

Research Article

Magnetic Resonance Imaging (MRI) Brain Tumor Image Classification Based on Five Machine Learning Algorithms

Song Jiang^{1*} , Yuan Gu² , Ela Kumar³ 

¹Department of Biochemistry, Huzhou Institute of Biological Products Co., Ltd., Huzhou, China

²Department of Statistics, The George Washington University, Washington, DC, USA

³Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation Deemed to be University, Vaddeswaram, India

E-mail: songjiang@hzbio.net

Received: 28 March 2023; Revised: 11 April 2023; Accepted: 5 May 2023

Abstract: With the emergence of new technologies, vast amounts of data have become pervasive in various aspects of social life, including public transportation, community services, and scientific research. As the population ages, healthcare has become increasingly crucial, and reducing the public burden, especially in hospitals, has become an urgent issue. For instance, manually managing vast electronic medical files, such as MRI images, based on their types is practically impossible. However, accurate classification is fundamental and critical for subsequent tasks, such as diagnosis. In this article, we utilized machine learning techniques to classify MRI brain tumor images. We employed a range of machine learning models, including k-Nearest Neighbors (k-NN), decision tree, Support Vector Machine (SVM), logistic regression, and Stochastic Gradient Descent (SGD). The performance of each model type was measured by True Skill Statistics (TSS), based on the results obtained from the confusion matrix. The results showed that k-NN works most efficiently among all those classification models. However, due to the constraints of limited running time and computational power, further investigation of the models and parameter optimization are necessary for future work.

Keywords: MRI, brain tumor, image, classification, model

1. Introduction

Electronic Health Records (EHR) are modern digital records of health information that have revolutionized the healthcare industry such as by replacing outdated and error-prone handwritten medical records [1-3]. With the help of different forms of electronic devices such as computers, health cards, the EHR enables healthcare professionals to securely store, manage, transmit, and reproduce up-to-date confidential medical information with ease and accuracy [2-5]. EHR contains all the information present in paper medical records regardless of location and time, including well-structured data and personalized data such as doctors' notes, ambulance records, authorization details, and medication steps, thus this ensures the patients' medical history is accessible easily [4].

However, managing the diverse and often complex nature of this data can be challenging due to its subjective and mixed nature. To extract information from EHR, manual methods can be used for small datasets, but for larger datasets reaching the petabyte scale, manual methods become time-consuming, labor-intensive, error-prone, and cannot be traced. Additionally, manual methods lack transparency, making it challenging to trace how data was

Copyright ©2023 Song Jiang, et al.

DOI: <https://doi.org/10.37256/ccds.4220232740>

This is an open-access article distributed under a CC BY license
(Creative Commons Attribution 4.0 International License)
<https://creativecommons.org/licenses/by/4.0/>

collected and processed [5]. Therefore, automated and adaptive methods such as machine learning algorithms such as Natural Language Processing (NLP), Convolutional Neural Network (CNN), and Recurrent Neural Network (RNN) have emerged as very promising solutions for processing and building accurate and robust models for EHR data [6, 7].

Among those widely used algorithms, NLP a sub-field of artificial intelligence, aims to facilitate the comprehension and production of human natural language by computers [8]. Its tasks encompass a broad range of activities, including information extraction, machine translation, sentiment analysis, and abstract extraction. NLP leverages various technologies to understand and process human languages, such as name entity recognition, semantic analysis, reference resolution, speech tagging, and syntactic analysis. In the realm of medicine, NLP offers significant opportunities by analyzing a vast repositories of medical texts containing information on medical history, diagnosis, treatment methods, drugs, and other terms to uncover hidden meanings, medicine can advance significantly [6-8].

However, when dealing with some imaging data such as MRI for brain tumors, NLP has limited abilities. For example, although NLP can convert patient reports into disease labels, it has limitations in extracting tumor information from MRI images such as from Brain tumors. This task is made even more challenging by the uncommon practice of MRI image annotation [9].

Brain tumors are a very prevalent condition, and they have different types and pose a significant threat to people's health [10]. MRI is currently the most widely used method for brain tumor examination due to its speed and affordability. As a result, it has become the primary screening and diagnostic technique for various brain tumors, including glioma, meningioma, pituitary tumors, and others.

Gliomas are the most common type of brain tumor in adults, accounting for 78% of malignant brain tumors [11]. Meningiomas can occur in various locations and are also a serious concern. Although pituitary cancer is rare, it can cause significant problems by producing excess hormones [12, 13]. Early diagnosis and treatment are essential for reducing the mortality rate associated with brain tumors. However, understanding MRI brain tumor images requires a high level of expertise, typically provided by radiologists. Despite this, misdiagnosis can occur due to the varying pathological features of the brain and the potential fatigue or inexperience of radiologists. Therefore, there is a growing demand for MRI brain tumor image-assisted diagnosis algorithms that can support radiologists in treating brain tumor patients effectively and efficiently [14, 15]. Hence, the classification of MRI brain tumor image is very important. So far, most image classification works are based on neural networks, such as CNN, which is very time-consuming if it contains many hidden layers. The second reason is in the medical field, for example, before diagnosis, doctors must distinguish normal tissue from abnormal tissue, or benign from malignant tumor. Therefore, image classification is more basic and important to some extent. In the last, through searching websites, like Kaggle or GitHub, most works to classify medical images, are based on machine learning models. But few of them compared the performances. So, this article applied a statistical index to compare the performance of each machine learning model.

Due to the limitation of NLP models to handle MRI data, some other machine learning methods should be considered. Currently, in biomedical fields, such as skin cancer diagnosis, fetal ultrasound standard plane detection and localization, and lung nodule detection [16-18]. There are different machine learning techniques have shown promise in the classification of brain tumors [19], such as CNN, logistic regression, SVM, random forest, decision tree, k-means, etc. [20-22].

Although AI-based classification has advantages, it has drawbacks such as difficulty recognizing low-level features, handling small tumor spots, and dealing with changing tumor locations [23, 24]. These issues arise due to the coarse-grained recognition of AI-based systems when compared to real images. The convolutional network architecture is typically used for classification techniques, allowing for the independent processing of multiple feature maps [25]. But it is challenging to establish connections between features and identify the most representative features for fine-grained brain tumor classification [20]. For the purpose of this report, we focus on logistic regression and SVM. Logistic regression is a generalized linear regression analysis model that is widely used in data mining, disease diagnosis, economic prediction, and other fields. On the other hand, SVM is a supervised learning algorithm that seeks to find the optimal separation hyperplane in the feature space, maximizing the intervals between positive and negative samples [26, 27]. It is commonly used to solve binary classification problems. In this report, we compare the performances of logistic regression and SVM.

Another difficulty is due to the imbalance between the number of normal MRI brain tumor images and the number of brain tumor images. A large number of normal images can dominate the gradient direction and cause the algorithm to optimize towards normal images, leading to potential misdiagnosis [24]. While the algorithm can detect whether an image belongs to a diseased category or not, it is challenging to identify the type of disease and the subtle differences between tumors [25]. To address this issue, binary cross-entropy is applied to compensate for the imbalance in the loss function between MRI brain tumor images and normal MRI images [28].

In the following sections, we will provide an overview of relevant research pertaining to the topic at hand. Subsequently, we will delve into the methodology employed, including a discussion on data extraction and preprocessing, the application of various models, and a comparative analysis of the models' performance. Finally, in conclusion, we will present a comprehensive evaluation of the model results, and highlight potential avenues for future research.

2. Related research and literature review

Although MRI images are very common, the classification is still rare based on machine learning techniques. Previously, as we know, MRI images classification has already been applied in the medical field. For example, J. A. Sim and others investigated the relationship between graft appearance on follow-up MRI and knee stability after double-bundle Anterior Cruciate Ligament (ACL) reconstruction [29].

Another example of MRI images classification is conducted by Qinghua Yang and others. They investigated the association of Vitamin D Receptor (VDR) gene polymorphisms and VDR levels with lumbar disc degeneration based on the MRI spine image classification. It shows the classification of patient's spine based on biomedical methods. However, all of them are classified the MRI images based on the traditional methods based on the knowledge from doctors [30].

Interestingly, many papers have already investigated the MRI image segmentation based on the deep learning network, such as Artificial Neural Network (ANN). For example, Corin Willers and his colleagues automated lung segmentation for functional lung MRI based on Deep learning algorithms [31, 32].

3. Methodology

3.1 Dataset

The brain MRI image dataset consists of four categories: normal tissue, glioma, meningioma, and pituitary tumors (Figure 1), with thousands of images available. To enhance model efficiency and reduce processing time, we randomly selected 250 images in each category. The images are in.m format and can be converted to.jpg format directly using MATLAB software. The whole dataset can be downloaded from the link https://figshare.com/articles/brain_tumor_dataset/1512427.

3.2 Procedure

The following Figure 2 illustrates the procedure described in this article exhibits. Firstly, all needed images are imported. In this step, each image has its unique label for identifying which type of brain tumor it should be. These images are resized and followed by Principal Component Analysis (PCA) for dimension reduction. The data are then split into 2 parts: training and testing. The split ratio is 0.8. After splitting, different machine learning models are applied to the training data and training labels to train the model. The trained models are used to predict labels based on the testing features. Finally, the real testing label and predicted label are used to quantify and compare the performances of models.

3.3 Classification

The classification models used in this article include k-NN, decision tree, SVM, logistic regression and stochastic gradient descent. Each model type has different advantages and disadvantages Which results in varying performances.

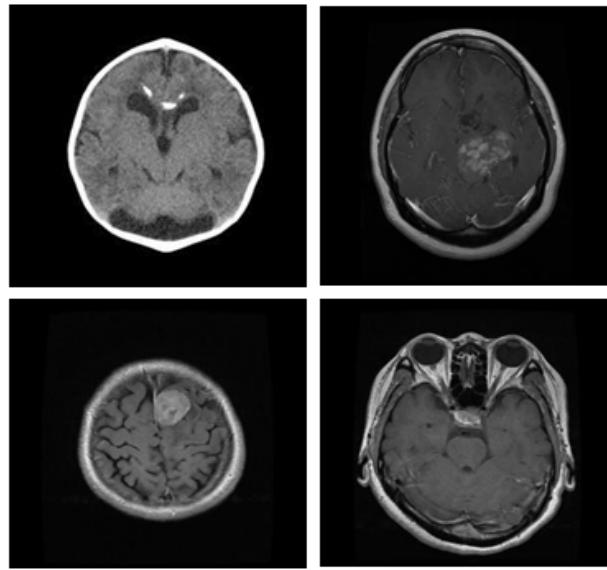


Figure 1. normal tissue (top left), glioma (top right), meningioma (bottom left) and pituitary (bottom right)

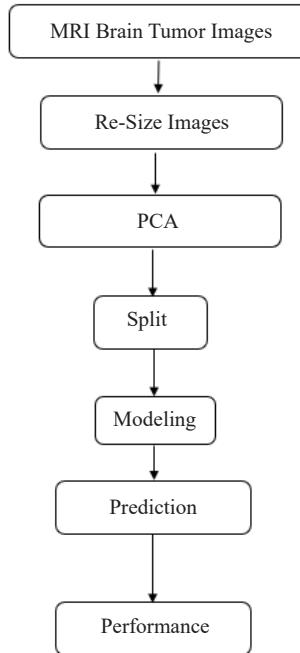


Figure 2. The flowchart of Spectral Clustering Procedure

3.3.1 *k*-NN

The first model utilized in this article is *k*-NN. In this model type, three hyperparameters, weights, number of neighbors and *p* (Table 1), are tuned. Table 1 provides detailed information about these parameters. In the table, weights can be set to uniform and distance. Uniform means All points in each neighborhood are weighted equally. Distance means points are weighted by the inverse of their distance from the query point. For example, closer neighbors of a query point will have a greater influence than neighbors who are further away. The number of

neighbors is set to default as 5. Here, we set it to 3. The last parameter p is the power parameter for the Minkowski metric. When p = 1, this is equivalent to using the Manhattan distance and when p = 2 this is equivalent to Euclidean distance.

Table 1. Parameters of k-NN

Model Number	weights	n_neighbors	p
1	uniform	3	1
2	uniform	3	2
3	uniform	5	1
4	uniform	5	2
5	distance	3	1
6	distance	3	2
7	distance	5	1
8	distance	5	2

3.3.2 Decision tree

The second type of model used is the decision tree. The parameters tuned include criterion, max depth, and splitter p (see Table 2). The splitting criteria can be set to “Gini” for the Gini impurity and “entropy” for the information gain. The splitter strategy can be “best” to choose the best split or “random” to choose the best random split. Max depth refers to the maximum depth of the tree, where if set to None, nodes are expanded until all leaves are pure or until all leaves contain less than the minimum samples split.

Table 2. Parameters of SVM

Model Number	kernel	class_weight
1	linear	None
2	linear	balanced
3	poly	None
4	poly	balanced
5	rbf	None
6	rbf	balanced
7	sigmoid	None
8	sigmoid	balanced

3.3.3 SVM

Another image classification model type used is SVM. The parameters tuned only has two types: kernel and class_weight p (Table 3). kernel type is used in the algorithm that includes ‘linear’, ‘poly’, ‘rbf’ and ‘sigmoid’. The default kernel type is ‘rbf’. The class weight is used to set the parameter C of class i to class_weight[i]*C for SVC. If not provided, all classes are assumed to have a weight of one. The “balanced” mode automatically adjusts weights inversely proportional to class frequencies in the input data based on the values of y.

Table 3. Parameters of Decision Tree

Model Number	criterion	max_depth	splitter
1	gini	5	best
2	gini	5	random
3	gini	None	best
4	gini	None	random
5	entropy	5	best
6	entropy	5	random
7	entropy	None	best
8	entropy	None	random

3.3.4 Logistic regression

The parameters of the logistic regression model type include multi_class, solver and fit_intercept p (Table 4). The multi_class parameter specifies if a binary problem should be fit for each label (option ‘ovr’) or if the entire probability distribution should be used to minimize the multinomial loss (option ‘multinomial’), even when the data is binary. Solver is used for optimization. For multiclass problems, ‘newton-cg’ and ‘saga’ should be used. Fit_intercept is used to specify if a constant should be added to the decision function.

Table 4. Parameters of Logistic Regression

Model Number	multi_class	solver	fit_intercept
1	ovr	newton-cg	True
2	ovr	saga	True
3	ovr	newton-cg	False
4	ovr	saga	False
5	multinomial	newton-cg	True
6	multinomial	saga	True
7	multinomial	newton-cg	False
8	multinomial	saga	False

3.3.5 SGD

The last model used is SGD. The parameters tuned include loss, eta0 and learning rate p (Table 5). For loss function, ‘squared_hinge’ is like a hinge but is quadratically penalized. ‘Perceptron’ is the linear loss used by the perceptron algorithm. eta0 represents the initial learning rate, and in this study, the model used optimal and invscaling learning rates.

Table 5. Parameters of SGD

Model Number	loss	eta0	Learning rate
1	perceptron	10	invscaling
2	perceptron	10	constant
3	perceptron	1	invscaling
4	perceptron	1	constant
5	squared_hinge	10	invscaling
6	squared_hinge	10	constant
7	squared_hinge	1	invscaling
8	squared_hinge	1	constant

Table 6. Confusion matrix

		Predicted Class	
		Positive	Negative
Actual Class	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

3.4 Measurement

To better access and quantify the classification performance of each model type, True Skill Statistics (TSS) is utilized as TSS is simple and intuitive measure for the model performance and it compensates for the shortcoming of other metrics such as Kappa while keeping its own advantages [33]. The formula is presented in equation 1. In this article, we use a confusion matrix to summarize the following metrics: True Negative Rate (TNR) and True Positive Rate (TPR), where TPR is recall and TNR is specificity (Table 6). In the formula of TPR and TNR, TP means True Positive, FP means False Positive, TN means True Negative and FN means False Negative. The TSS is calculated by the sum of TPR and TNR and minus 1.

$$\text{TSS} = \text{TPR} + \text{TNR} - 1,$$

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$TNR = \frac{TN}{TN + FP}$$

Equation 1. Formular of TSS.

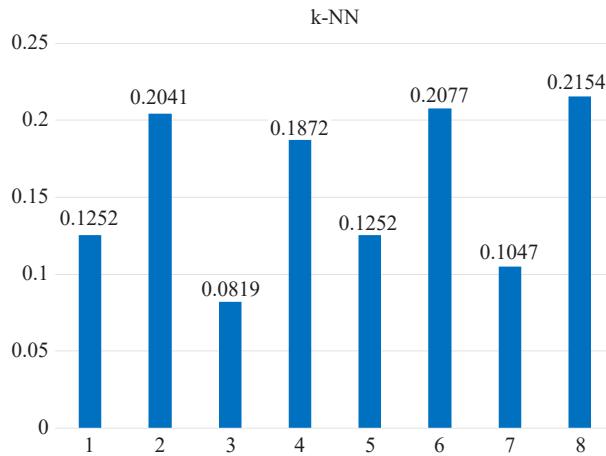


Figure 3. TSS results of k-NN

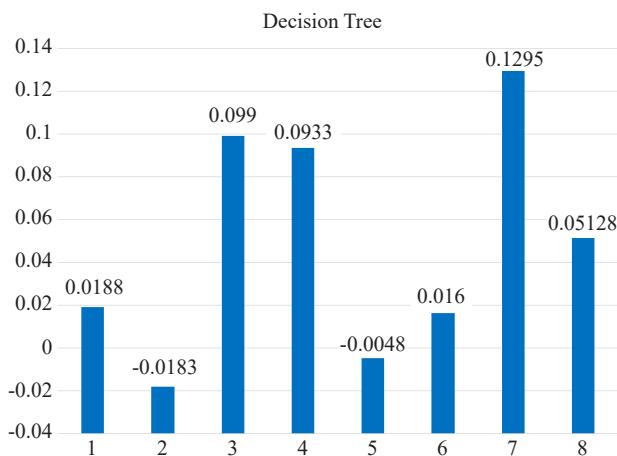


Figure 4. TSS results of Decision Tree

4. Experiments and results

Figures 3 through 8 provide a comprehensive visual representation of the quantified TSS results obtained from various models. Figure 3 shows that the last model has the highest TSS value in k-NN, while the third model has the lowest. Similarly, in Figure 4, it is observable that the second last model has the highest value of decision tree TTS, whereas the second and fifth models have negative values, which are the lowest. Figure 5 highlights that the first, second, and sixth models exhibit the highest value of TSS in SVM, while the last ones display the lowest value. Similarly, in Figure 6, the third model has the highest value of logistic regression TTS, while the second model has the lowest value. In contrast, the first and fourth models demonstrate the highest value of TTS in SGD in Figure 7, while the last model has the lowest value. Finally, Figure 8 provides an overall comparison of all model types, with k-NN showing

the best performance as it has the highest TTS value, while the decision tree displays the worst performance as its value is the lowest. In summary, the figures indicate that different models have varying levels of effectiveness in TTS, with k-NN emerging as the most effective model, while the decision tree model displays the least effectiveness.

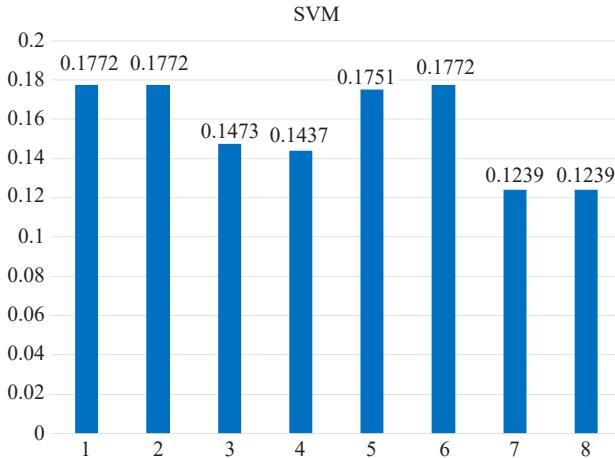


Figure 5. TSS results of SVM

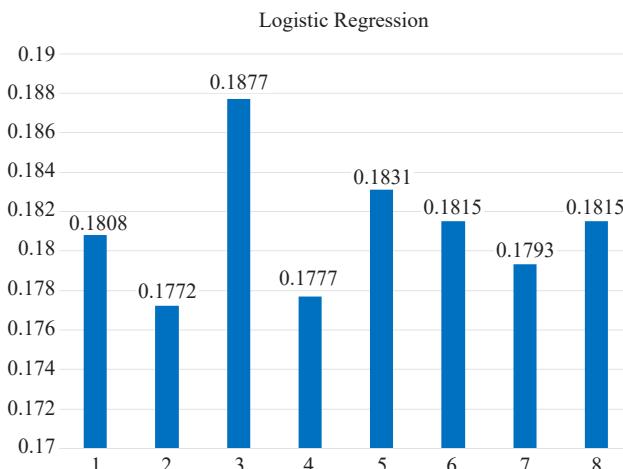


Figure 6. TSS results of Logistic Regression

5. Conclusion and future works

Based on the aforementioned analysis, it is interesting to note that each type of model exhibits its own distinct optimal model number. In particular, for the k-NN model, it has been observed that the 8th model achieves the highest level of performance. On the other hand, for the decision tree model, the 7th model has been identified as the best. Furthermore, the 1st, 2nd, and 6th models of the SVM model are recognized as the top performers for this model type. In terms of logistic regression, it has been determined that the 3rd model is the most effective. When considering the SGD model, it has been revealed that the 1st and 4th models exhibit the highest level of performance. Upon analyzing the performance of all the aforementioned best model types, it has been concluded that k-NN is the most efficient. Specifically, the optimal parameters for the best k-NN model include uniform weights, 3 n_neighbors and 3 p. Notably, this model can be utilized for the purpose of diagnosing tumors based on MRI images. As there is

limited time available, several tasks need to be completed to perfect this article. Firstly, the machine learning model's running time can be optimized, particularly when dealing with extensive data. Secondly, the number of brain tumor types used in this article is not comprehensive, and additional types of MRI brain tumor images should be included to enhance the classification's significance and realism. Furthermore, each model type should have more parameters added. Lastly, since two decision tree models have negative TSS values, they should be evaluated to determine whether they require modification or are correct.

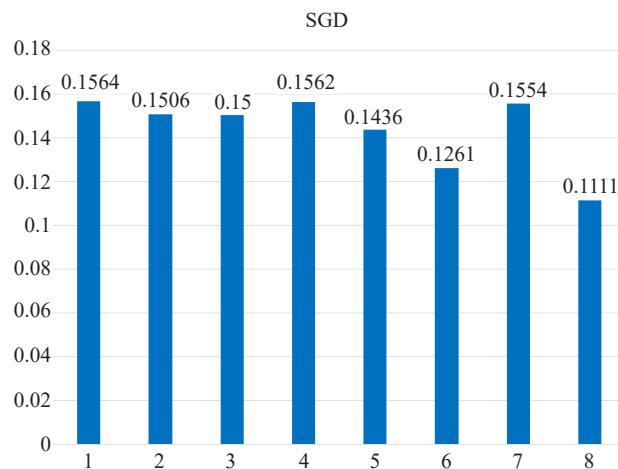


Figure 7. TSS results of SGD

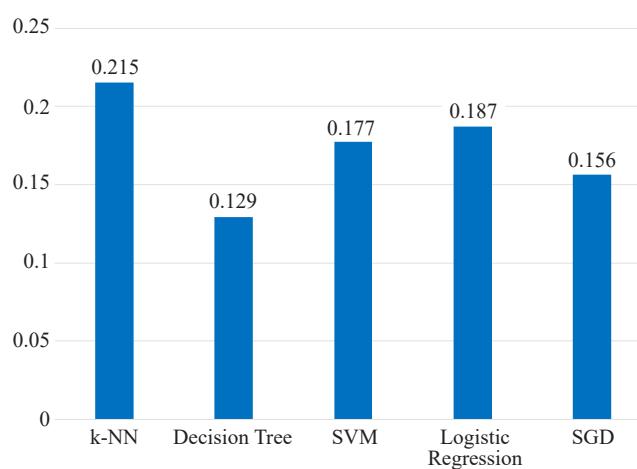


Figure 8. TSS results comparison of all best models

Conflict of interest

The authors declare no competing financial interest.

References

- [1] Ansell P, Johnston T, Simpson J, Crouch S, Roman E, Picton S. Brain tumor signs and symptoms: analysis of primary health care records from the UKCCS. *Pediatrics*. 2010; 125(1): 112-119.
- [2] Jie S, Qin X, Ying L, Gengying L. Home-built magnetic resonance imaging system (0.3 T) with a complete digital spectrometer. *Review of Scientific Instruments*. 2005; 76(10): 105101.
- [3] Evans RS. Electronic health records: then, now, and in the future. *Yearbook of Medical Informatics*. 2016; 25(S 01): S48-61.
- [4] Harman LB, Flite CA, Bond K. Electronic health records: privacy, confidentiality, and security. *AMA Journal of Ethics*. 2012; 14(9): 712-719.
- [5] Reisman M. EHRs: the challenge of making electronic data usable and interoperable. *Pharmacy and Therapeutics*. 2017; 42(9): 572.
- [6] Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. *Future Healthcare Journal*. 2019; 6(2): 94.
- [7] Rieke N, Hancox J, Li W, Milletari F, Roth HR, Albarqouni S, et al. The future of digital health with federated learning. *NPJ Digital Medicine*. 2020; 3(1): 119.
- [8] Juhn Y, Liu H. Artificial intelligence approaches using natural language processing to advance EHR-based clinical research. *Journal of Allergy and Clinical Immunology*. 2020; 145(2): 463-469.
- [9] Horowitz AL, Horowitz AL. *MRI Physics for Radiologists*. Springer; 1992.
- [10] Zülch KJ. *Brain Tumors: Their Biology and Pathology*. Springer-Verlag; 2013.
- [11] Thomas TL, Stolley PD, Stemhagen A, Fontham ET, Bleecker ML, Stewart PA, et al. Brain tumor mortality risk among men with electrical and electronics jobs: a case-control study. *Journal of the National Cancer Institute*. 1987; 79(2): 233-238.
- [12] Goodenberger ML, Jenkins RB. Genetics of adult glioma. *Cancer Genetics*. 2012; 205(12): 613-621.
- [13] Knight R, Brazier M, Collins SJ. Human prion diseases: cause, clinical and diagnostic aspects. *Contributions to Microbiology*. 2004; 11: 72-97.
- [14] Sharma K, Kaur A, Gujral S. Brain tumor detection based on machine learning algorithms. *International Journal of Computer Applications*. 2014; 103(1): 7-11.
- [15] Schlemper J, Oktay O, Chen L, Matthew J, Knight C, Kainz B, et al. Attention-gated networks for improving ultrasound scan plane detection. *arXiv*. 2018. Available from: doi: 10.48550/arXiv.1804.05338.
- [16] Lee CY, Ho J, Chow SN, Yasojima K, Schwab C, Mcgeer PL. Immunoidentification of gonadotropin releasing hormone receptor in human sperm, pituitary and cancer cells. *American Journal of Reproductive Immunology*. 2000; 44(3): 170-177.
- [17] Er O, Yumusak N, Temurtas F. Chest diseases diagnosis using artificial neural networks. *Expert Systems with Applications*. 2010; 37(12): 7648-7655.
- [18] Alemzadeh M, Boylan C, Kamath M. Review of texture quantification of CT images for classification of lung diseases. *Critical Reviews™ in Biomedical Engineering*. 2015; 43(2-3): 183-200.
- [19] Dunnmon JA, Yi D, Langlotz CP, Ré C, Rubin DL, Lungren MP. Assessment of convolutional neural networks for automated classification of chest radiographs. *Radiology*. 2019; 290(2): 537-544.
- [20] Qin R, Wang Z, Jiang L, Qiao K, Hai J, Chen J, et al. Fine-grained lung cancer classification from PET and CT images based on multidimensional attention mechanism. *Complexity*. 2020; 2020: 1-2.
- [21] Sachdev V, Tian X, Gu Y, Nichols J, Sidenko S, Li W, et al. A phenotypic risk score for predicting mortality in sickle cell disease. *British Journal of Haematology*. 2021; 192(5): 932-941.
- [22] Cure-Cure CA, Cure P, Gu Y, Tian X, Patel T, Wu CO, et al. Predictors of all cause mortality and their gender differences in a hispanic population from barranquilla-colombia using machine learning with random survival forests. *Circulation*. 2018; 138(Suppl_1): A16252.
- [23] Tonpho T, Leelasantitham A, Kiattisin S. Investigation of chest x-ray images based on medical knowledge and balanced histograms. In *2010 International Symposium on Intelligent Signal Processing and Communication Systems*. IEEE; 2010. p. 1-4.
- [24] Butnor KJ. Avoiding underdiagnosis, overdiagnosis, and misdiagnosis of lung carcinoma. *Archives of Pathology & Laboratory Medicine*. 2008; 132(7): 1118-1132.
- [25] Singh V, Misra AK. Detection of plant leaf diseases using image segmentation and soft computing techniques. *Information Processing in Agriculture*. 2017; 4(1): 41-49.
- [26] Karunakaran B, Misra D, Marshall K, Mathrawala D, Kethireddy S. Closing the loop-Finding lung cancer patients

- using NLP. In *2017 IEEE international conference on big data (big data)*. IEEE; 2017. p. 2452-2461.
- [27] Datta S, Si Y, Rodriguez L, Shooshan SE, Demner-Fushman D, Roberts K. Understanding spatial language in radiology: Representation framework, annotation, and spatial relation extraction from chest X-ray reports using deep learning. *Journal of Biomedical Informatics*. 2020; 108: 103473.
- [28] Zhang J, Xie Y, Li Y, Shen C, Xia Y. COVID-19 screening on chest x-ray images using deep learning based anomaly detection. *arXiv*. 2020. Available from: doi: 10.48550/arXiv.2003.12338.
- [29] Sim JA, Kwak JH, Lee YS, Kim KH, Nam SW, Jun SS, et al. Relationship between graft appearance on follow-up MRI and knee stability after double bundle ACL reconstruction. *Journal of the Korean Arthroscopy Society*. 2012; 16(2): 128-133.
- [30] Yang Q, Liu Y, Guan Y, Zhan X, Xiao Z, Jiang H, et al. Vitamin D Receptor gene polymorphisms and plasma levels are associated with lumbar disc degeneration. *Scientific Reports*. 2019; 9(1): 7829.
- [31] Armato SG, Li F, Giger ML, MacMahon H, Sone S, Doi K. Lung cancer: performance of automated lung nodule detection applied to cancers missed in a CT screening program. *Radiology*. 2002; 225(3): 685-692.
- [32] Willers C, Bauman G, Andermatt S, Santini F, Sandkühler R, Ramsey KA, et al. The impact of segmentation on whole-lung functional MRI quantification: repeatability and reproducibility from multiple human observers and an artificial neural network. *Magnetic Resonance in Medicine*. 2021; 85(2): 1079-1092.
- [33] Allouche O, Tsoar A, Kadmon R. Assessing the accuracy of species distribution models: prevalence, kappa and the True Skill Statistic (TSS). *Journal of Applied Ecology*. 2006; 43(6): 1223-1232.