes and mechanisms of various brain disorders like Multiple Sclerosis (MS), autism, dementia, Alzheimer's Disease (AD), gliomas, schizophrenia and epilepsy [30]. The AI has been now very commonly used in many areas of research but mostly in creating the latest developments of computer-aided diagnosis (CAD) systems by the application of image data and the manual help of a medical professional, which is a negative area, with the latter being a limiting factor in these cases, technology will help to understand the problem more accurately and will make the doctor 's diagnosis easier because of the automatic systems that are used.[30] The traditional ML-based CAD systems that the engineers come up with may use multiple learning techniques that often necessitate a lot of tuning and are just designed for some certain applications, which may find it hard to be used in cases that fall outside their training data. The development of AI methods in recent years, especially the end-to-end deep learning technique, and the evolution in the neuroimaging field (example: the use of diffusion-weighted MRI and other brain and nervous imaging technologies) have made new possibilities.[36] These developments not only improve the classical ML approaches but also they will introduce new ones for the brain diseases prediction and diagnosis.

Artificial intelligence (AI) has shown significant potential in improving surgical procedures, making them safer and more efficient. Research suggests that AI-assisted surgery can reduce operating time, docking time, staging time, and estimated blood loss compared to conventional surgery.[31] However, there is no significant difference in the length of hospital stay, recovery time, or the number of lymph nodes harvested between AI- assisted and traditional surgeries.[31] These findings indicate that AI-assisted surgery can enhance surgical

outcomes by minimizing complications, thereby offering a more effective alternative to conventional surgical methods.

C. Machine Learning and Deep Learning Models

Machine learning, a branch of artificial intelligence, has turned into a crucial machine which identifies complicated patterns within large biomedical data like the symptoms of patients or their genetic information.[2] Machine learning is a family of algorithms that train algorithms by learning from data rather than being programmed explicitly. It is a common approach in the fields of bioinformatics, text processing, gene expression data, and others. They take large and complicated data sets that may have underlying patterns and help find those patterns which can be used for predictive analytics. We find that the core of bioinformatics lies in unearthing hidden messages in data. Below, we enumerate the mutual health benefits brought about by maintaining one's microbiome.

Unlike normal statistical methods, machine learning algorithms are capable of learning from data to make predictions or decisions on their own without being programmed explicitly. The improvement of those projects leads to new devices and the subsequent development of various industries. The turning point was given by the increasing use of semi-automated machinery, while robotics laid the groundwork of the rapid industrial growth.

For many people, the main focus of their life is on being healthy, that is why people try to take better care of themselves and avoid potentially harmful things to stay healthy. That the brains of the human beings are the natural information processors has made the automation of the aforementioned processes even more effective. Neural networks play the role of a cherry on top of the machine learning solution. Unlike traditional regression approaches, ANN can model complex nonlinear relationships. It also offers excellent fault tolerance, speed, and scalability with parallel processing [25].

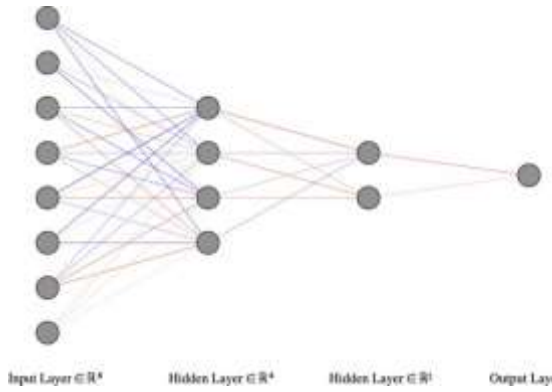


Fig 1. The visual representation of an Artificial Neural Networks structure

LITERATURE REVIEW: We have analyzed and reviewed the following papers that use Machine Learning and Deep Learning models for diagnosis and detection of various disorders that is used in medical technology [11][12][13][14][15].

Paper Number	Model Used	Accuracy	Paper Title
1	Random Forest	87.30%	Predicting Heart Disease Risk Factors Using Machine Learning Techniques
2	Support Vector Machine (SVM)	91.50%	Early Detection of Alzheimer's Disease Using Support Vector Machines
3	Convolutional Neural Network (CNN)	94.20%	Automated Diabetic Retinopathy Detection Using Deep Convolutional Neural Networks
4	Long Short-Term Memory (LSTM)	89.80%	Predicting Hospital Readmission Risk for Heart Failure Patients Using Long Short-Term Memory

			Networks

5	Gradient Boosting Machines (GBM)	92.60%	Predicting Stroke Risk with Gradient Boosting Machines
6	Logistic Regression	85.90%	Predicting Diabetes Onset Using Logistic Regression
7	K-Nearest Neighbors (KNN)	88.70%	Early Detection of Breast Cancer Using K-Nearest Neighbors
8	Decision Tree	90.10%	Predicting Parkinson's Disease Progression with Decision Trees
9	Gaussian Naive Bayes	84.60%	Predicting Kidney Disease Using Gaussian Naive Bayes

10	Recurrent Neural Network (RNN)	93.50%	Predicting Mortality in Intensive Care Unit Patients Using Recurrent Neural Networks

Although these models have shown impressive accuracy and precision, they can be prone to overfitting because they do not include algorithms specifically designed to prevent this issue. Overfitting occurs when a machine learning model learns too much from the training data, achieving high accuracy on that data but failing to generalize well to new data.[24] On the other hand, our integrated Artificial Intelligence model tackles this problem by incorporating a distinctive algorithm. While it may have slightly lower accuracy due to high dimensionality data and certain constrained system configurations, compared to some other models, it consistently delivers precise and reliable results, making it a more trustworthy option.[1]

METHODOLOGY: In this section, we outline the methodologies used in our study to improve predictive accuracy. We start with data pre- processing, emphasizing the division of data into training and testing sets to support model learning [3]. Next, we address dimensionality reduction using Autoencoders, a crucial step for handling high-dimensional datasets.[9] We then examine the architecture and activation functions of Artificial Neural Networks, which form the foundation of our predictive model [8]. Additionally, we discuss the optimizer, loss function, and metrics used in training the Artificial Neural Network [23]. Lastly, we cover the essential mathematical calculations that underpin our deep learning model [24].

A. Abbreviations and Acronyms

ADAM	Adaptive Moment Estimation
AI	Artificial Intelligence
ANN	Artificial Neural Networks
CPU	Central Processing Unit
DL	Deep Learning
GPU	Graphical Processing Unit
KOI	Kepler Object of Interest
ML	Machine Learning
PCA	Principal Component analysis
ReLU	Rectified linear activation function
TPU	Tensor Processing Unit

B. Theory And Calculations

Mathematics of Principal Component Analysis

PCA is a statistical method used for dimensionality reduction while preserving as much variance as possible in the dataset. The mathematics behind PCA involves several key steps, including standardization, covariance matrix computation, eigenvalue decomposition, and projection.

First, the input data needs to be standardized to have zero mean and unit variance [46].

$$X_{std} = \frac{X - \mu}{\sigma}$$

where is the X original data matrix, μ is the mean, and σ is the standard deviation of each feature.

Next, we compute the covariance matrix to understand how the features in the dataset vary with respect to each other.

$$C = \frac{1}{n-1} X_{std}^{T} X_{std}$$

where C is the covariance matrix, and n is the number of data points.

The covariance matrix is then decomposed through its eigenvalues and eigenvectors. Eigenvectors define the directions of the new feature space, whereas eigenvalues calculate their magnitude.

$$C\,v = \lambda v$$

where v is an eigenvector and λ is the corresponding eigenvalue. This equation essentially transforms the covariance matrix C into a set of eigenvectors and eigenvalues.

The principal components are then sorted in descending manner of their respective values. The top k eigenvectors form a new feature space.

$$W = [v_1, v_2, \ldots, v_k]$$

Where W is the matrix of selected eigenvectors.

Finally, the original data is projected onto the new feature space to obtain the lower-dimensional representation [47].

$$Y = X_{std} \cdot W$$

where Y is the transformed data in the new feature space.

Mathematics of Artificial Neural Networks

In artificial neural networks (ANNs), weights are represented as matrices and input data as vectors. The dot product of these matrices and vectors calculates the weighted sum of inputs, which is used to compute activations. The equation for matrix multiplication is:

$$C = A \cdot B$$

where C is the dot product of vectors A and B.

ANNs use vector operations like the dot product, element-wise multiplication, and addition to compute activations and gradients during forward and backward propagation. The dot product of a weight matrix and an input vector is used to compute activations. The equation for matrix-vector multiplication is:

$$\vec{c} = A \cdot \vec{v}$$

where c is the result vector, A is the weight matrix, and v is the input vector.

This equation updates the model's parameters to minimize the loss function. It is expressed as:

$$\theta = \theta - \eta \nabla_\theta J(\theta)$$

where,

θ represents the parameters of the model

η represents the learning rate of the model

$J(\theta)$ represents the loss function

$\nabla\theta\ J(\theta)$ represents the gradient of the loss function with respect to the model parameters.

Neural networks often apply the mathematics of Stochastic Gradient Descent (SGD). SGD updates the model parameters using a random subset of training data at every iteration:

$$\theta = \theta - \alpha \cdot \nabla_\theta J(\theta_i)$$

where θi is a random sample from the training data.

The chain rule is used to calculate the gradient of the loss function with respect to the model parameters. It is formulated as:

$$\frac{d}{dx}\left[f(g(x))\right] = f'(g(x)) \cdot g'(x)$$

where $f(x)$ and $g(x)$ represent any two functions, and $f'(x)$, $g'(x)$ represent the derivatives of $f(x)$ and $g(x)$ respectively.

Backpropagation involves calculating the gradient of the loss function [43] with respect to the model parameters using the chain rule. The gradient of the loss function with respect to the weights W1 and W2 can be calculated as follows:

$$\frac{dL}{dW_1} = \frac{dL}{dy} \cdot \frac{dy}{dW_1}, \quad \frac{dL}{dW_2} = \frac{dL}{dy} \cdot \frac{dy}{dW_2}$$

where dL/dy is the derivative of the loss function with respect to y, and dy/dW1 and dy/dW2 are the derivatives of y with respect to W1 and W2, respectively.
The updated parameters are calculated as:

$$W_1 = W_1 - \backslash textitlearning\_rate \cdot \frac{dL}{dW_1}$$

$$W_2 = W_2 - \backslash textitlearning\_rate \cdot \frac{dL}{dW_2}$$

where learning_rate is a hyperparameter that controls the size of the update.
The optimization algorithm updates the network's weights to minimize the loss function. Common algorithms include gradient descent, stochastic gradient descent (SGD), and ADAM [44] :

$$w_n = w_0 - \alpha \cdot \nabla \backslash textitLoss$$

where w_o and w_n are the old and new weights, respectively, and ∇Loss is the gradient of the loss with respect to the weights.
The categorical cross-entropy loss function [43] is given by:

$$L(y, \hat{y}) = -\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c} \log(\widehat{y_{i,c}})$$

where:

- N is the number of samples.
- C is the number of classes.
- $y_{i,c}$ is the true label for the i^{th} sample and the c^{th} class (one-hot encoded).
- $\hat{y}_{i,c}$ is the predicted probability for the i^{th} sample and the c^{th} class produced by the model.
- The summation is performed over all samples and all possible classes.
-

The SoftMax function [45] is formulated as:

$$softmax(z) = \frac{\exp(z_i)}{\sum(\exp(z_j))}$$

where z is a vector of arbitrary real-valued inputs and exp is the exponential function. The SoftMax function computes a probability distribution over K possible classes, where each class is represented by a node in the output layer of a neural network. The numerator of the formula computes the exponential of each input, and the denominator computes the sum of these exponentials. The resulting probabilities are then normalized such that they sum to 1.

C. Dataset and Materials

The dataset utilized in this project originates from a comprehensive collection of medical records detailing brain tumour symptoms. This dataset includes demographic information, symptoms, and treatment details of patients diagnosed with various types of brain tumors. The data encompasses parameters such as gender, age, ethnicity, biopsy site, histology, recurrence, prior treatment history, therapy type, and a range of symptoms including seizures, headaches, motor and sensory changes, and cognitive impairments.

Each entry in the dataset has been meticulously evaluated and categorized, providing a robust foundation for analyzing the prevalence and characteristics of different brain tumor types. The dataset serves as a critical resource for developing and validating our AI model which is aimed at improving the accuracy and reliability of brain tumor detection and
diagnosis. By efficiently utilizing this extensive dataset, we aim to identify key patterns and correlations that can inform effective diagnostic and therapeutic strategies for future patient care.

System Requirements

ARM-based CPU, 8-core processing chip with a base frequency of 3.2 GHz. The RAM Configuration comprises of 8 GB unified memory (configurable to 16 GB).
Additionally, the configuration can be upgraded to Integrated 8-core GPU for enhanced graphical processing and faster training and data processing.

Open Source Libraries

Pandas [b]: It is a python library which is used for analyzing, cleaning and manipulation of data. It uses proper and effective data structures to deal with the files easy.
NumPy [c]: It is a python library which is used to perform various operations on arrays and matrices according to the purpose
Seaborn [d]: It is a data visualization library in python which is built over Matplotlib to provide great visuals. It is used to plot the data and gives amplified results
Scikit Learn [e]: It is a robust python library which is used to create statistical modelling.

Algorithm Used

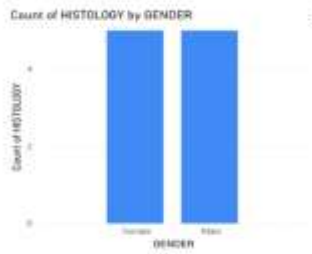
The below algorithm represents the working of the Artificial Intelligence model - NeuroNet.

Algorithm 1: NeuroNet – An Integrated Model of PCA and Neural Networks

1. **Procedure**: BrainTumorDetection
2. Get features X and target y.
3. **Initialise function**: Compute PCA.
4. **Start**: Initialise reduction of dimensionality of features.
5. Standardize the input data.
6. X_{std} is the standardized matrix.
7. Compute the covariance matrix C.
8. Perform eigen-decomposition.
9. Sort the eigenvectors in descending order of eigenvalues.
10. Project the data onto the new feature space using the matrix of selected eigenvectors.
11. **Build Neural Network Model**.
12. Define a sequential model.
13. Construct the dense layers.
14. The predicted output $y_{predicted}$ is computed using activation functions and weight matrices.
15. **Compute activation function**.
16. Compute the dot product of two vectors.
17. Update model parameters to minimize the loss function using Gradient Descent.
18. Update the model parameters θ, where θ represents the model parameters and η represents the learning rate.
19. Compute the gradient of the loss function using backpropagation.
20. Calculate the updated weights using the learning rate and the gradient of the loss function.
21. **Calculate the optimizer and loss function**.
22. The cross-entropy loss function is used to compute the error between the predicted output and the true labels.

GRAPHxCAL INTERPRETATIONS

The following section presents a detailed analysis of the dataset through various graphical representations. These visualizations provide insights into the distribution and relationships of different parameters, such as gender, symptoms, treatment history, and tumor types, among patients diagnosed with brain tumors. By examining these graphs, we aim to highlight significant patterns and correlations that can inform the development of effective diagnostic and therapeutic strategies. The visual analysis serves as a foundational step in understanding the characteristics of the patient population, ultimately contributing to the enhancement of brain tumor detection and treatment methodologies.



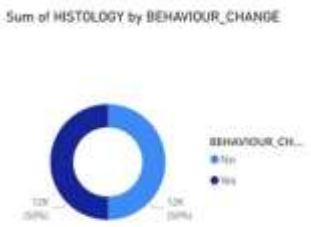
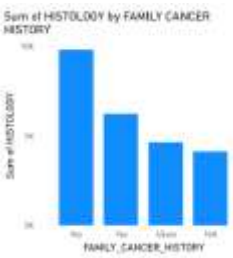


Figure 1(Left Hand Side): Sum of Histology by Gender **Figure 2 (Right Hand Side):** Sum of Histology by Behavior Change



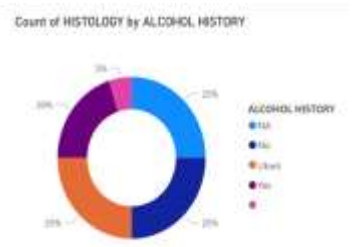
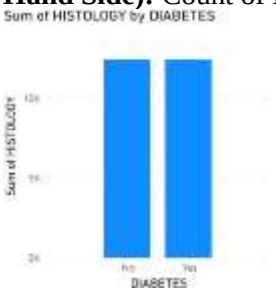


Figure 3 (Left Hand Side): Sum of Histology by Family Cancer History **Figure 4 (Right Hand Side):** Count of Histology by Alcohol History



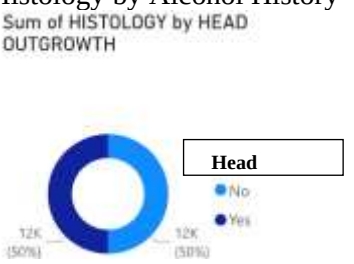


Figure 5(Left Hand Side): Sum of Histology by Diabetes **Figure 6 (Right Hand Side):** Sum of Histology by Head Outgrowth

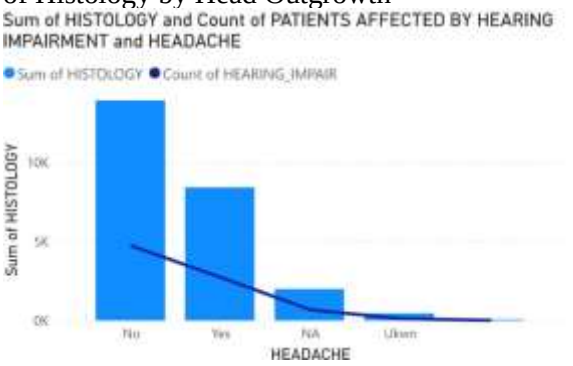


Figure 7(Top): Sum of Histology and Count of Patients Affected by Hearing Impairment and Headache

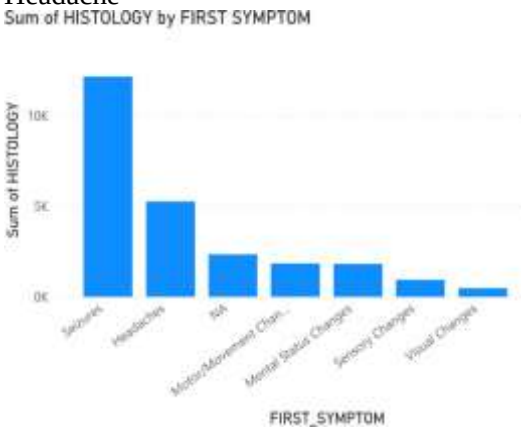


Figure 8(Bottom): Sum of Histology by First Symptom

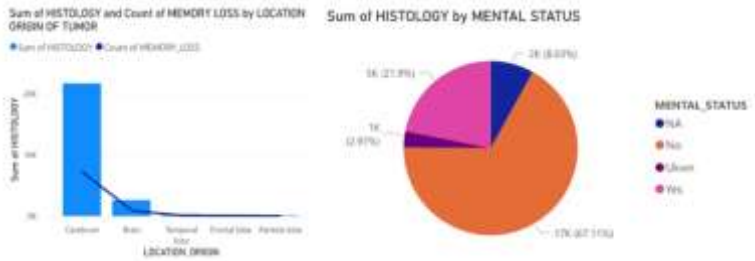


Figure 9 (Left Hand Side): Sum of Histology and Count of Memory Loss by Location of Tumor Origin **Figure 10(Right Hand Side):** Sum of Histology by Mental Status

The graphical analysis provides comprehensive insights into the demographic and symptomatic distribution of histology cases. Figure 1 shows an equal distribution of cases between genders, suggesting that gender does not influence the prevalence of histology types. Figure 2 also displays an equal split for behavior change, indicating it is not a definitive diagnostic symptom. In Figure 3, most cases lack a family cancer history, implying sporadic tumor occurrence, although some genetic influence cannot be ruled out. Figure 4 presents an even distribution of alcohol history, showing no significant correlation with histology types. Figure 5 indicates an equal distribution for diabetes, suggesting it is not a major factor. Similarly, Figure 6 reveals that head outgrowth is not a predominant symptom. Figure 7 highlights that headaches are not common among patients with hearing impairments. Figure 8 shows seizures and headaches as the most common first symptoms, emphasizing their importance in early diagnosis. Figure 9 demonstrates that most

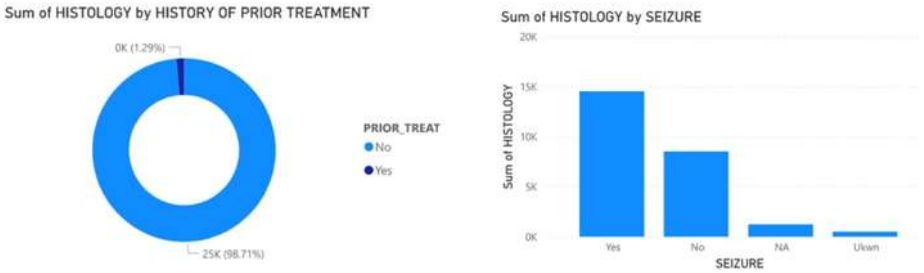


Figure 11 (Left Hand Side): Sum of Histology by History of Prior Treatment Figure 12 (Right Hand Side): Sum of Histology by Seizure

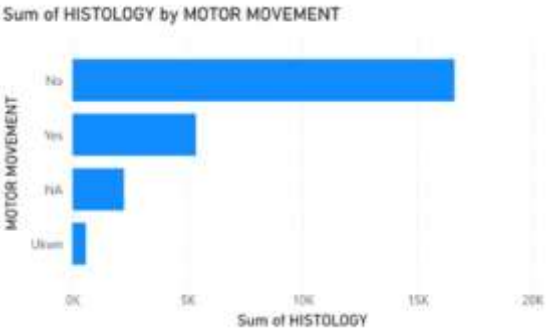


Figure 13: Sum of Histology by Motor Movement

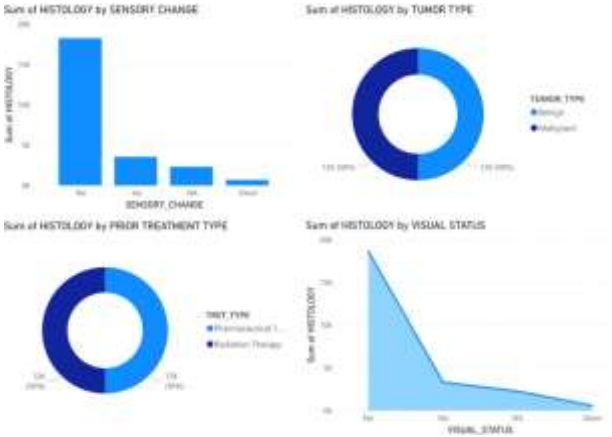
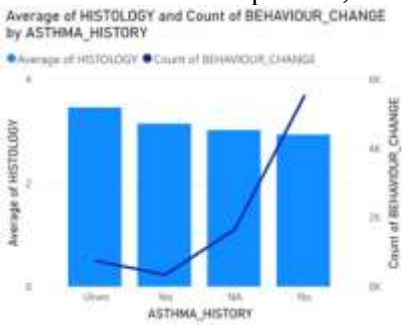


Figure 14 (Top Left-Hand Side): Sum of Histology by Sensory Change **Figure 15 (Top Right-Hand Side):** Sum of Histology by Tumor Type
Figure 16 (Bottom Left-Hand Side): Sum of Histology by Prior Treatment Type **Figure 17 (Bottom Right-Hand Side):** Sum of Histology by Visual Status

The graphical analysis of Figures 11 through 17 provides insights into the various aspects of the patient data. Figure 11 reveals that the vast majority of patients (98.71%) have no history of prior treatment, suggesting most are newly diagnosed. Figure 12 shows that seizures are a

common symptom, with 15K cases reporting seizures, highlighting the need for careful monitoring. Figure 13 indicates that motor movement issues are not prevalent, but a significant number of patients (5K) do experience these problems, pointing to the tumor's impact on motor functions. Figure 14 shows that most patients (15K) do not experience sensory changes, making it an uncommon symptom. Figure 15 reveals an equal distribution of benign and malignant tumors, suggesting a balanced occurrence in the dataset. Figure 16 indicates an equal utilization of pharmaceutical and radiation therapies, reflecting diverse treatment approaches. Finally, Figure 17 shows that visual changes are not prevalent, with most patients (15K) reporting no visual changes. These visualizations underscore the importance of comprehensive data analysis to understand the characteristics and treatment needs of brain tumor patients, informing effective diagnostic and therapeutic strategies.



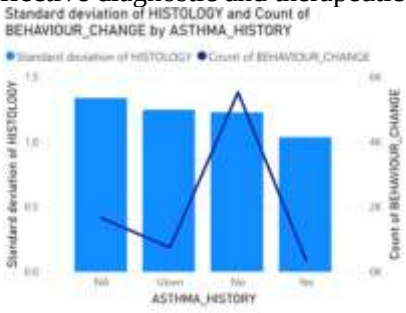
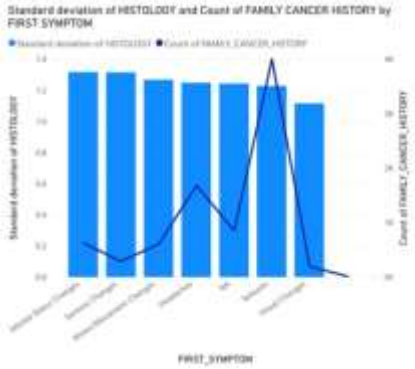


Figure 18(Left Hand Side): Average of Histology and Count of Behavior Change by Asthma History **Figure 19 (Right Hand Side):** Standard Deviation of Histology and Count of Behavior Change by Asthma History



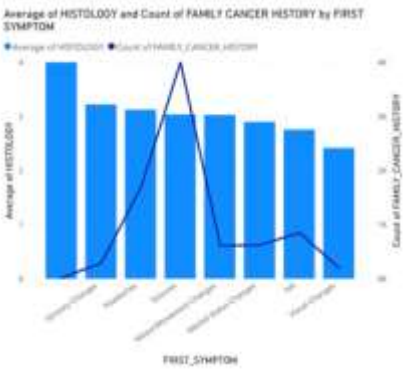


Figure 20 (Left Hand Side): Standard Deviation of Histology and Count of Family Cancer History by First Symptom **Figure 21 (Right Hand Side):** Average of Histology and Count of Family Cancer History by First Symptom

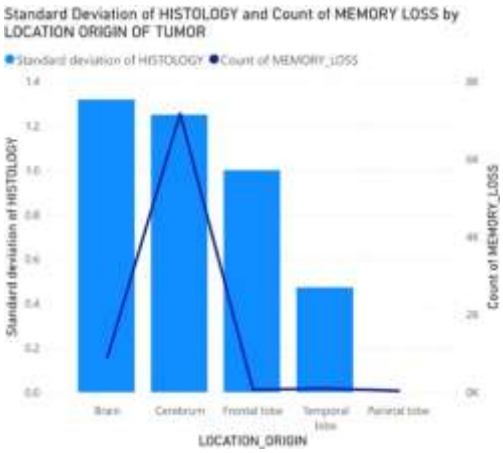


Figure 24: Standard Deviation of Histology and Count of Memory Loss by Location of Tumor Origin

The graphical analysis of Figures 18 through 24 provides valuable insights into the relationship between various symptoms and demographic factors. Figure 18 shows that the average histology and count of behavior changes are relatively consistent across different asthma histories, suggesting no strong correlation. Figure 19 reveals similar trends in the standard deviation of histology and behavior changes, indicating that asthma history does not significantly affect behavior changes. Figure 20 highlights that the standard deviation of histology and family cancer history is consistent across first symptoms, with a notable peak for mental status changes, suggesting variability in this symptom group. Figure 21 shows that the average histology and family cancer history are relatively consistent across first symptoms, indicating no strong correlation between family cancer history and specific initial symptoms. Figure 22 reveals that the standard deviation of histology and memory loss varies with the presence of headaches, with the highest variability observed in patients without headaches, suggesting that memory loss is more variable in this group. Figure 23 shows that the average histology and count of memory loss are higher for tumors originating in the cerebrum and parietal lobe, highlighting the cognitive impact of tumors in these locations. Figure 24 reveals that the standard deviation of histology and memory loss is highest for tumors in the brain and cerebrum, indicating greater variability in cognitive symptoms for these tumor origins. These visualizations underscore the importance of considering multiple factors when analyzing brain tumor characteristics and symptoms, as well as the need for comprehensive data analysis to inform effective diagnostic and treatment strategies.

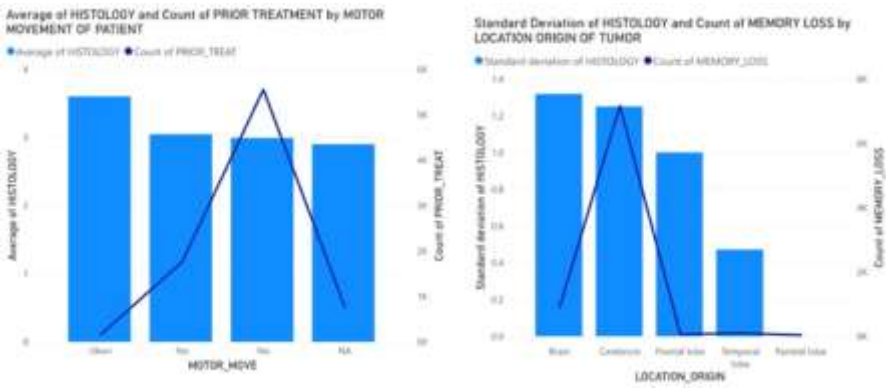


Figure 25 (Left Hand Side): Average of Histology and Count of Prior Treatment by Motor Movement of **Patient Figure 26 (Right Hand Side):** Standard Deviation of Histology and Count of Memory Loss by Location Origin of Tumor

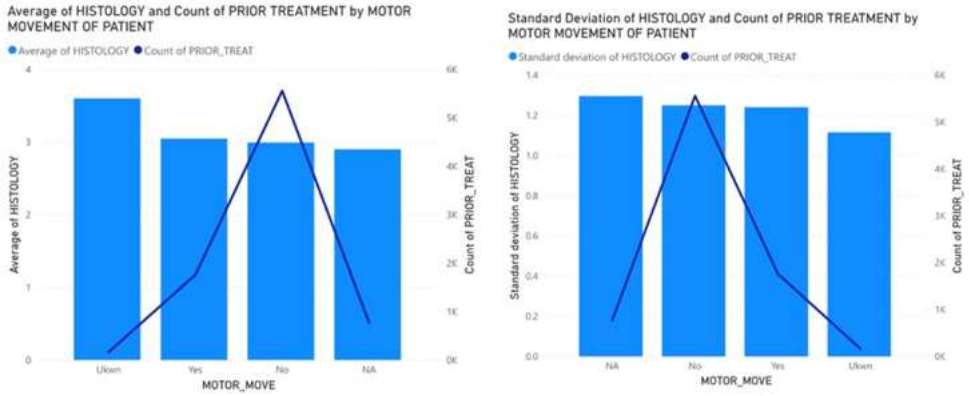


Figure 27 (Left Hand Side): Average of Histology and Count of Prior Treatment by Motor Movement of Patient **Figure 28 (Right Hand Side):** Standard Deviation of Histology and Count of Prior Treatment by Motor Movement of Patient

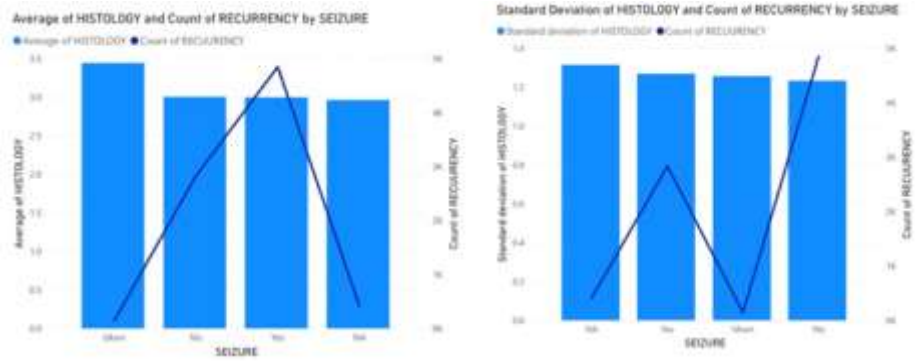


Figure 29 (Left Hand Side): Average of Histology and Count of Recurrence by Seizure Figure 30 (Right Hand Side): Standard Deviation of Histology and Count of Recurrence by Seizure

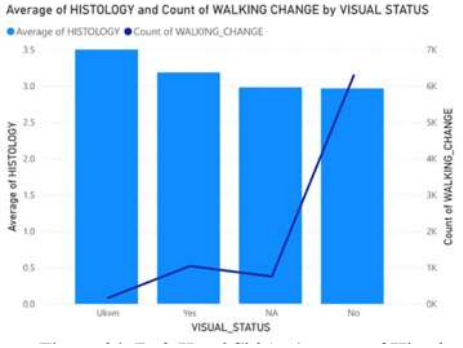


Figure 31 (Left Hand Side): Average of Histology and Count of Walking Change by Visual Status **Figure 32 (Right Hand Side):** Standard Deviation of Histology and Count of Walking Change by Visual Status

The graphical analysis of Figures 25 through 32 provides insights into the relationship between treatment history, motor movement, tumor location, and symptoms. Figures 25 and 27 indicate that the average histology and count of prior treatment arerelatively consistent across different motor movement statuses, with a notable peak for unknown motor movement. This suggests no strong correlation between motor movement and prior treatment. Figures 26 and 28 show that the standard deviation of histology and memory loss is highest for tumors originating in the brain and cerebrum, indicating greater variability in cognitive symptoms for these tumor origins. Figures 29 and 30 reveal that the average histology and count of recurrence are higher for patients with seizures, highlighting the importance of monitoring seizures for recurrence risk. Figures 31 and 32 indicate that the average histology and count of walking changes are higher for patients with no visual changes, suggesting a potential link between visual status and mobility. The standard deviation data in these figures also show variability in walking changes based on visual status.

These visualizations emphasize the importance of considering multiple factors when analyzing brain tumor characteristics andsymptoms, as well as the need for comprehensive data analysis to inform effective diagnostic and treatment strategies.

RESULTS: Our AI Model showed approximately 81.24% accuracy, which stands out as well. Despite the fact that the model is actually capable of being more accurate, its present arrangement suggests an enormous number of parameters as well as 8,256 patients on whom data was collected, but it still remains viable from a technical standpoint.[2] The great number of patient as well as the scenario complexity of data flow are the leading issues, which in turn makes the path to high accuracy quite challenging. By entangling methods of the kinds such

as regularization for overfitting avoidance, dropout, PCA as a test feature reducer and cross-validation, the model managed to keep a balance in terms of accuracy and generalization.
The complexity and model performance improvement were dealt with by having some of the most advanced methodologies included. The following are the highlights of some of these applications.
[4] Regularization is a technique used to keep models from overfitting. It is the situation when a model performs well with the training data but after performance is released on the actual test set, it performs very poorly.

L1 Regularization is known as Lasso regularization is the extra cost of the absolute value of the size of coefficients to be added to the loss function, and thus, it favors models with fewer coefficients that still can predict data accurately.
Ridge regularization which is a lesser-used name L2 regularization, it means adding a penalty equal to the square of the magnitude of coefficients which in turn causes all coefficients to become small and thus a simple model is obtained.
An overview of these regularization techniques is that they introduce a penalty on the loss function proportional to the magnitude of the model coefficients, hence controlling the large weights that bring about overfitting. In our model, [28] Dense layers have both L1 and L2 regularization which keep the weights small and by that means reduce the likelihood of the model becoming too entangled.

Dropout is one more regularization method in which some neurons are deliberately disconnected randomly during the training. It should be noted that in each training step, different sets of neurons that are "dropped out" or not used are created one after another, making them independent of the existing weights. Therefore, the network can be learned more general features that are essentially valid across different random subsets of the rest of the neurons. Here, the dropout layers are appended to the existing Dense layers. Hence, by switching on only a group of haphazard neurons in the course of training, dropout enhances the generalization of the model, and thus performance on the new data.
PCA is a dimensionality reduction method which represents the data in a set of orthogonal (uncorrelated) components, which are the original variables weighted with some coefficients. This method is primarily executed to reduce the number of features but still maximize the variability of the data. The negative effect of the dimensionality reduction technique can be the loss of the randomness and other irrelevant characteristics, which can cause the training set to be time-consuming through the elimination of noise and irrelevant features. Moreover, such operations can possibly improve the model's accuracy.
Cross-validation is a technique used to measure the model's generalizability. The steps involved in it are parallelly splitting the data into folds, training on each fold, and testing on the rest of the data. This way the model's performance does not correspond to only a single split of the dataset and also a more robust estimate of the accuracy and generalization is reached.
The main objective of the method is to minimize the train set's accuracy and maximize the generalization power of the model. A high training accuracy might be a sign of overfitting where the learning process unwinds the training data too much but can't extend it to new data. The measures such as regularization, L1 and L2, and dropout layers are employed which in

turn decreases the effect of overfitting, makes the model more robust and more widely applicable to the new data.

In summary, our AI model, with an accuracy of 81.24%, effectively handles the complexity of a high-dimensional dataset from a large number of patients. The use of regularization, dropout, PCA for dimensionality reduction, and cross-validation helps in balancing accuracy and generalization, ensuring that the model performs well not only on training data but also on unseen data.

DISCUSSIONS: The brain tumor detection model uses the technology of Deep Learning and Machine Learning to effectively learn the high-dimensional data and detect the presence of Gliomas in the patients exhibiting certain symptoms and some who have a family history of cancer.

Usually when symptoms first occur, the patient tends to ignore it could be a flu. But this could be early stages of Glioma. This ignorance of small symptoms, increase drastically, which leads to severe consequences with the patients. Patient then is sent for MRI, where the Glioma is diagnosed and, in most cases, it would have reached last end stages.

NeuroNet prevents this by allowing the patient to check the chances of presence of Glioma by inputting the basic symptoms they are experiencing. This is depicted by figure 33.

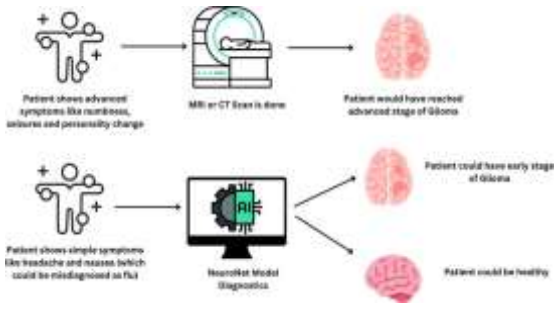


Figure 33: The application of NeuroNet model. Bottom: Depicts how NeuroNet of the symptoms of the patient before the implementation of AI Model. Bottom: Depicts how NeuroNet helps to confirm the early stages of Glioma.

This model comprises of Principal Component Analysis model for dimensionality reduction and Artificial Neural Networks to predict the presence of glioma.

As a way to achieve dimensionality reduction since the data has a high dimension the PCA was conducted. The PCA method gave the opportunity to keep 95% of the data which was important to the separation of the features so the model would not be overfitted or take a long time to learn for such a case.[32] As can be seen, this stage was essential to the complexity and variability interries of the data.

The architectural design of the model of Artificial Neural Network (ANN) is dependent on, which is a very tough measure of the computer's ability to pick out rampant patterns amid data, in the data. The ADLs are more suitable for the classification tasks like brain tumor detection which is the ability to learn from the massive datasets by distinguishing the types of data and can be used for augmentation of the same. There were various methods applied to enhance the performance and inhibit overfitting. To wit, L1 and L2 penalties and the dropout layers were utilized that is to say the modules have not been over complicated and have not captured the noise in the testing of the model. Additionally, normalization layers were used to stabilize and speed up the training process thereby improving the models' generalization power.

One of the greatest advantages ANNs has is that they can learn from the data and make forecasts. This ability is sufficient for the model to make forecasts on account of the recurring patterns that are observed during the training, which is very useful for complex tasks such as disease diagnosis. The exceptional characteristic of ANN is its versatility to be employed in all sorts of applications, which are, for instance, image acknowledgment, natural language processing, the stock market, and medical diagnosis. [33]

Besides, ANNs are the most supported technique in the field of machine learning since they can generalize any problem that has a non-linear dependent variable. [33] Typically, machine learning models will just assume a linear relationship between the input and the output, and the way they will move to the best model is by finding relations among the outputs and the inputs. On the other hand, ANNs can solve them more gracefully and can capture more nonlinear relationships more easily. This makes them more versatile, and they can model complex interactions more accurately.

This integrated model of the PCA and ANN is one of the effective integrated models to detect the disease or abnormality using patient data.

Our aim in application of AI in the field of medicine lies with the development of NeuroNet, an integrated model of PCA and ANN is highly effective for detecting diseases or abnormalities using patient data. Its applications extend to various fields of medicine, enabling quick diagnosis of conditions such as gliomas by analyzing a patient's medical history and exhibited symptoms. This rapid diagnosis capability can significantly aid in designing targeted treatments and conducting research, ultimately benefiting society by leveraging the power of artificial intelligence.
In conclusion, the brain tumor detection model combines the strengths of PCA and ANN to efficiently process high- dimensional medical data and accurately detect gliomas.

But the applications of this model is not limited only to detection of Gliomas, but can extend to various fields of medicine in order to detect the medical disorders and diagnose the disease. The integration of dimensionality reduction and deep learning techniques results in a robust tool that can significantly enhance diagnostic accuracy and speed, contributing to improved patient outcomes and advancing medical research. NeuroNet can be used to diagnose the

cancer/tumor like Gliomas much quickly by going through the patient's medical history and current status of the patient.

ACKNOWLEDGMENT: We, Nidhi Shakhapur, Sakshi Lele and Somya Mishra, the authors of this paper, would like to express our heartfelt gratitude to Dr. Sathya K, for her invaluable mentorship and unwavering support throughout the research process. Her guidance and expertise have been instrumental in shaping this work. We also extend our sincere thanks to VIT University for providing us with the necessary resources and an opportunity to conduct our research and complete the research successfully. Their support has been influential in our academic journey. This collaboration and guidance have enriched our research experience and added depth to our knowledge. Our acknowledgement goes out to all those who have contributed to this endeavor in various ways. We acknowledge the use of OpenAI's ChatGPT in the preparation of this manuscript. ChatGPT has assisted in generating text on the topic of brain tumor diagnosis using artificial intelligence. Our gratitude and thanks go out to everyone whose support has been pivotal in the successful completion of this research paper.

REFERENCES:

[1] Woźniak, M., Siłka, J., & Wieczorek, M. (2023). Deep neural network correlation learning mechanism for CT brain tumor detection. Neural Computing and Applications, 35(20), 14611-14626.

[2] Tandel, G. S., Biswas, M., Kakde, O. G., Tiwari, A., Suri, H. S., Turk, M., ... & Suri, J. S. (2019). A review on a deep learning perspective in brain cancer classification. Cancers (Basel), 11(1), 111.

[3] Saleh, A., Sukaik, R., & Abu-Naser, S. S. (2020, August). Brain tumor classification using deep learning. In 2020 International Conference on Assistive and Rehabilitation Technologies (iCareTech) (pp. 131-136). IEEE.

[4] Taher, F., Shoaib, M. R., Emara, H. M., Abdelwahab, K. M., Abd El Samie, F. E., & Haweel, M. T. (2022). Efficient framework for brain tumor detection using different deep learning techniques. Frontiers in Public Health, 10, 959667.

[5] Hemanth, G., Janardhan, M., & Sujihelen, L. (2019, April). Design and implementing brain tumor detection using machine learning approach. In 2019 3rd international conference on trends in electronics and informatics (ICOEI) (pp. 1289-1294). IEEE.

[6] Khawaldeh, S., Pervaiz, U., Rafiq, A., & Alkhawaldeh, R. S. (2017). Noninvasive grading of glioma tumor using magnetic resonance imaging with convolutional neural networks. Applied Sciences, 8(1), 27.

[7] Han, L., & Kamdar, M. R. (2018). MRI to MGMT: predicting methylation status in glioblastoma patients using convolutional recurrent neural networks. In Pacific Symposium on Biocomputing 2018: Proceedings of the Pacific Symposium (pp. 331-342).

[8] Gutta, S., Acharya, J., Shiroishi, M. S., Hwang, D., & Nayak, K. S. (2021). Improved glioma grading using deep convolutional neural networks. American Journal of Neuroradiology, 42(2), 233-239.

[9] Raja, P. S. (2020). Brain tumor classification using a hybrid deep autoencoder with Bayesian fuzzy clustering-based segmentation approach. Biocybernetics and Biomedical Engineering, 40(1), 440-453.

[10] Vamvakas, A., Williams, S. C., Theodorou, K., Kapsalaki, E., Fountas, K., Kappas, C., ... & Tsougos, I. (2019). Imaging biomarker analysis of advanced multiparametric MRI for glioma grading. Physica Medica, 60, 188-198.

[11] Abdusalomov, A. B., Mukhiddinov, M., & Whangbo, T. K. (2023). Brain tumor detection based on deep learning approaches and magnetic resonance imaging. Cancers, 15(16), 4172.

[12] ZainEldin, H., Gamel, S. A., El-Kenawy, E. S. M., Alharbi, A. H., Khafaga, D. S., Ibrahim, A., & Talaat, F. M. (2022). Brain tumor detection and classification using deep learning and sine-cosine fitness grey wolf optimization. Bioengineering, 10(1), 18.

[13] Saeedi, S., Rezayi, S., Keshavarz, H., & R. Niakan Kalhori, S. (2023). MRI- based brain tumor detection using convolutional deep learning methods and chosen machine learning techniques. BMC Medical Informatics and Decision Making, 23(1), 16.

[14] Aamir, M., Rahman, Z., Dayo, Z. A., Abro, W. A., Uddin, M. I., Khan, I., ... & Hu, Z. (2022). A deep learning approach for brain tumor classification using MRI images. Computers and Electrical Engineering, 101, 108105.

[15] Pritanjli, & Doegar, A. (2020). Brain tumor detection and classification using machine learning. In Recent Trends in Communication and Intelligent Systems: Proceedings of ICRTCIS 2019 (pp. 227-234). Springer Singapore.

[16] UshaVela. (n.d.). Brain tumor dataset [Data set]. Kaggle. Retrieved from https://www.kaggle.com/datasets/ushavela/brain-tumor

[17] McKinney, W. (2010, June). Data structures for statistical computing in Python. In SciPy (Vol. 445, No. 1, pp. 51-56).

[18] Waskom, M. L. (2021). Seaborn: statistical data visualization. Journal of Open Source Software, 6(60), 3021.

[19] Oliphant, T. E. (2006). Guide to numpy (Vol. 1, p. 85). USA: Trelgol Publishing.

[20] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. The Journal of Machine Learning Research, 12, 2825-2830.

[21] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Zheng, X. (2016). TensorFlow: A system for large-scale machine learning. In 12th USENIX symposium on operating systems design and implementation (OSDI 16) (pp. 265-283).

[22] Chollet, F., & others. (2015). Keras. Retrieved from https://keras.io

[23] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

[24] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press.

[25] Zou, J., Han, Y., & So, S. S. (2009). Overview of artificial neural networks. Artificial neural networks: methods and applications, 14-22.

[26] Kapoor, M., & Gupta, V. (2020). Astrocytoma.

[27] Grimm, S. A., & Chamberlain, M. C. (2016). Anaplastic astrocytoma. CNS oncology, 5(3), 145-157.

[28] Hart, M. N., Petito, C. K., & Earle, K. M. (1974). Mixed gliomas. Cancer, 33(1), 134-140.

[29] Van den Bent, M. J., Reni, M., Gatta, G., & Vecht, C. (2008). Oligodendroglioma. Critical reviews in oncology/hematology, 66(3), 262-272.

[30] Shalaby, A., Soliman, A., Elaskary, S., Refaey, A., Abdelazim, M., & Khalifa, F. (2023). Artificial intelligence

[31] based computer-aided diagnosis applications for brain disorders from medical imaging data. Frontiers in Neuroscience, 17, 998818.

[32] Nwoye, E., Woo, W. L., Gao, B., & Anyanwu, T. (2022). Artificial intelligence for emerging technology in surgery: Systematic review and validation. IEEE Reviews in Biomedical Engineering, 16, 241-259.

[33] Abdi, H., & Williams, L. J. (2010). Principal component analysis. Wiley interdisciplinary reviews: computational statistics, 2(4), 433-459.

[34] Zou, J., Han, Y., & So, S. S. (2009). Overview of artificial neural networks. Artificial neural networks: methods and applications, 14-22.

[35] Grimm, S. A., & Chamberlain, M. C. (2016). Anaplastic astrocytoma. CNS Oncology, 5(3), 145-157.

[36] Han, L., & Kamdar, M. R. (2018). MRI to MGMT: Predicting methylation status in glioblastoma patients using convolutional recurrent neural networks. In Pacific Symposium on Biocomputing 2018: Proceedings of the Pacific Symposium (pp. 331-342).

[37] Nwoye, E., Woo, W. L., Gao, B., & Anyanwu, T. (2022). Artificial intelligence for emerging technology in surgery: Systematic review and validation. IEEE Reviews in Biomedical Engineering, 16, 241-259.

[38] Shalaby, A., Soliman, A., Elaskary, S., Refaey, A., Abdelazim, M., & Khalifa, F. (2023). Artificial intelligence-based computer-aided diagnosis applications for brain disorders from medical imaging data. Frontiers in Neuroscience, 17, 998818.

[39] Tandel, G. S., Biswas, M., Kakde, O. G., Tiwari, A., Suri, H. S., Turk, M., & Suri, J. S. (2019). A review on a deep learning perspective in brain cancer classification. Cancers, 11(1), 111.

[40] Van den Bent, M. J., Reni, M., Gatta, G., & Vecht, C. (2008). Oligodendroglioma. Critical Reviews in Oncology/Hematology, 66(3), 262-272.

[41] Abdusalomov, A. B., Mukhiddinov, M., & Whangbo, T. K. (2023). Brain tumor detection based on deep learning approaches and magnetic resonance imaging. Cancers, 15(16), 4172.

[42] Text Generated from: OpenAI. (2024). ChatGPT (Version 4.0) [Large language model]. OpenAI. Retrieved from https://chatgpt.com/

[43] Kelvin K. L. Wong (2023). Backpropagation Neural Network. Wiley.

[44] Shachar, A. (2024). Introduction to Algogens. arXiv. https://doi.org/10.48550/arXiv.2403.01426

[45] Kelvin K. L. Wong. "Cybernetical Intelligence", Wiley, 2023

[46] Mehdi Ghayoumi. "Generative Adversarial Networks in Practice", CRC Press, 2023

[47] Seyed Navid Mashhadi Moghaddam, Huhua Cao. "Chapter 2 Artificial Intelligence", Springer Science and Business Media LLC, 2024

Websites:

[a] Pandas.(n.d.). Pandas documentation. Retrieved from https://pandas.pydata.org/docs/

[b] NumPy. (n.d.). NumPy documentation. Retrieved from https://numpy.org/doc/

[c] Seaborn. (n.d.). Seaborn documentation. Retrieved from https://seaborn.pydata.org/

[d] Scikit-Learn. (n.d.). Scikit-Learn documentation. Retrieved from https://scikit-learn.org/stable/index.html