Brain Tumor Prediction using Machine Learning

**Anurag Dutta 1[**[0000 – 0002 – 5787 – 3860**]**](https://orcid.org/0000-0002-5787-3860) **and Samrat Ray 2**[**[**0000 – 0002 – 9845 – 2974**]**](https://orcid.org/0000-0002-9845-2974)

1 Undergraduate, Department of Computer Science and Engineering, Government College of Engineering and Textile Technology, Serampore, India

2 Research Scholar, Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

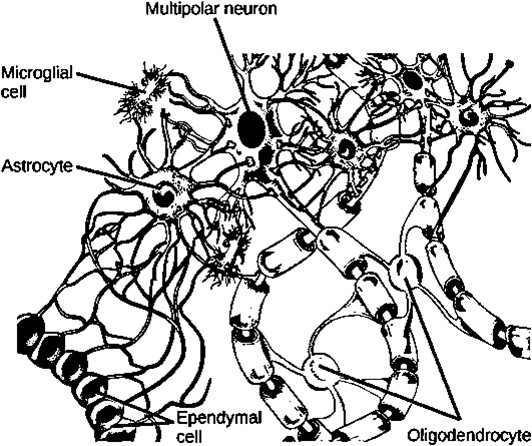
**Abstract.** In the current modernization phase, Web Technologies have caught great heights. Crawling up, and cradling, came from important and relevant technological stacks like IoT (Internet of Things), ML (Machine Learning), and AI (Artificial Intelligence). The usability of these domains is immense. In this work, we would focus on classifying Brain Tumors by taking use of Magnetic Resonance Imaging data. Brain tumors are considered one of the most aggressive diseases affecting both children and adults. Brain tumors account for 85-90% of all primary CNS disorders. About 11,700 people are diagnosed with brain tumors each year. The 5-year survival rate for people with cancerous brain or CNS tumors is about 34% for men and 36% for women. Brain tumors are classified into benign, malignant, and pituitary tumors. Appropriate treatment, planning, and accurate diagnosis are essential to prolonging patient life. The best technique for detecting brain tumors is Magnetic Resonance Imaging (MRI). Scanning creates a huge amount of image data. These images are reviewed by a radiologist. Manual inspection can be error-prone due to the complexity of brain tumors and their characteristics. The application of automatic classification techniques using machine learning (ML) and artificial intelligence (AI) consistently demonstrates higher accuracy than manual classification.

**Keywords:** Brain Tumor **·** Magnetic Resonance Imaging **·** Machine Learning Algorithms **·** Support Vector Machines **·** Brain Cells

# Introduction

Brain cells form the functional tissue of the brain. The rest of the brain tissue is structural or connective and contains blood vessels called the stroma. The main types of cells in the brain are neurons, also known as nerve cells, and glial cells, also known as glia. Neurons are excitatory cells of the brain that function by communicating (via synapses) with other neurons and interneurons in neural circuits

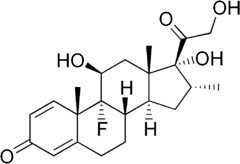
and larger brain networks. The two major neuronal classes in the cerebral cortex (Shipp, 2007) are excitatory projection neurons and inhibitory interneurons (Tepper et al., 2010). About 70 – 80 percent are excitatory projection neurons and 20 – 30 percent are inhibitory interneurons. Neurons are often grouped into clusters called nuclei and usually have similar connections and functions. Nuclei are connected to other nuclei by white matter pathways. Glia is the support cells of neurons and has many functions, not all clearly understood, including providing support and nutrition to neurons. Cellular, ependymal, oligodendrocyte macroglia (James et al., 2021), and much smaller microglia (Schwabenland et al., 2021). Astrocytes (Qin et al., 2021) can communicate with neurons through a neurotransmission - like signaling process called gliotransmission (Agulhon et al., 2010).



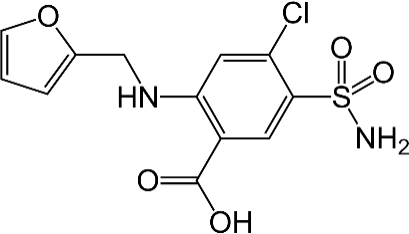
**Fig. 1.** Labeled Diagram of a Brain Cell (Neurons and Glial Cells · Biology, n.d.)

Brain tumors (Penoncello et al., 2022) occur when abnormal cells mark their presence in the brain. There are two types of tumors, Malignant tumors, and Benign tumors. These can be further divided into primary tumors that originate in the brain and secondary tumors that most spread from tumors located outside the brain known as brain metastases. All types of brain tumors can cause different symptoms de- pending on the size of the tumor and the part of the brain affected. Symptoms, if present, include headaches, seizures, blurred vision, vomiting, and mental changes. Treatment includes a combination of surgery, radiation therapy, and chemical therapy. Since the brain is the only irreplaceable organ in the body, surgery carries the risk of tumor recurrence. If seizures occur, anticonvulsants may be needed. Dexamethasone and furosemide are drugs that can be used to reduce swelling around the tumor. Some tumors grow slowly and require only monitoring and sometimes no further intervention. Treatments that use the human immune system are being re- searched. The outcome of malignant tumors varies greatly depending on the type of tumor and how widespread it is at the time of diagnosis. A benign tumor grows only in one area but can be life-threatening depending on its size and location. The out- come of malignant glioblastoma (McKenney et al., 2022) is usually very poor, whereas the outcome of benign meningioma is usually good. The 5-year survival

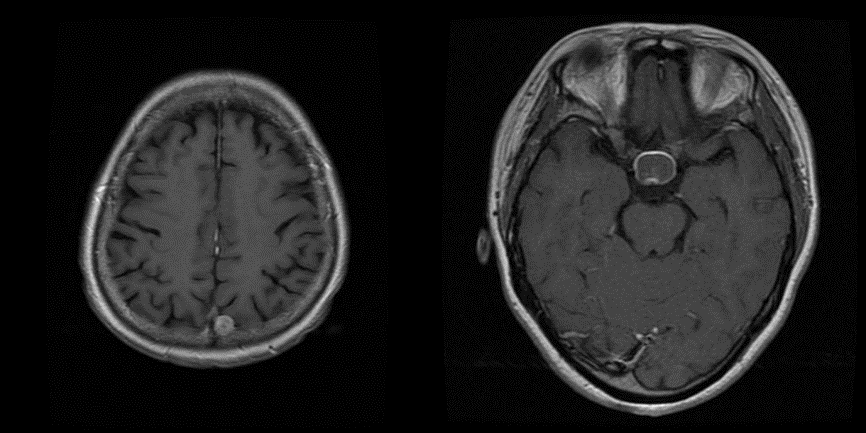
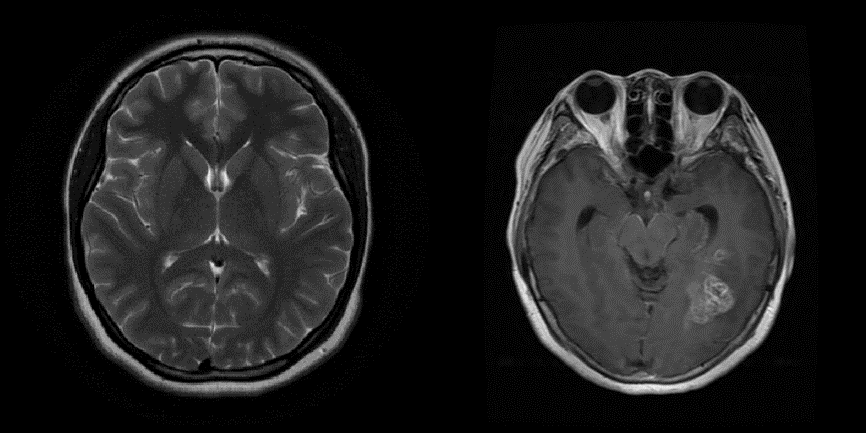
rate for people with cancerous brain or CNS tumors is about 34% for men and 36% for women.



**Fig. 2.** Dexamethasone (Sinha et al., 2021) is a glucocorticoid medication first synthesized by Philip Showalter Hench in 1957 and approved for medical use in 1958. It is used for the treatment of rheumatic problems, many skin diseases, severe allergies, asthma (Farne et al., 2022), chronic obstructive pulmonary disease (Martini & Frauenfelder, 2020), croup, brain swelling, eye pain after eye surgery, superior vena cava syndrome, a complication of certain cancers, and concomitant tuberculosis with antibiotics. Long-term use of dexamethasone can cause thrush, bone loss, cataracts, easy bruising, or muscle weakness. In the United States, it is classified as Pregnancy Cate- gory C (Sannerstedt et al., 1996). This means it should only be used where the benefits are expected to outweigh the risks. In Australia, it is Category A for oral use, meaning it is widely used during pregnancy (Taylor & James, 2011) (Bekaert & SmithBattle, 2016) and has not caused problems for the baby. Dexamethasone has anti-inflammatory and immunosuppressive effects.



**Fig. 3.** Furosemide become patented in 1959 and accredited for scientific use in 1964. It is a loop diuretic used to treat fluid buildup due to heart failure, liver scarring, or kidney disease. It can be injected into a vein or taken by mouth. Usually, when taken orally, it starts working within an hour, but when given intravenously, it usually starts working within 5 minutes. Common side effects include drowsiness when standing, tinnitus, and sensitivity to light. Potentially serious side effects include electrolyte imbalances, low blood pressure, and hearing loss (Arlinger, 2003). Regular blood tests are recommended during treatment. Common side effects of furosemide injections in- clude hypokalemia (Firth, 2010) (low potassium), low blood pressure (low blood pressure), and dizziness.



**Fig. 4.** While moving in Row Major manner, the first image shows an MRI Image of the Brain without any tumor. The second figure shows Pituitary Tumor. Pituitary cancer (also known as pituitary cancer or metastatic PitNET) is extremely rare. Although it can occur at any age, it occurs most often in older people. These cancers, like many adenomas, usually produce hormones. Pituitary cancers look like pituitary adenomas (Melmed, 2020) under a microscope, so doctors have difficulty distinguishing between them. The only way to tell for sure if a pituitary tumor is a carcinoma rather than an adenoma is if the tumor has spread to other parts of the body away from the pituitary gland. The most common pituitary cancers spread to the brain, spinal cord, meninges (the lining of the brain and spinal cord), or bones around the pituitary gland. These cancers rarely spread to other organs, such as the liver, heart (Neubauer, 2007) (Flaws, 1994), or lungs. The third image shows Meningioma Tumor. A meningioma is a tumor that forms in the membranes that cover the brain and spinal cord just inside the skull. Specifically, tumors form in a three-layered membrane called the meninges. These tumors often grow slowly. Up to 90% are benign (noncancerous). Most meningiomas form in the brain. However, they can also grow in parts of the spinal cord. Meningiomas often cause no symptoms and do not require immediate treatment. However, benign meningioma growth can cause serious problems. In some cases, such growth can be fatal. Meningiomas are the most common type of tumor originating from the central nervous system. It occurs more often in females than in males. Some meningiomas are classified as atypical. They are not considered benign or malignant (cancer). But they can be malicious. A small number of meningiomas are cancerous. They tend to grow rapidly. The fourth figure shows Glioma Tumor. Glioma is a type of tumor that forms in the spinal cord and brain. Gliomas arise from glial cells that surround nerve cells and help them function. Three types of glial cells form tumors, and gliomas are generally classified according to the type of glial cells involved in the tumor. Types of gliomas include astrocytoma (Ponnapalli et al., 2020), ependymoma (Mehrjardi et al., 2017), and

oligodendroglioma (Lassman, 2009). Glioma is one of the most common forms of primary brain tumor. They can impair brain function and can even be life-threatening depending on their location and rate of growth. Conventional treatments such as radiation therapy, targeted therapy, and surgery are useful in treating gliomas. is proven.

Machine learning algorithms build models based on sample data, called training data, to make predictions and decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of applications such as medicine, email filtering, speech recognition, agriculture, computer vision, etc., where it is difficult or impossible to develop traditional algorithms that perform the required tasks.

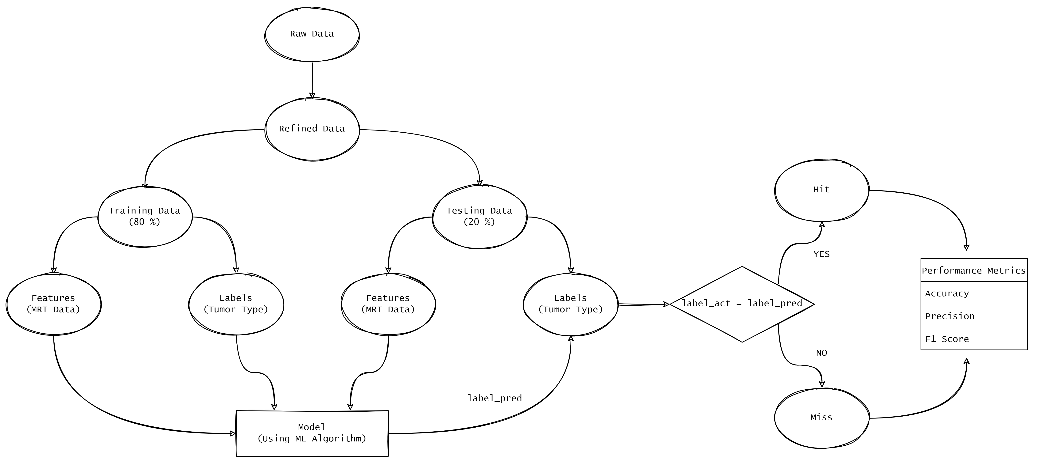
Here, in this work, we will rely on the same. We have gathered a wide range of MRI Samples and their respective tumor variant. We will try to build upon that and will make use of certain pre-existent Machine Learning Algorithms. Data, code, and supporting materials are publicly available via <https://github.com/Anurag-Dutta/BrainTumor> publicly available.

# Architecture

In this section, we describe all the details related to the proposed framework shown in Figure 5 and implement a learning technique using pre-trained ML models for classification tasks. A pre-trained classification model with learned parameters is available after training. But firstly, we will refine the raw data.

Some of them are:

1. Resizing all the MRI Data Images to a specific resolution
2. Removing any kind of noise.
3. Adjusting the picture corrections like Brightness, Contrast, etc.



**Fig. 5.** Proposed Framework for the Brain Tumor Classification

Further, the Hit Rate and Miss Rate could be used to calculate useful metrics like Accuracy, Precision, and F1 Score.

# ML Algorithms

We have used Machine Learning Algorithms, for preparing the model. We made use of a Support Vector Machine (SVM) (Cuingnet et al., 2011), Random Forrest Classifier (RF) (Smith et al., 2013) , KNN Classifier (Samworth, 2012), Gaussian Naïve Bayes (GNB) (Niculescu-Mizil & Caruana, 2005), Logistic Regression (LR) (Tolles & Meurer, 2016).

In machine learning, a support vector machine is a supervised learning model with an associated learning algorithm that analyses data for classification and regression analysis. Developed by Vladimir Vapnik and colleagues at AT&T Bell Laboratories. SVM is one of the most robust forecasting methods based on statistical learning frameworks. Given a set of training examples, each of which is marked as belonging to one of the binary categories, the SVM training algorithm creates a model that assigns new examples to either category and calls it a non-probabilistic binary linear classifier. SVM maps training samples to points in space to maximize the width of the gap between two categories. New examples are then mapped to the same space and predicted to belong to a category based on which side of the gap they are on. In addition to performing linear classification, SVMs can efficiently perform nonlinear classification using so-called kernel tricks by implicitly mapping the input to a high-dimensional feature space.

Random Forest (Random Decision Forest) is an ensemble learning method for classification, regression, and other tasks that works by building different decision trees during training. For classification tasks, the output of a random forest is the classes chosen by most trees. For regression tasks, the mean or mean prediction for each tree is returned. Random Decision Forest corrects the decision tree's habit of overfitting the training set. Random forests generally outperform decision trees but are less accurate than gradient-boosted trees. However, data properties can affect performance.

K-Nearest Neighbor is one of the simplest nonparametric machine learning algorithms based on supervised learning techniques. Assuming similarities between new cases/data and available cases, assigning new cases to categories that most closely resemble available categories, storing all available data, and based on similarity Classifying new data points. This means that the K-NN algorithm can be used to easily classify new data into appropriate categories. K-NN algorithms can be used for both regression and classification, but are primarily used for classification problems. That is, no assumptions are made about the underlying data. It is also called a delayed learning algorithm because it does not learn from the training set immediately, but instead saves the dataset and performs some action on it during classification. The k-NN algorithm only stores the training phase dataset, and as it receives new data, it classifies the data into categories that closely resemble the new data.

Naive Bayes is a simple technique for creating classifiers. A model that assigns class labels to problem instances, represented as a vector of feature values, class labels are drawn from a finite set. There is no single algorithm for training such classifiers, but there is a family of algorithms based on common principles.

All naive Bayesian classifiers assume that the value of a particular feature is independent of the value of another feature given the class variables. For example, a red, round fruit about 10 cm in diameter can be considered an apple. A simple Bayesian classifier considers each of these features independently of the probability that the fruit is an apple, independent of possible correlations between the color, roundness, and diameter features. Many practical applications use maximum likelihood methods for parameter estimation of simple Bayesian models. In other words, we can use simple Bayesian models without accepting Bayesian probabilities or using Bayesian methods. Despite its naive design and seemingly oversimplified assumptions, naive Bayesian classifiers perform well in many complex real-world situations. In 2004, an analysis of Bayesian classification problems showed that there are good theoretical reasons for the seemingly incredible effectiveness of simple Bayesian classifiers. However, a comprehensive comparison with other classification algorithms in 2006 showed that Bayesian classification outperforms other approaches such as boosted trees and random forests. The advantage of Naive Bayes is that it requires only a small amount of training data to estimate the parameters needed for classification.

A logistic model is a statistical model that models the probability of an event occurring by taking the log-odds of the event as a linear combination of one or more independent variables. In regression analysis, logistic regression estimates the parameters of a logistic model, the coefficients of a linear combination. Formally, in binary logistic regression, there is a single binomial dependent variable encoded by an indicator variable, with the two values labelled '0' and '1', while the independent variables are It can be a binomial variable or a continuous variable. The corresponding probability of a value marked '1' can vary between 0 and 1, so it is marked. The function that converts log odds to probabilities is a logistic function, hence the name.

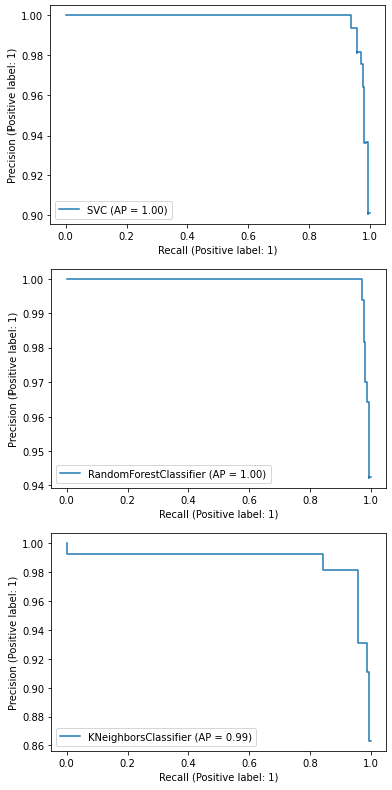
# Comparative Analysis

Predictions were made following the Machine Learning Algorithms mentioned in the previous section. And the resultant label was compared with the actual label, which helped to generate some performance metrics to compare the course of action, and to be specific the nature of their action of the following Algorithms. The Metrics ticked in our work are, Precision, Recall, Confusion Matrix.

Some of the terms worth mentioning hereby before moving further are,

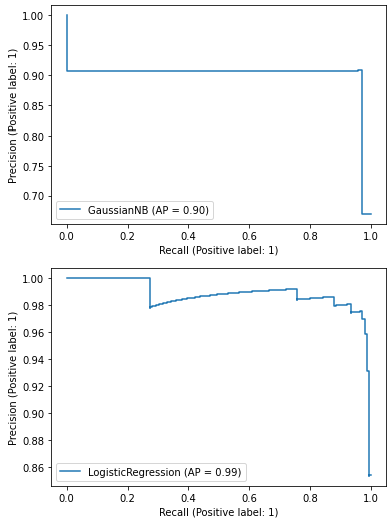
1. **True positive (TP)**: A test result that correctly indicates the existence of a state or property.
2. **True Negative (TN)**: A test result that correctly indicates the absence of a condition or feature.
3. **False Positive (FP)**: Test results that falsely indicate the existence of certain conditions or attributes
4. **False Negative (FN)**: Test results that falsely indicate that certain conditions or properties are missing

Accordingly, and .

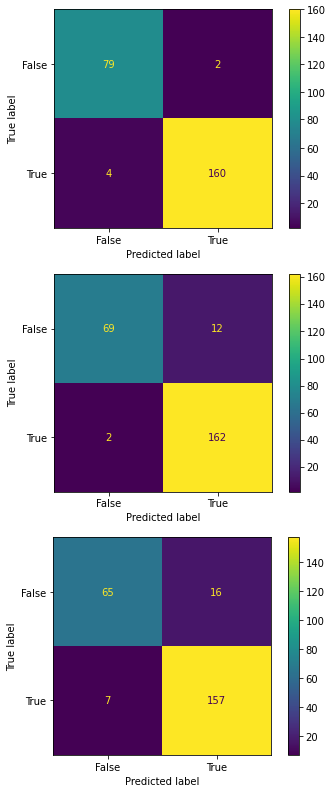
**Table 1.** Precision, Recall, and the corresponding Precision Recall Curve for the aforementioned Machine Learning Algorithms.

|  |  |  |  |
| --- | --- | --- | --- |
| *Algorithms* | *Precision* | *Recall* | *Precision Recall Curve* |
| SVM | 0.964 | 0.982 |  |
| RF | 0.988 | 0.976 |  |
| k-NN | 0.931 | 0.988 |  |

|  |  |  |  |
| --- | --- | --- | --- |
| G-NB | 0.908 | 0.957 |  |
| LR | 0.953 | 0.988 |  |

**Table 2.** Confusion Matrices for the respective Machine Learning Algorithms.

|  |  |
| --- | --- |
| *Algorithms* | *Confusion Matrix* |
| Support Vector Machine |  |



|  |  |
| --- | --- |
| Random Forest |  |
| k-NN Classifier |  |
| Gaussian Naïve Bayes |  |

|  |  |
| --- | --- |
| Logistic Regression |  |

# Conclusion

Now, we have reached to the rear part of our work. Here, we would conclude our work, and would try to draw out summarizations from the paper. In this work, we have proposed an Architecture to predict the possibility of Brain Tumor, by reading out the MRI Scan imagery. Further, the same have been acquainted with a variety of Machine Learning Algorithms, namely – “Support Vector Machine”, “Random Forest Classification”, “k – NN Classifier”, “Gaussian Naïve Bayes Classification”, “Logistic Regression”.

We have also brought about a close comparison on the performance of these Machine Learning Algorithms by drawing out the metrics, namely Precision, Recall, Precision Recall Curve, Confusion Matrices. Comparing all of the algorithm, we could probably rely on the fact that “Random Forest Classifier” would turn out the most suitable for the realization of Brain Tumor.

Further scopes of research include inclusion of latest, and stronger Ma- chine Learning Algorithms for realization of the same. Also, few other metrics could also be brought into action.

# References

*Neurons and Glial Cells* · Biology. (n.d.-b). [https://philschatz.com/biology-](https://philschatz.com/biology-book/contents/m44747.html) [book/contents/m44747.html](https://philschatz.com/biology-book/contents/m44747.html)

Shipp, S. (2007). Structure and function of the cerebral cortex. *Current Biology*, 17(12), R443–R449. <https://doi.org/10.1016/j.cub.2007.03.044>

Tepper, J. M., Tecuapetla, F., Koós, T., & Ibáñez-Sandoval, O. (2010). Heterogeneity and Diversity of Striatal GABAergic Interneurons. *Frontiers in Neuroanatomy*, 4. <https://doi.org/10.3389/fnana.2010.00150>

James, O. G., Mehta, A. R., Behari, M., & Chandran, S. (2021). Centenary of the oligoden- drocyte. *The Lancet Neurology*, 20(6), 422. [https://doi.org/10.1016/s1474-](https://doi.org/10.1016/s1474-4422(21)00136-8) [4422(21)00136-8](https://doi.org/10.1016/s1474-4422(21)00136-8)

Schwabenland, M., Brück, W., Priller, J., Stadelmann, C., Lassmann, H., & Prinz, M. (2021). Analyzing microglial phenotypes across neuropathologies: a practical guide. *Acta Neuropa- thologica*, 142(6), 923–936. <https://doi.org/10.1007/s00401-021-02370-8>

Qin, D., Wang, J., Le, A., Wang, T. J., Chen, X., & Wang, J. (2021). Traumatic Brain Injury: Ultrastructural Features in Neuronal Ferroptosis, Glial Cell Activation and Polarization, and Blood–Brain Barrier Breakdown. *Cells*, 10(5), 1009.

<https://doi.org/10.3390/cells10051009>

Agulhon, C., Fiacco, T. A., & McCarthy, K. D. (2010). Hippocampal Short- and Long-Term Plasticity Are Not Modulated by Astrocyte Ca+2 Signaling. Science, 327(5970), 1250–1254. <https://doi.org/10.1126/science.1184821>

Penoncello, G. P., Gagneur, J. D., Vora, S. A., Yu, N. Y., Fatyga, M., Mrugala, M. M., Bendok, B. R., & Rong, Y. (2022). Comprehensive Commissioning and Clinical Implemen- tation of GammaTiles STaRT for Intracranial Brain Cancer. *Advances in Radiation Oncol- ogy*, 7(4), 100910. <https://doi.org/10.1016/j.adro.2022.100910>

McKenney, A. S., Weg, E., Bale, T. A., Wild, A. T., Um, H., Fox, M. J., Lin, A., Yang, J. T., Yao, P., Birger, M. L., Tixier, F., Sellitti, M., Moss, N. S., Young, R. J., & Veeraraghavan,

H. (2022). Radiomic Analysis to Predict Histopathologically Confirmed Pseudoprogression in Glioblastoma Patients. *Advances in Radiation Oncology*, 100916. <https://doi.org/10.1016/j.adro.2022.100916>

Sinha, S., Rosin, N. L., Arora, R., Labit, E., Jaffer, A., Cao, L., Farias, R., Nguyen, A. P., de Almeida, L. G. N., Dufour, A., Bromley, A., McDonald, B., Gillrie, M. R., Fritzler, M. J., Yipp, B. G., & Biernaskie, J. (2021). Dexamethasone modulates immature neutrophils and interferon programming in severe COVID-19. *Nature Medicine*, 28(1), 201–211. <https://doi.org/10.1038/s41591-021-01576-3>

Farne, H. A., Wilson, A., Milan, S., Banchoff, E., Yang, F., & Powell, C. V. (2022). Anti- IL-5 therapies for asthma. *Cochrane Database of Systematic Reviews*, 2022(7). <https://doi.org/10.1002/14651858.cd010834.pub4>

Martini, K., & Frauenfelder, T. (2020). Advances in imaging for lung emphysema. *Annals of Translational Medicine*, 8(21), 1467–1467. <https://doi.org/10.21037/atm.2020.04.44>

Taylor, D., & James, E. A. (2011). An Evidence‐Based Guideline for Unintended Pregnancy Prevention. *Journal of Obstetric, Gynecologic &Amp; Neonatal Nursing*, 40(6), 782–793. <https://doi.org/10.1111/j.1552-6909.2011.01296.x>

Bekaert, S., & SmithBattle, L. (2016). Teen Mothers’ Experience of Intimate Partner Vio- lence. *Advances in Nursing Science*, 39(3), 272–290. <https://doi.org/10.1097/ans.0000000000000129>

Sannerstedt, R., Lundborg, P., Danielsson, B. R., Kihlström, I., Alván, G., Prame, B., & Rid- ley, E. (1996). Drugs During Pregnancy An Issue of Risk Classification and Information to Prescribers. *Drug Safety*, 14(2), 69–77. [https://doi.org/10.2165/00002018-](https://doi.org/10.2165/00002018-199614020-00001)

[199614020-00001](https://doi.org/10.2165/00002018-199614020-00001)

Firth, J. (2010). Disorders of potassium homeostasis. *Oxford Textbook of Medicine*, 3831– 3845. <https://doi.org/10.1093/med/9780199204854.003.210202_update_001>

Arlinger, S. (2003). Negative consequences of uncorrected hearing loss—a review. *Interna- tional Journal of Audiology*, 42(sup2), 17–20. <https://doi.org/10.3109/14992020309074639>

Melmed, S. (2020). Pituitary-Tumor Endocrinopathies. *New England Journal of Medicine*, 382(10), 937–950. <https://doi.org/10.1056/nejmra1810772>

Neubauer, S. (2007). The Failing Heart — An Engine Out of Fuel. *New England Journal of Medicine*, 356(11), 1140–1151. <https://doi.org/10.1056/nejmra063052>

Flaws, B. (1994). *Statements of Fact in Traditional Chinese Medicine*. Amsterdam Univer- sity Press.

Ponnapalli, S. P., Bradley, M. W., Devine, K., Bowen, J., Coppens, S. E., Leraas, K. M.,

Milash, B. A., Li, F., Luo, H., Qiu, S., Wu, K., Yang, H., Wittwer, C. T., Palmer, C. A., Jensen, R. L., Gastier-Foster, J. M., Hanson, H. A., Barnholtz-Sloan, J. S., & Alter, O. (2020). Retrospective clinical trial experimentally validates glioblastoma genome-wide pattern of DNA copy-number alterations predictor of survival. *APL Bioengineering*, 4(2), 026106. <https://doi.org/10.1063/1.5142559>

Mehrjardi, M. Z., Mirzaei, S., & Haghighatkhah, H. R. (2017). The many faces of primary cauda equina myxopapillary ependymoma: clinicoradiological manifestations of two cases and review of the literature. *Romanian Neurosurgery*, 31(3), 385–390. <https://doi.org/10.1515/romneu-2017-0062>

Lassman, A. B. (2009). Retrospective analysis of outcomes among more than 1,000 patients with newly diagnosed anaplastic oligodendroglial tumors. *Journal of Clinical Oncology*, 27(15\_suppl), 2014–2014. <https://doi.org/10.1200/jco.2009.27.15_suppl.2014>

Cuingnet, R., Rosso, C., Chupin, M., Lehéricy, S., Dormont, D., Benali, H., Samson, Y., & Colliot, O. (2011). Spatial regularization of SVM for the detection of diffusion alterations associated with stroke outcome. *Medical Image Analysis*, 15(5), 729–737. <https://doi.org/10.1016/j.media.2011.05.007>

Smith, P. F., Ganesh, S., & Liu, P. (2013). A comparison of random forest regression and multiple linear regression for prediction in neuroscience. *Journal of Neuroscience Methods*, 220(1), 85–91. <https://doi.org/10.1016/j.jneumeth.2013.08.024>

Samworth, R. J. (2012). Optimal weighted nearest neighbour classifiers. *The Annals of Sta- tistics*, 40(5). <https://doi.org/10.1214/12-aos1049>

Niculescu-Mizil, A., & Caruana, R. (2005). Predicting good probabilities with supervised learning. *Proceedings of the 22nd International Conference on Machine Learning - ICML ’05*. <https://doi.org/10.1145/1102351.1102430>

Tolles, J., & Meurer, W. J. (2016). Logistic Regression. *JAMA*, 316(5), 533.

<https://doi.org/10.1001/jama.2016.7653>