International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 12 Issue: 2
***Article Received:** 25 November 2023 **Revised:** 12 December 2023 **Accepted:** 30 January 2024*

Machine Learning Methods for Prediction of Brain Tumors and Pneumonia Diseases

Khadija EL Haddad
Cadi Ayyad University, National School of Applied Sciences
Safi, Morocco
khadija.elhaddad@ced.uca.ma

Aissam Bekkari
Cadi Ayyad University, National School of Applied Sciences
Marrakech, Morocco
A.BEKKARI@UCA.MA

Walid Bouarifi
Cadi Ayyad University, National School of Applied Sciences
Marrakech, Morocco
W.BOUARIFI@UCA.MA

Abstract— Pneumonia and brain tumors are considered critical diseases due to the substantial challenges related to accurate prediction and diagnosis at an early stage. Machine learning (ML) methods are used in medical imaging processing to detect specific patterns and features within input images and automatically classify various medical conditions. This paper aims to predict and classify pneumonia and brain tumors diseases, to compare the ML performance of methods: Decision Tree (DT), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forests (RF), Logistic Regression (LR), and Naïve Bayes (NB), and to analyze the impact of dataset increasing size on the classification performance. This study reveals that the Random Forest algorithm achieves the best performance, with 90% accuracy in the brain tumors dataset and 79% accuracy in pneumonia disease prediction.

Keywords- ML, Classification, Medical Images, RF, SVM, NB, KNN, LR, DT, Machine Learning

I. INTRODUCTION

Nowadays many factors such as the significant growth of medical data amount, noise, and texture make manual image processing slow and inefficient. Machine learning techniques have proven efficiency and contribute to improving medical analysis solutions. Machine learning as an artificial intelligence subcategory is one of the rapidly growing domains [1] of computer science. Reinforcement, supervised, unsupervised, and semi-supervised learning are the four main groups of machine learning [2]. The process of machine learning is iterative and based on five stages:

- Data Collection: consists of collecting data from diverse sources and formats
- Data Processing: in this stage, the data is refined and features are extracted
- Model Building: for this step, the machine learning model is selected and data is trained and validated.
- Model evaluating: based on the model's evaluation, following each model's approach.
- Model Deployment: in the deployment phase, the model is integrated into the production environment for decision-making.

Figure 1 illustrates the steps of the machine learning process:

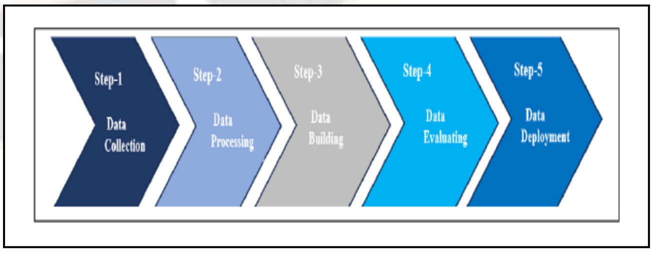

Figure 1: Machine Learning Process

Machine learning techniques have widespread applications [1] the most fundamental of which is data mining including classification, clustering, regression anomalies detection, and more in various fields such as business, economy, marketing, computer vision, healthcare, and others. In medical fields, machine learning techniques classification consists of extracting valuable data and features for accurate medical conditions prediction.

II. Related Work

[3] applied the machine learning techniques: Extreme Gradient Boost (XGB) classifier, SVM, Decision Tree, Gaussian Naive Bayes, Random Forest, Stochastic Gradient Descent (SGD) classifier, Bagging classifier, and LGBM classifier for Brain Tumors presence prediction and demonstrated that all classifiers are giving good results.

Pediatric brain tumor segmentation and classification in [4] explores the efficiency of combining two innovative texture characteristics alongside multimodal magnetic resonance image intensity, the Self Organizing Map (SOM) has proved 90% for this study.

In the [5] study, the classification of brain X-rays using classifiers SVM and KNN into malignant Vs. Benign categories use the GLCM method for extracting features from images. The results demonstrated 96% and 86% accuracy for SVM and KNN respectively.

In the research paper [6], the classification of Pneumonia is based on Random Forests, Support Vector Machines, and KNN for pulmonary disease classification. This study utilized performance metrics: Precision 89%, Recall 88%, Fi-Score 87%, and Accuracy 85%. This result demonstrates that the Random Forests method is the most efficient in classifying lung diseases.

[7] is a pneumonia prediction study combining machine learning methods and mRMR Feature detection. The comparison in this study demonstrates that the deep models are more robust in detecting pneumonia diseases.

The [8] study used a new approach based on the ultrasound video analysis application to detect pneumonia. This technique relies on the analysis of small video segments, and each part is analyzed using an algorithm of image processing to extract statistics from the entire video. This study results achieve the AUC metric values between 0.7851 and 0.9177.

III. Methodology

1. Research Approach

This study focused on the prediction and classification of pneumonia and brain tumors diseases using the ML supervised methods: Naïve Bayes, K-Nearest Neighbors, Random Forests, Support Vector Machine, Decision Tree, and Logistic Regression, and the comparison of those methods' performance using two datasets: Brain Tumors MRI dataset and Pneumonia MRI dataset. The machine learning classification is implemented under Python, the programming language widely utilized within the machine learning domain.

The classification is a supervised technique of machine learning used to determine the group association of dataset samples [9]. This technique classifies the dataset by associating the data input features with the output results approximately. The objective of the classification is to predict the target class with high precision [10], and the process for the classification in this study is through many steps: pre-processing, feature extraction, feature selection, data training, data validation, and predicting the result classes.

- Pre-processing: consists of data cleaning and denoising. The dataset is partitioned into testing and training datasets with 20% and 80% sizes respectively. The MRI images size is set to 224×224px.
- Feature extraction: in this step, the row data is transformed into numerical features for processing while maintaining the original dataset. In this paper, the PCA (Principal Component Analysis) technique is used for feature extraction.
- Feature Selection: the forward Selection approach is employed to collect the most relevant information from the data row.
- Training and Validation Data: in this stage, the model is trained and the dataset is validated to make predictions.
- Classification Outcome: the classification result and the output of the prediction of brain tumors diseases target three output classes: Normal (No Tumor), Pitularity Tumor, and Meningioma Tumor. The Pneumonia is classified as Normal (No Pneumonia), Virus Pneumonia, and Bacteria Pneumonia.
- Evaluation: the performance comparison is based on many evaluation metrics: F1-Score, accuracy, Recall, Mean Absolute Error (MAE), Precision, and Mean Squared Error (MSE). The classification experimental results will demonstrate the machine learning algorithm with the highest prediction accuracy.

Figure 2 shows the classification workflow for this study:

Figure 2: Classification Workflow

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 12 Issue: 2
***Article Received:** 25 November 2023 **Revised:** 12 December 2023 **Accepted:** 30 January 2024*

1.1 Evaluation Metrics

Many performance metrics are used for machine learning models evaluation, the metrics used in this research paper are: [23]

- Accuracy: refers to the proportion of classes accurately classified to the sum of evaluated items:

$$\text{Accuracy}=\frac{TP+TN}{TP+FP+TN+FN}$$

- F1-Score: is measured as shown below using the Precision and the Recall metrics:

$$F1=2*\frac{\text{Precision} * \text{Rcall}}{\text{Precision}+\text{Recall}}$$

- Mean Squared Error**:** the MSE is calculated as shown in the next equation**:**

$$MSE = \frac{1}{N}\sum_{I=1}^{N}(Y_I-\hat{Y}_I)^2$$

- Precision: measures the accurate positive items by all positive class predicted items:

$$\text{Precision}=\frac{TP}{TP+TF}$$

- Recall or Sensitivity: determines the fraction of the positive samples correctly classified:

$$\text{Sensitivity}=\frac{TP}{P}=\frac{TP}{TP+FN}$$

- Mean Absolute Error: is calculated as the mean of the absolute error values:

$$MAE=\frac{TP}{P}=\frac{TP}{TP+FN}$$

1.2 Datasets

In this research, the classification is performed on two medical MRI datasets:

- Brain tumors dataset: including 409 images to predict brain tumor disease.
- Pneumonia dataset: including 1047 images to predict pneumonia disease.

Both datasets are taken from the Kaggle Repository. Machine learning classifiers are applied to these datasets having different sizes to compare the performance behavior of each method and then evaluated to detect the algorithm with the best performance.

2. Machine Learning Classifiers

Machine learning methods as mentioned are classified under four major categories, each category includes subcategories.

Figure 3 illustrates machine learning techniques classification and commonly used algorithms for each category:

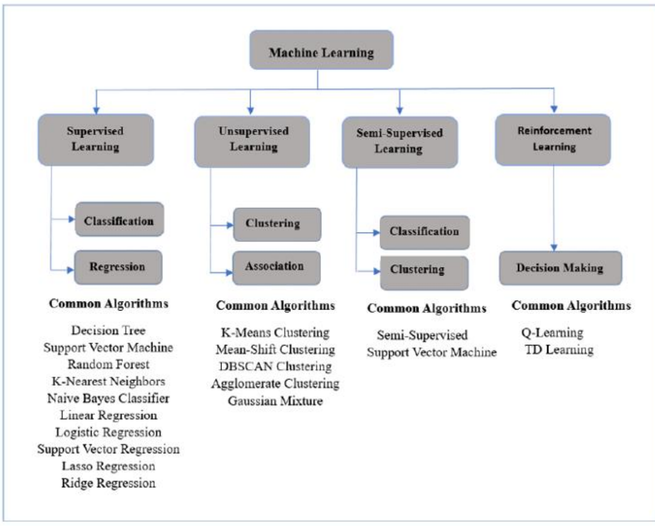

Figure 3: Machine Learning Algorithms

In this research study, our focus will be on the supervised machine-learning techniques with various applications, dealing especially with classification problems including Random Forests, Naïve Bayes, K-nearest neighbors, Support Vector Machines, Logistic Regression, and Decision Trees.

2.1 Support Vector Machine

Support vector machine is a classifier centered on a non-probabilistic binary approach that classifies data into distinct categories and uses pre-labeled data for model training and classification [14]. SVM is an innovative machine learning technique based on statistical learning theory proving efficiency in solving problems having limited samples, nonlinearities, high dimensions, and local minima [15].

The SVM algorithm relies on the discovery of a max-margins separation in the n-dimension feature space.

SVM linear is used to solve linear problems however for no linear data the kernel functions are usually more adapted. The SVM aim is the finding of a hyperplane classifying a space of n-dimension into distinct classes.

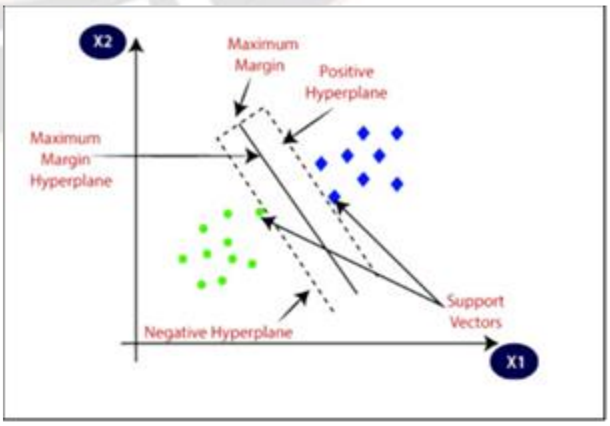

Figure 4: Illustration of a hyperplane in SVM [24]

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 12 Issue: 2
***Article Received:** 25 November 2023 **Revised:** 12 December 2023 **Accepted:** 30 January 2024*

Support vector machines are generally classified into two types, Linear SVM and Nonlinear SVM [24]:

- Linear SVM: a linear SVM refers to the SVM category used for separable data linearly classification, which signifies that data is segregated into two classes divided by a unique line.
- Non-linear SVM: generally utilized in case the data cannot be segmented into two categories having a single line as a separator, in this situation, kernel functions are used to handle non-separable datasets.

2.2 Decision Tree

Decision Tree is a technique of learning having a tree structure, to solve classification problems. For this learning approach, each tree is constituted by nodes and branches where internal nodes stand for attributes, branches define rules and the leaf node specifies the result decision [3]. In the DT algorithm, data is segmented at least into two analogous collections and then applied decision-making on that data. Used variables are commonly unrelated to make distinctive categories.

There are various types of decision tree algorithms among others: Multivariate Adaptive Regression Splines (MARS), Conditional Inference Trees (CTREE), and classification and regression Tree (CART). [10]

- Entropy and Information Gain: Entropy and information gain are metrics used in the decision tree algorithm for the dataset's impurity and data segmentation. Entropy is utilized to calculate randomness and always has a value between 0 and 1, the best performance is when the value equals 0.[10]
- Information gain: is referred to as mutual information and it represents entropy opposite since the higher its value is the better: [10]

2.3 Random Forests

Random Forests classifier, operates as a collection of classification algorithms generating various decision trees by selecting randomly training items and variables [16]. Information gain and the Gini index are key factors considered in every data split during the tree-building process [17]. Random Forests as a supervised technique is one of the most accurate algorithms for classification and regression problems.

The results of all decision trees contribute to constructing the final outcome, furthermore, the predicted result depends on the vote of every individual tree [17].

The random forests algorithm is useful to deal with data having null or missing data and to solve the problem of overfitting since the out is focused on the voting of the majority or averaging. However, it is one of the algorithms highly complex in comparison with other algorithms.

2.4 Naïve Bayes

Naïve Bayes Classifier belongs to supervised ML techniques for classification problems and utilizing the Bayes Theorem and assuming that each predictor operates independently of others [11]. It is a probabilistic classifier for which the presence of a special feature is unrelated to any other feature [12]. Naïve Bayes Classifier is a simple linear, particularly useful for large datasets. For this classifier, data training is only focused on estimating the probabilities P(c/G) and P(G) from the trained images, using the Laplace probability estimate.

The probability of a group G=[x1, x2, ... xn] in class c: [13]:

$$p(c/G)=\frac{p(G/c)p(c)}{p(G)}$$

The Naïve Bayes algorithm is classified under three categories:

- Gaussian Naïve Bayes: the Gaussian NB is used to solve classification problems and the likelihood of the feature is supposed to be Gaussian
- Multinomial Naïve Bayes: the Multinomial NB utilized with multinomially distributed data where parameters are estimated by maximum likelihood, this type of NB algorithm is used in text classification.
- Bernoulli Naïve Bayes: the Bernoulli NB is adapted with binary features.

2.5 K-Nearest Neighbors

K-Nearest Neighbors is a widely used algorithm easy to implement and a machine learning technique based on a non-parametric operational principle [18] utilized for solving regression and classification problems.

KNN operates as an instance-based approach where the classification of an item is done by its neighbor's majority vote. [19]

The classification performance of the KNN algorithm is impacted by three major indicators: the distance function utilized for neighbors’ determination, the number of nearest neighbors denoted as (k value), and the neighbor's classes level or features [18]. The equations below calculate: Euclidean Eu, Manhattan Ma, and Minkowski Mi, representing the distance functions: [12]

$$\text{Eu}=\sqrt{\sum_{k=1}^{n}(\text{pk-qk})^2}$$

$$\text{M}i=\left(\sum_{k=1}^{n}|\text{pk-qk}|^{r}\right)^{1/r}$$

$$\text{Ma}=\sum_{k=1}^{n}|\text{pk-qk}|$$

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 12 Issue: 2
***Article Received:** 25 November 2023 **Revised:** 12 December 2023 **Accepted:** 30 January 2024*

2.6 Logistic Regression

Logic Regression is a statistical process, used mostly in the biomedicines field for classification and regression purposes. LR is a predictive approach to analyzing datasets for binary variables and extracting the relation between dependent and independent binary variables. [20]

The variants of the LR algorithm are numerous and some examples are: [21]

- Basic multiple logistic regression Binary Logistic regression is the basic model used when the response is binary to target variables of many different types. The following linear relationship is used to calculate the mean [21]:

$$E(yi) = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{pi}$$

This model is based on the hypothesis that an observed target value *yi* is normally distributed. The estimated probabilities in the interval (0,1), for the following reasons:

- The value of the numerator is a power of a positive value, as a result, it is positive,
- The denominator value is (1 + numerator), so the value of pi is less than 1.

- Ridge logistic regression: the ridge logistic regression model, to overcome problems of linear regression related to poor prediction results of instability when having multiple predictors, can either use the variable selection method or adapt the logistic ridge regression estimator. [21]

The value of this estimator depends on a tuning parameter $\lambda \geqslant 0$, and are values maximizing the log-likelihood function as shown in the next equation: [22]

$$I_{\lambda}^{R}(\beta) = \sum_{i=1}^{n} + \left[y_i x_i \beta - \log\left(1 + e^{x_j \beta}\right)\right] - \lambda \sum_{j=1}^{P} \beta_j^2$$

IV. RESULTS

The experimental results of brain tumors and pneumonia disease classification using machine learning, NB, SVM, RF, DT, KNN, and LR are evaluated based on various performance metrics already explained.

Figure 5 shows a sample of the classification of the brain tumors dataset using the Random Forest Classifier:

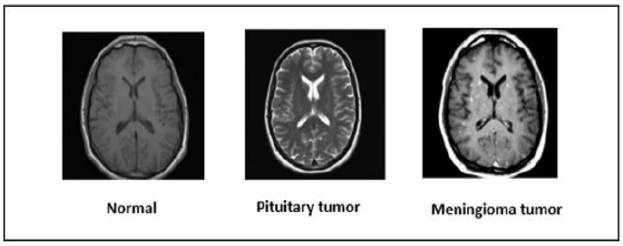

Figure 5: Random Forests algorithm classification sample output

Figure 6 illustrates a sample of pneumonia dataset classification using the NB classifier:

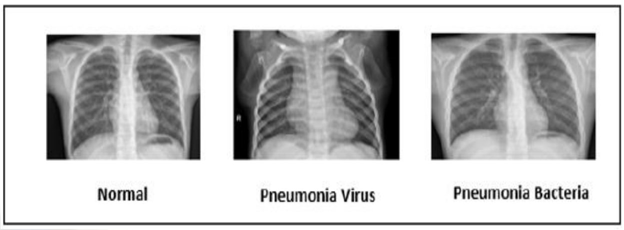

Figure 6: NB algorithm classification sample output

TABLE I AND TABLE II SHOW THE COMPARISON OF THE RESULTS FOR VARIOUS CLASSIFIERS AND BOTH DATASETS:

TABLE I. RESULTS FOR BRAIN TUMORS DATASET

Classifier	F1-Score (%)	Recall (%)	Precision (%)	MAE (%)	MSE (%)
SVM	0.89	0.88	0.92	0.17	0.30
DT	0.82	0.81	0.83	0.25	0.39
RF	0.89	0.89	0.93	0.16	0.28
NB	0.74	0.74	0.75	0.37	0.62
KNN	0.81	0.81	0.83	0.26	0.43
LR	0.83	0.84	0.85	0.25	0.44

TABLE II. RESULTS FOR PNEUMONIA DATASET

Classifier	F1-Score (%)	Recall (%)	Precision (%)	MAE (%)	MSE (%)
SVM	0.76	0.78	0.80	0.22	0.24
DT	0.63	0.64	0.66	0.42	0.56
RF	0.76	0.78	0.81	0.24	0.29
NB	0.69	0.69	0.70	0.35	0.44
KNN	0.71	0.73	0.75	0.29	0.34
LR	0.74	0.76	0.78	0.25	0.30

TABLE III illustrates the results of time execution for both datasets for various classifiers:

TABLE III. EXECUTION TIME OF CLASSIFIERS

	Execution Time(s)	
Classifier	Brain Tumors Dataset	Pneumonia Dataset
SVM	149.59	1343.26
DT	70.37	429.69
RF	31.36	231.36
NB	9.18	180.10
KNN	8.74	490.11
LR	8.66	342.90

The graphs of execution time for both datasets are shown in Figure 7 and Figure 8:

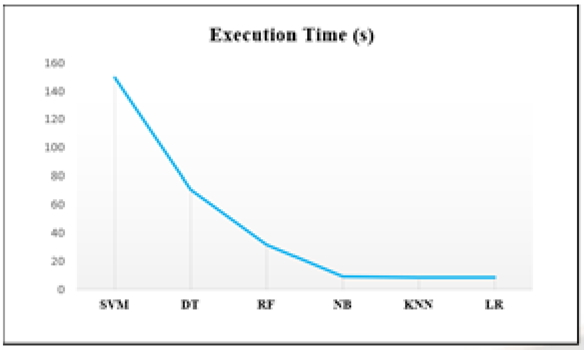

Figure7. Execution time of classifiers for brain Tumors MRI dataset

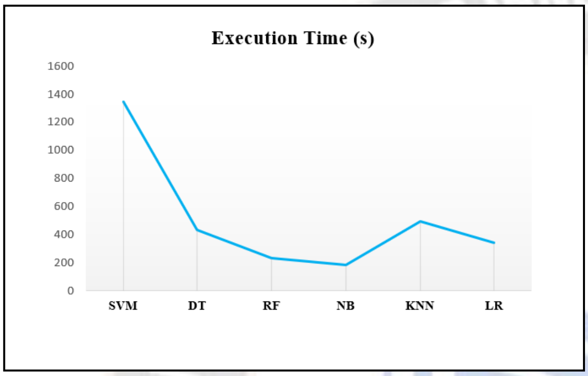

Figure 8. Execution time of classification for pneumonia dataset

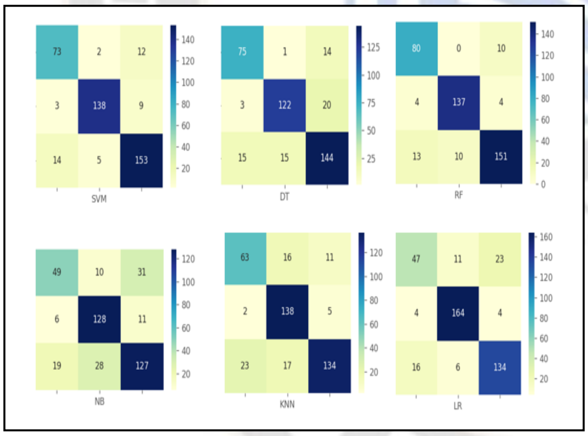

Figure 9. Confusion Matrix of various classifiers for brain tumors dataset

The accuracy of classifiers for brain tumors and pneumonia datasets is illustrated in Table 4:

TABLE IV. ACCURACY OF CLASSIFIERS

	Accuracy (%)	
Classifier	**Brain Tumors Dataset**	**Pneumonia Dataset**
SVM	0.89	0.79
DT	0.82	0.64
RF	0.90	0.79
NB	0.75	0.69
KNN	0.81	0.73
LR	0.84	0.76

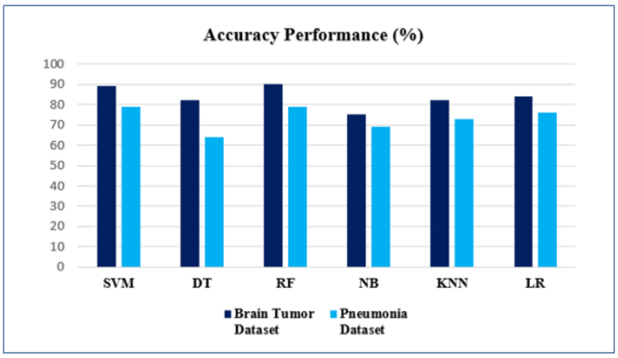

Figure 10: Comparison of accuracy performance of brain tumors dataset and pneumonia dataset

V. Discussion

In this study, the prediction and classification of brain tumors and pneumonia using ML classifiers DT, RF, SVM, KNN, NB, and LR contribute to multi-classes classification based on feature extraction, and feature selection to enhance accuracy performance.

From this study's experimental results, the Random Forest algorithm attains high performance with 90%-accuracy, 93%-precision in brain tumors prediction and 79%-accuracy, 81%-precision in pneumonia prediction. The machine learning classifiers achieve the best results for the brain tumors dataset; however, the performance has generally decreased for all ML algorithms on the pneumonia dataset having a large data size compared to the brain tumors dataset.

VI. Conclusion

Machine learning techniques have proven successful achievements in medical images analysis improving early diagnosis of brain tumors and pneumonia which is primordial for patient treatment. This research paper is based on a machine learning approach using algorithms DT, RF, SVM, KNN, NB, and Logistic Regression to make predictions. The results of experimented machine learning techniques demonstrate good results however limited to large-size datasets. The future work will focus on using deep learning techniques to make predictions with large amounts of medical data and high-quality medical images.

References

1. Osisanwo F.Y., Akinsola J.E.T., Awodele O., Hinmikaiye J. O., Olakanmi O., Akinjobi J. Supervised Machine Learning Algorithms: Classification and Comparison. International Journal of Computer Trends and Technology, 48(3):128,2017. https://www.researchgate.net/publication/318338750_Supervised_Machine_Learning_Algorithms_Classification_and_Comparison
2. Iqbal H. Sarker. Machine Learning: Algorithms, Real-World Applications and Research Directions, SN Computer Science journal,160(2):1,2021. doi:https://doi.org/10.1007/s42979-021-00592-x

3. M. Khuntia, P. Kumar Sahu and Swagatika Devi. Prediction of Presence of Brain Tumor Utilizing Some State-of-the-Art Machine Learning Approaches. International Journal of Advanced Computer Science and Applications,13(5):833, 2022.

   https://thesai.org/Downloads/Volume13No5/Paper_95-Prediction_of_Presence_of_Brain_Tumor_Utilizing_Some_State_of_the_Art_Machine.pdf

4. Iftekharuddin, K. M., Zheng, J., Islam, M. A., and Ogg, R. J. (2009). Fractal-based brain tumor detection in multimodal MRI. Applied Mathematics and Computation, 207(1):23-41,2009.

   doi: https://doi.org/10.1016/j.amc.2007.10.063

5. V. Wasule and P. Sonar. Classification of brain MRI using SVM and KNN classifier. Third International Conference on Sensing, Signal Processing and Security (ICSSS), Chennai, pages. 218-223, 2017.

   doi: https://doi.org/10.1109/SSPS.2017.8071594

6. S. Heranurweni, A. Kuniawan Nugroho and B. Destyningtias. A New Approach Method for Multi Classification of Lung Diseases using X-Ray Images. International Journal of Advanced Computer Science and Applications, 14(7):468-474, 2023.

   doi: http://dx.doi.org/10.14569/IJACSA.2023.0140751

7. B. Ergen and Z. Cömert. A Deep Feature Learning Model for Pneumonia Detection Applying a Combination of mRMR Feature Selection and Machine Learning Models. IRBM, 1:1–11, 2019.

   doi: https://doi.org/10.1016/j.irbm.2019.10.006

8. P. Cisneros-Velarde et al. Automatic pneumonia detection based on ultrasound video analysis. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Int. Conf., 2016-October:4117–4120, 2016.

   doi: https://doi.org/10.1109/embc.2016.7591632

9. Aized Amin Soofi and Arshad Awan. Classification Techniques in Machine Learning: Applications and Issues.Journal of Basic & Applied Sciences, 13:459-465, 2017.

   doi: http://dx.doi.org/10.6000/1927-5129.2017.13.76

10. Bahzad Taha Jijo, Adnan Mohsin Abdulazeez. Classification Based on Decision Tree Algorithm for Machine Learning.Journal of Applied Science and Technology Trends. 2(1):22,2021. https://www.researchgate.net/publication/350386944_Classification_Based_on_Decision_Tree_Algorithm_for_Machine_Learning

11. Batta Mahesh. Machine Learning Algorithms -A Review. International Journal of Science and Research (IJSR), 9(1):382,2019. https://www.researchgate.net/publication/344717762_Machine_Learning_Algorithms_-A_Review

12. Ch Anwar ul Hassan, Muhammad Sufyan Khan and Munam Ali Shah. Comparison of Machine Learning Algorithms in Data Classification. Proceedings of the 24th International Conference on Automation & Computing, Newcastle University, Newcastle upon Tyne, UK, pages 6-7 2018.

    https://www.researchgate.net/publication/334244465_Comparison_of_Machine_Learning_Algorithms_in_Data_classification

13. Wanli Xing, Xin Chen, Bodong Chen and Bob Wadholm. Learning Analytics at "Small" Scale: Exploring a Complexity- Grounded Model for Assessment Automation. Journal of Universal Computer Science, 21(1):78, 2015.

    https://lib.jucs.org/article/22880/download/pdf/587741

14. Imam, A. Alhroob and W. Alzyadat. SVM Machine Learning Classifier to Automate the Extraction of SRS Elements.International Journal of Advanced Computer Science and Applications, 12(3):174, 2021.

    doi: 10.14569/IJACSA.2021.0120322

15. Chaoyong Wang, Yanfeng Sun, Yanchun Liang. An Improved SVM Based on Similarity Metric. Journal of Universal Computer Science, 13(10):1462, 2007.

    https://lib.jucs.org/article/28868/

16. Mariana Belgiu and Lucian Drăguţ . Random forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing, 114:24-31, 2016.

    doi: https://doi.org/10.1016/j.isprsjprs.2016.01.011

17. Ranjan Kumar Behera, Sushree Das, Santanu Kumar Rath, Sanjay Misra, Robertas Damasevicius. Comparative Study of Real Time Machine Learning Models for Stock Prediction through Streaming Data. Journal of Universal Computer Science, 26(9):1133, 2020.

    https://www.jucs.org/jucs_26_9/comparative_study_of_real/jucs_26_09_1128_1147_behera.pdf

18. S. Memiş, S. Enginoğlu, U. Erkan. Fuzzy parameterized fuzzy soft k-nearest neighbor classifier, Neurocomputing, 23(8):351-378, 2022.

    doi: https://doi.org/10.1016/j.neucom.2022.05.041

19. Akrivi Krouska, Christos Troussas, Maria Virvou. Comparative Evaluation of Algorithms for Sentiment Analysis over Social Networking Services. Journal of Universal Computer Science, 23(8):763, 2017.

    https://www.jucs.org/jucs_23_8/comparative_evaluation_of_algorithms/jucs_23_08_0755_0768_krouska.pdf

20. Riham Alsmariy, Graham Healy, Hoda Abdelhafez. Predicting Cervical Cancer Using Machine Learning Methods. International Journal of Advanced Computer Science and Applications, 11(7): 175,2020.

    doi: https://dx.doi.org/10.14569/IJACSA.2020.0110723

21. Claudia Perlich, Foster Provost, Jeffrey S. Simonoff. Tree Induction vs. Logistic Regression: A Learning-Curve Analysis. Journal of Machine Learning Research, 4:211-255,2003.

    https://www.jmlr.org/papers/volume4/perlich03a/perlich03a.pdf

22. Jose Manuel Pereira, Mario Bastoa, Amelia Ferreira da Silvab. The logistic lasso and ridge regression in predicting corporate failure. 3rd GLOBAL CONFERENCE on BUSINESS, ECONOMICS, MANAGEMENT, and TOURISM, 39: 637,2016.

    doi: http://dx.doi.org/10.1016/S2212-5671(16)30310-0

23. Laith Alzubaidi, Jinglan Zhang, Amjad J. Humaidi, Ayad Al-Dujaili, Ye Duan, Omran Al-Shamma, J. Santamaría, Mohammed A. Fadhel, Muthana Al-Amidie and Laith Farhan. Review of deep learning: concepts, CNN architectures, challenges, applications, future

24. directions. Journal of Big Data, pages 60-61,2021.

    https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00444-8

25. Malti Bansal,Apoorva Goyal and Apoorva Choudhary. A comparative analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory algorithms in machine learning. Decision Analytics Journal, 3: 100071,2022.

    doi: https://doi.org/10.1016/j.dajour.2022.100071