INTERNATIONAL JOURNAL OF SOCIAL HEALTH
https://ijsh.ph/index.php/rp
Volume 4 Number 9 September 2023 E-ISSN: 2984-7079

Classification of Glioma Grades on Diffusion-Weighted MRI Using Random Forest in Machine Learning

Kristina Naralyawan*, Sugiyanto, Tri Asih Budiati, Edy Susanto, M.Choiroel Anwar, Gatot Murti Wibowo
Politeknik Kesehatan Kemenkes Semarang, Indonesia
Email: kristinanaralyawan@gmail.com*, gieksugiyanto@yahoo.com, budiati.triasih@gmail.com, edysusanto@poltekkes-smg.ac.id, choirul1960@gmail.com, gatotmurtiw@poltekkes-smg.ac.id

KEYWORDS	ABSTRACT
MRI; DWI; Machine Learning; Random Forest; Digital Imaging Processing	Glioma, a type of brain tumor, is classified into two grades, LGG and HGG, and requires early detection due to the high mortality rate from brain tumors in Indonesia, with 5,323 new cases and 4,229 deaths. For patients with kidney disorders, gadolinium-based contrast MRI is unsuitable, making DWI images and AI-based systems viable alternatives for detection. This study aims to develop a machine learning (ML) model using the random forest (RF) algorithm for automatic glioma grade classification from DWI MRI brain images and to evaluate its performance. Purposive sampling was used to collect 1,848 DWI MRI brain images. A quasi-experimental design with a post-test control group was employed. The model's validity was assessed by media experts using a Likert scale to evaluate aspects such as user interface design, system performance, flexibility, usability, and reliability. Reliability was tested using Fleiss' Kappa for inter-rater reliability. The ML model achieved 82% accuracy and a micro-AUC of 0.96, excelling in Normal and Grade 4 classifications but needing better recall for Grades 1 and 2. IT experts rated it positively: 84% for User Interface Design, 82% for Usability, 84.4% for System Performance, and 68.9% and 73.3% for Flexibility and Reliability, respectively. The Wilcoxon test found no significant differences among respondents ($p > 0.05$), with Fleiss' Kappa at 0.85 and 92% Observed Agreement. This study successfully developed and tested an RF model for glioma classification, demonstrating consistent and accurate results.

DOI:

Corresponding Author: KristinaNaralyawan[1*]
Email: kristinanaralyawan@gmail.com

INTRODUCTION

Brain tumors refer to abnormal cell growths within the brain or the *Central Nervous System* (CNS). Based on terminology, primary brain tumors refer to *intracranial* central nervous system tumors. Structurally, they are categorized into two major groups: neuroepithelial and non-neuroepithelial. *Intracranial* tumors consist of two types, intra-axial and extra-axial. Intra-axial tumors are located within the brain *parenchyma*, involving neural and glial cells, while extra-axial tumors are found outside the brain *parenchyma*. One common example of intra-axial tumors is glioma (Chougule et al., 2020).

Gliomas are the most common type of neuroepithelial tumors in the brain's nervous system. The World Health Organization (WHO) classifies gliomas into four grades (I, II, III, and IV) (Bray F et al., 2018) based on aggressiveness. Gliomas are divided into two categories: Low-Grade Gliomas (LGG), comprising Grades I and II, and High-Grade Gliomas (HGG), comprising Grades III and IV. Grade I gliomas are benign and can often be treated surgically, while Grade II gliomas are low-grade malignancies that may not be entirely resolved. Grade III gliomas are invasive and aggressive with a poor prognosis, whereas Grade IV gliomas are the most invasive type, associated with a very poor prognosis (Louis DN et al., 2016).

In 2013, the American Cancer Society estimated that 23,130 individuals in the United States were diagnosed with brain and other nervous system diseases, resulting in 14,080 deaths (Siegel et al., 2013). In 2016, a study by Aman et al. recorded 2,142 cases of brain tumors per 100,000 population annually in the United States, with a global rate of 34 per 100,000 population annually (Aman et al., 2016). In Indonesia, the same year saw 5,405 deaths from 6,337 cases of brain tumors, representing 85.29% of cases (Patel et al., 2016). Globally, brain tumors rank as the 10th leading cause of death among men and women. By 2020, it was estimated that 251,329 people had died due to brain cancer (Suta et al., 2019). Data from the Global Cancer Observatory (GCO) in 2020 showed that brain tumors had significant prevalence in Indonesia, ranking 15th among diseases with the highest case numbers, with an average of 15,310 deaths over the past five years (Ferlay et al., 2020).

The treatment options for brain tumors depend on their location, size, morphology, and type, including radiation therapy, surgery, chemotherapy, or combinations of various methods. Medical imaging plays a vital role in diagnosing brain tumors due to their non-specific symptoms. Radiological examinations for brain tumors include Computed Tomography (CT), Positron Emission Tomography (PET), Magnetic Resonance Spectroscopy (MRS), Single-Photon Emission Computed Tomography (SPECT), and Magnetic Resonance Imaging (MRI) (J.R. Fink et al., 2020).

Brain tumors can be identified using MRI, which offers high sensitivity and excellent image quality (Riley et al., 2018). MRI effectively displays both soft and hard tissues, enabling accurate identification. It utilizes magnetic fields and radio waves to produce images and analyze structural, vascular, and metabolic functions. The advantages of MRI include clearer images, high spatial resolution, sharp contrast between tissue types, the absence of ionizing radiation exposure, and the ability to produce multiplanar images (Nizar et al., 2019). However, MRI contrast media usage has contraindications, such as adverse reactions and issues related to gadolinium, including *post-contrast acute kidney injury* (PC-AKI) and myeloma. Patients with renal dysfunction (eGFR < 60 mL/min/1.73 m²) should avoid gadolinium-based contrast media (Thomsen et al., 2006).

MRI diagnosis of brain tumors involves contrast-enhanced axial T1 spin echo, T2 fast recovery fast spin echo, T2 Fluid Attenuated Inversion Recovery (*FLAIR*), and coronal T1 and contrast-enhanced T1 imaging. Additional imaging techniques include Magnetic Resonance Angiography (MRA), Magnetic Resonance Spectroscopy (MRS), and Diffusion-Weighted Imaging (DWI) (Elmaoğlu et al., 2012). DWI is an advanced MRI technique that evaluates brain tumors by analyzing the movement of water molecules under abnormal conditions. Quantitative DWI parameters, such as the Apparent Diffusion Coefficient (ADC), correlate with tumor cellularity (Chandra et al., 2024). Research shows that malignant tumors generally have lower ADC values compared to benign tumors.

Digital image processing involves manipulating visual signals in the form of images using various techniques to generate new images. Detection methods using digital MRI images enable accurate tumor identification and classification, accelerating diagnoses through automated segmentation (Sensusiati et al., 2020). Advances in computer technology have enabled the development of Machine Learning (ML) systems, a branch of artificial intelligence (AI) that integrates human and machine intelligence to perform specific tasks (Winangun et al., 2020). The Random Forest (RF) algorithm is a classification method that combines multiple trees to improve accuracy, robustness against outliers, and storage efficiency. RF also performs feature selection to enhance classification performance, making it suitable for large datasets with complex parameters (Jakhar et al., 2024).

Brain tumor classification using MRI images often relies on manual segmentation by experts, which is time-consuming and prone to errors. Radiologists face challenges when interpreting complex MRI data, and the variability in glioma heterogeneity or subjective interpretations can lead to diagnostic inaccuracies.

This research evaluates the performance of ML-based methods compared to radiologist expertise in analyzing glioma tumors. The proposed ML method involves using the RF algorithm to classify digital MRI images into normal brain, LGG, and HGG categories. This approach aims to achieve automatic, accurate, and efficient glioma tumor detection with diagnostic predictions comparable to radiologists, facilitating faster and more precise diagnoses. The benefits of this study are to provide a non-invasive diagnostic alternative for patients with renal dysfunction who cannot receive gadolinium contrast media, reduce diagnosis time and subjectivity through automation, support radiologists with reliable tools for consistent glioma grading, and improve the accessibility of cutting-edge diagnostic tools in resource-constrained environments.

METHOD

This research employs a *Research and Development* (R&D) methodology. R&D is a research approach utilized to develop or validate products used in education and learning contexts. It focuses on creating specific products and testing their effectiveness. The development in this research involves designing and validating methods or tools to ensure they meet the intended objectives. The study design is a quasi-experimental design, specifically a post-test with control group design, which includes categorical independent and dependent variables. The structure of this design is illustrated in the diagram below:

Figure 1. R&D Process with The Steps
Source: Research Paper, 2024

RESULT AND DISCUSSION

Model/ Design of the Machine Learning Random Forest Algorithm

This study was approved by the Health Research Ethics Committee of Poltekkes Kemenkes Semarang on July 22, 2024. The research utilized a machine learning model to classify DWI MRI Brain images. The dataset used consisted of images in DICOM format that were converted to JPG format to reduce computation time during model training and testing. These images were then labeled and grouped into folders based on their types: Normal, Grade 1, Grade 2, Grade 3, and Grade 4.

1. Flowchart of the Machine Learning Model
   The overall machine learning process using the Random Forest algorithm is illustrated in a flowchart (Figure 2).
2. Dataset Preparation
   DWI MRI Brain images in DICOM format were converted to JPG for easier processing. Image preprocessing was conducted to optimize model performance. Images were labeled into five classes: Normal (Grade 0), Grade 1, Grade 2, Grade 3, and Grade 4.
3. Data Cleaning
   Created a main folder ("data dir") with subfolders for each image class. Removed patient identifiers (e.g., name, age, DOB), and replaced them with unique image IDs (e.g., I0000001).
4. Feature Extraction
   Loading Dataset: DICOM images were read and converted to pixel arrays. Image Processing: Converted images into structured pixel arrays for feature extraction. Feature Types Extracted:
   - Texture Features: GLCM, LBP, and HOG.
   - Shape Features: Binary region detection from grayscale images.
   - Intensity Features: Intensity values using HOG.
5. Data Splitting
   Dataset split into 70% for training and 30% for testing.
6. Model Configuration
   Random Forest Classifier was configured with key hyperparameters such as:
   - n_estimators, max_depth, min_samples_leaf, min_samples_split

- max_features, bootstrap, class_weight, and criterion.

7. Model Training and Hyperparameter Tuning
   Multiple hyperparameter combinations were tested to find the best-performing model. Results were visualized using Pandas and Matplotlib.
8. Model Evaluation
   The trained model was evaluated using: Classification Report, Confusion Matrix and ROC Curve and AUC Score.

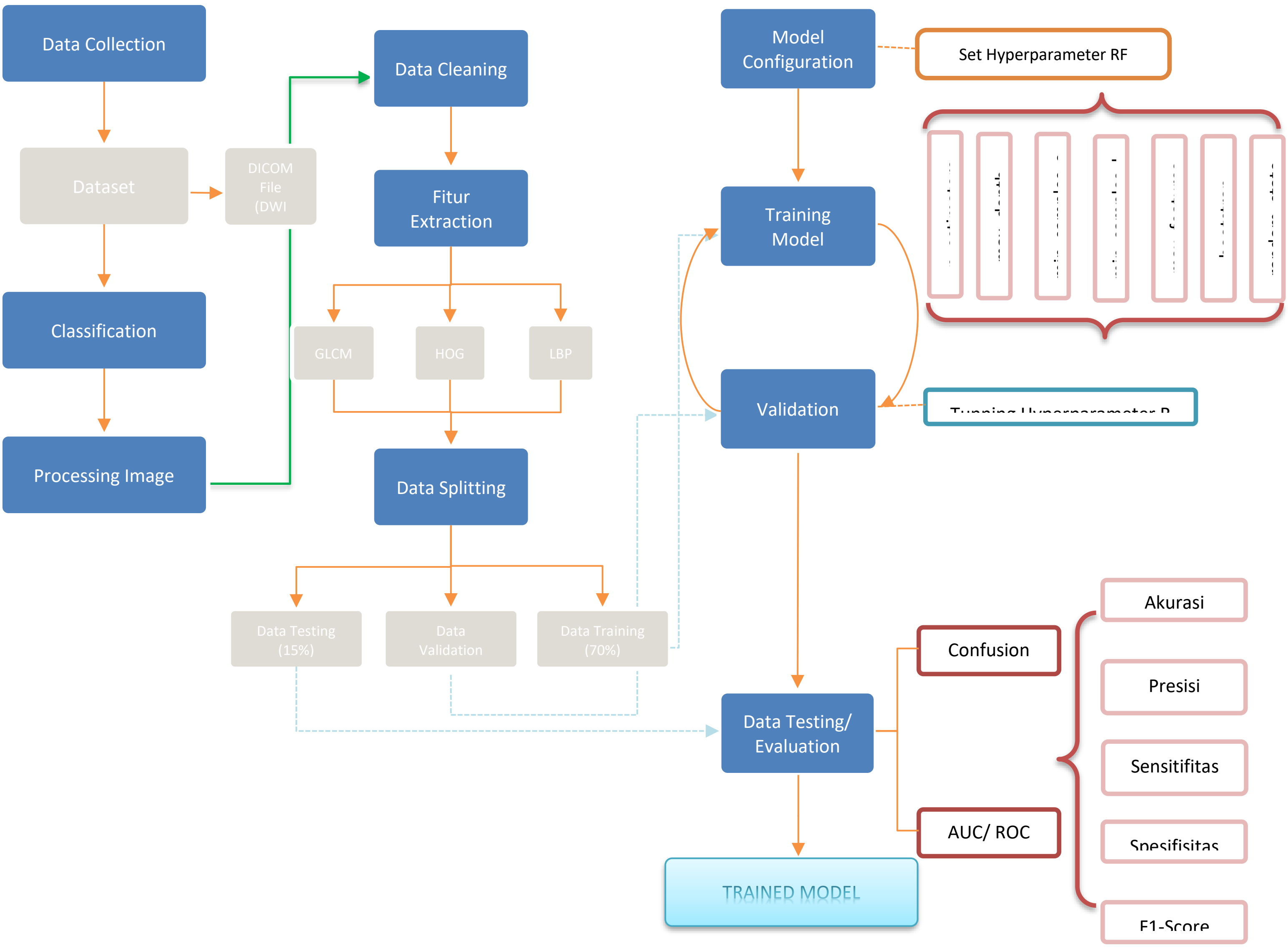


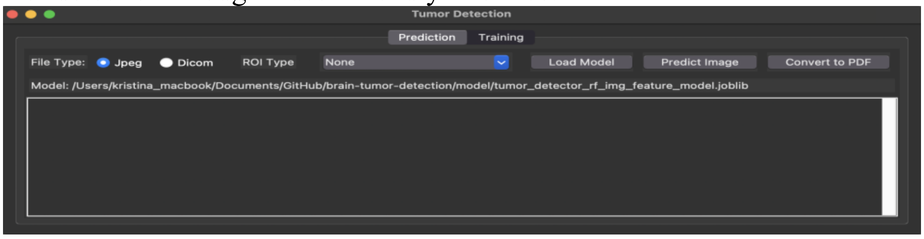
Figure 2. Flowchart of Machine Learning Model Design
Source: Research Document, 2024

9. Saving the Machine Learning Model
   The best-performing Random Forest model is saved using the save_model() function for future deployment.

```
# Save the best model
model_filename = MODEL_DIR + RF_MODEL_FILENAME
save_model(best_rf, model_filename)
```

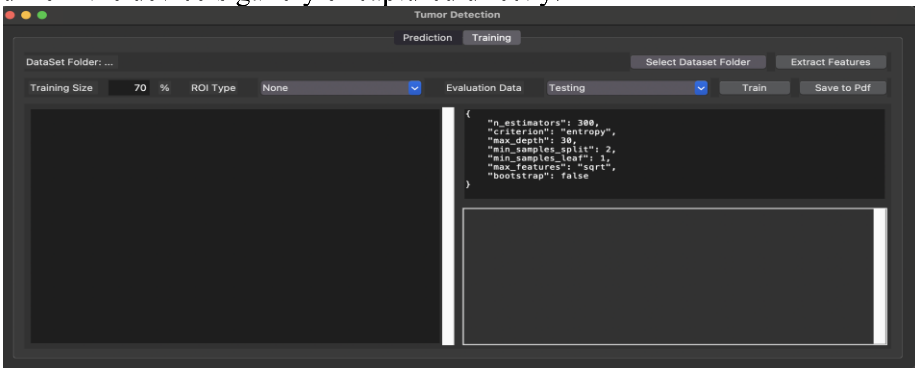
(**Figure 3.** – Saving the Machine Learning Model)
Source: Research Document, 2024

10. Model Deployment and Application Interface
    - Main Interface: The desktop application allows users to input their MRI Brain images and initiate tumor detection through a user-friendly interface.



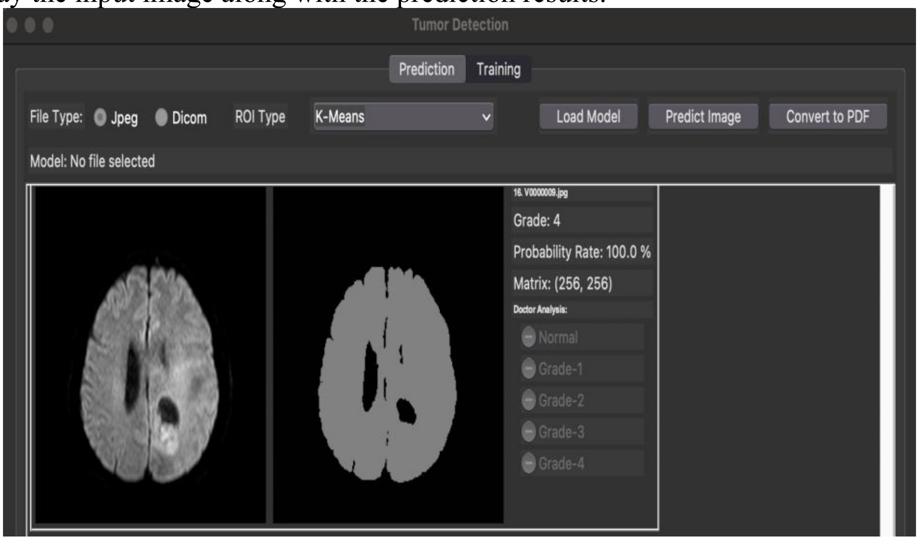
(**Figure 4.** – Main Application Interface)
Source: Research Document, 2024

- Image Upload Screen
  Users can choose to train a new model or upload an image for prediction. Images can be selected from the device's gallery or captured directly.



(**Figure 5.** – Upload Image Interface)
Source: Research Document, 2024

- Image Prediction Screen
  Once the model is loaded, users can click "Predict" and select an image. The system will display the input image along with the prediction results.



(**Figure 6.** –Prediction Interface)
Source: Research Document, 2024

Detailed procedures for model training, testing, and deployment are provided in the Appendix along with the GitHub source code.

Performance Evaluation of Machine Learning with Respondents

In this study, the performance of the machine learning model was analyzed with the involvement of respondents who were radiology specialists. Standardized interpretation of knee joint images was conducted by radiologists with more than five years of experience. Respondent characteristics indicated that two out of three respondents had over ten years of experience in radiology, ensuring their expertise and competence in assessing brain MRI results.

An inter-rater reliability test was performed to evaluate the level of agreement among the radiologists in interpreting MRI images. The methods used included Observer Agreement, Expected Agreement, and Fleiss' Kappa. The Observed Agreement was calculated based on the agreement among the three radiologists for 100 image data, resulting in 92% agreement. Expected Agreement, calculated using the total agreed-upon category ratings, yielded a value of 0.208. Fleiss' Kappa was computed using the formula K=(O−E)(1−E)K = \frac{(O - E)}{(1 - E)}, resulting in a value of 0.899, which indicates an almost perfect level of agreement among the raters.

Performance Metrics by Respondents

The classification results of brain MRI images by the three respondents showed variations in their performance. Accuracy, sensitivity, specificity, precision, and F1-score were calculated for each image grade. The following are the detailed results for each respondent:
Respondent 1

1. Accuracy**:** Grade 0: 99%, Grade 1: 94%, Grade 2: 92%, Grade 3: 94%, Grade 4: 100%.
2. Sensitivity**:** Grade 0: 100%, Grade 1: 96%, Grade 2: 72%, Grade 3: 86%, Grade 4: 100%.
3. Specificity**:** Grade 0: 99%, Grade 1: 96%, Grade 2: 96%, Grade 3: 95%, Grade 4: 100%.
4. Precision**:** Grade 0: 95%, Grade 1: 88%, Grade 2: 81%, Grade 3: 75%, Grade 4: 100%.
5. F1-Score**:** Ranged from 62% to 100%, depending on the grade.

Respondent 2

1. Accuracy**:** Grade 0: 99%, Grade 1: 95%, Grade 2: 92%, Grade 3: 92%, Grade 4: 99%.
2. Sensitivity**:** Grade 0: 95%, Grade 1: 88%, Grade 2: 81%, Grade 3: 81%, Grade 4: 100%.
3. Specificity**:** Grade 0: 100%, Grade 1: 97%, Grade 2: 95%, Grade 3: 95%, Grade 4: 99%.
4. Precision**:** Grade 0: 100%, Grade 1: 92%, Grade 2: 81%, Grade 3: 81%, Grade 4: 95%.
5. F1-Score**:** Ranged from 68% to 95%.

Respondent 3

1. Accuracy**:** Grade 0: 99%, Grade 1: 93%, Grade 2: 95%, Grade 3: 93%, Grade 4: 99%.
2. Sensitivity**:** Grade 0: 95%, Grade 1: 82%, Grade 2: 87%, Grade 3: 87%, Grade 4: 100%.
3. Specificity**:** Grade 0: 99%, Grade 1: 97%, Grade 2: 94%, Grade 3: 95%, Grade 4: 99%.
4. Precision**:** Grade 0: 95%, Grade 1: 92%, Grade 2: 81%, Grade 3: 75%, Grade 4: 95%.
5. F1-Score**:** Ranged from 72% to 95%.

Comparative Evaluation of Machine Learning Performance with Respondents

This study aimed to compare the performance of the Random Forest (RF) machine learning model with that of radiology specialists in detecting glioma brain tumors using digital MRI brain images. Radiology specialists acted as respondents, providing interpretations of the MRI images, with the highest-performing respondent (Respondent 1) selected as the primary benchmark.

To analyze the performance differences between the RF model and Respondent 1, the Wilcoxon Signed-Ranks Test was employed. This test was chosen because the study is comparative in nature, involving categorical, non-parametric, paired data with an ordinal scale (glioma grading severity). The data was measured once from DWI MRI brain images.
Wilcoxon Signed-Ranks Test Results
The test outcomes were as follows:

1. Negative Ranks: 4 cases where the RF model's performance was lower than Respondent 1.
2. Positive Ranks**:** 7 cases where the RF model's performance was higher than Respondent 1.
3. Ties**:** 89 cases where the RF model's performance was equal to Respondent 1.
4. Z Value**:** -0.663.
5. Asymptotic Significance (2-tailed)**:** 0.507.

The p-value of 0.507, which is greater than the significance level of 0.05, indicates that there is no statistically significant difference between the performance of the RF model and Respondent 1 in detecting glioma brain tumors.

Evaluation of Validity and Reliability Testing Results by IT Experts

The validity and reliability tests of the developed machine learning (ML) application model were conducted by IT experts using a Likert scale (1–5) questionnaire to assess five key aspects: User Interface Design (UID)**,** Usability**,** Flexibility**,** Reliability**,** and System Performance (SP)**.** The evaluation took place on July 30, 2024**,** with the following results:

User Interface Design (UID)

1. Score: 37 out of 45 (84%).
2. Category: Very Good.

The user interface design was rated highly, with the majority of respondents giving scores of 4 or 5. This reflects that the application's interface is visually appealing and user-friendly.

Usability

1. Score: 37 out of 45 (82%).
2. Category: Very Good.

Users found the model easy to use and functional, meeting their expectations. However, there is still room for minor improvements to enhance overall usability further.

Flexibility

1. Score: 31 out of 45 (68.9%).
2. Category: Good.

The application demonstrates adequate flexibility, but there is a need for improvement to better adapt to diverse requirements and contexts.

Reliability

1. Score: 33 out of 45 (73.3%).
2. Category**:** Good.

The reliability aspect received a "Good" rating, indicating acceptable performance consistency. However, certain weaknesses in maintaining consistent performance need to be addressed.

System Performance (SP)

1. Score: 38 out of 45 (84.4%).
2. Category: Very Good.

The application exhibits fast and responsive system performance, exceeding user expectations and delivering a high-quality experience.

CONCLUSION

Overall, the test results demonstrated that the machine learning model used in this study exhibited strong performance in classifying brain MRI images, with excellent agreement levels among the radiology specialists. These findings underline the effectiveness of the Random Forest algorithm in supporting clinical image interpretation, while also highlighting areas for potential improvement in sensitivity and positive predictive values. The results of the Wilcoxon Signed-Ranks Test show that the RF machine learning model's performance is statistically comparable to that of radiology specialists (represented by Respondent 1). This finding demonstrates that the RF model is a reliable tool for glioma brain tumor detection and has the potential to support radiologists in clinical decision-making processes. The ML application model performed exceptionally well in User Interface Design (UID), Usability, and System Performance (SP), earning a "Very Good" category in these aspects. However, Flexibility and Reliability require further improvements to enhance adaptability and consistency. Addressing these areas will ensure that the application operates optimally across various conditions and meets diverse user needs more effectively. By improving the Flexibility and Reliability aspects, the overall quality and performance of the ML application can reach an even higher standard, contributing to its success and user satisfaction. Based on the findings of the study, it is recommended to improve model recall in Grades 1 and 2 through more advanced feature engineering techniques, expand datasets from various healthcare institutions to improve model generalization, and integrate a multi-modal approach by combining different types of MRI imaging. For clinical implementation, it is necessary to conduct trials in routine diagnostic settings and develop training modules for radiologists so that the use of this system can be optimal.

REFERENCES

Aman, R. A., Soernarya, M. F., Andriani, R., Munandar, A., Tadjoedin, H., Susanto, E., & Nuhonni, S. A. (2016). *Panduan penatalaksanaan tumor otak*. Komite Penanggulangan Kanker Nasional, 1-79.

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 68(6), 394-424.

Chandra, S., Jeniyanti, N. P., & Budiati, T. A. (2024). Prosedur teknik pemeriksaan magnetic resonance spectroscopy pada kasus tumor otak di instalasi radiologi RS Pusat Pertamina Jakarta. *Jurnal Ilmu Kesehatan dan Gizi*, 2(1), 162-179.

Chougule, M. (2020). *Neuropathology of brain tumors with radiologic correlates*. Springer Nature.

Elmaoğlu, M., & Çelik, A. (2012). *MRI handbook: MR physics, patient positioning, and protocols*. Springer Science & Business Media.

Ferlay, J., Colombet, M., Soerjomataram, I., Parkin, D. M., Piñeros, M., Znaor, A., & Bray, F. (2021). Cancer statistics for the year 2020: An overview. *International Journal of Cancer*, 149(4), 778-789.

Fink, J. R., Muzi, M., Peck, M., & Krohn, K. A. (2015). Multimodality brain tumor imaging: MR imaging, PET, and PET/MR imaging. *Journal of Nuclear Medicine*, 56(10), 1554-1561.

Jakhar, D., & Kaur, I. (2020). Artificial intelligence, machine learning and deep learning: Definitions and differences. *Clinical and Experimental Dermatology*, 45(1), 131-132.

Louis, D. N., Perry, A., Reifenberger, G., Von Deimling, A., Figarella-Branger, D., Cavenee, W. K., Ohgaki, H., Wiestler, O. D., Kleihues, P., & Ellison, D. W. (2016). The 2016 World Health Organization classification of tumors of the central nervous system: A summary. *Acta Neuropathologica*, 131, 803-820.

Nizar, S., Fatimah, & Katili, M. I. (2019). Jurnal imejing diagnostik. *Jurnal Imejing Diagnostik*, 6.

Patel, A. P., Fisher, J. L., Nichols, E., Abd-Allah, F., Abdela, J., Abdelalim, A., Abraha, H. N., Agius, D., Alahdab, F., Alam, T., & Allen, C. A. (2019). Global, regional, and national burden of brain and other CNS cancer, 1990–2016: A systematic analysis for the Global Burden of Disease Study 2016. *The Lancet Neurology*, 18(4), 376-393.

Riley, R., Murphy, J., & Higgins, T. (2018). MRI imaging in pediatric appendicitis. *Journal of Pediatric Surgery Case Reports*, 31, 88-89.

Sensusiati, A. D. (2020). Apparent diffusion coefficient of diffusion weighted imaging have strong correlation with the malignancy grading of intracranial tumor. In *Journal of Physics: Conference Series* (Vol. 1445, No. 1, p. 012019). IOP Publishing.

Siegel, R., Naishadham, D., & Jemal, A. (2013). Cancer statistics, 2013. *CA: A Cancer Journal for Clinicians*, 63(1), 11-30.

Suta, I. B., Hartati, R. S., & Divayana, Y. (2019). Diagnosa tumor otak berdasarkan citra MRI (Magnetic Resonance Imaging). *Majalah Ilmiah Teknologi Elektro*, 18(2), 149-154.

Thomsen, H. S. (2006). European Society of Urogenital Radiology (ESUR) guidelines on the safe use of iodinated contrast media. *European Journal of Radiology*, 60(3), 307-313.

Winangun, P. P., Widyantara, I. M., & Hartati, R. S. (2020). Pendekatan diagnostik berbasis extreme learning machine dengan kernel linear untuk mengklasifikasi kelainan paru-paru. *Majalah Ilmiah Teknologi Elektro*, 19(1), 83-90.