Original Article

Check for updates

Available online at
www.heca-analitika.com/ijds

Infolitika Journal of Data Science

Vol. 3, No. 2, 2025

An Interpretable Machine Learning Framework for Predicting Advanced Tumor Stages

Teuku Rizky Noviandy [1] Mohsina Patwekar [2], Faheem Patwekar [3] and Rinaldi Idroes [4,*]

[1] Department of Information Systems, Faculty of Engineering, Universitas Abulyatama, Aceh Besar 23372, Indonesia; rizky_si@abulyatama.ac.id (T.R.N.)
[2] Department of Pharmacology, Luqman College of Pharmacy, Karnataka 585102, India; mohsina.patwekar@gmail.com (M.P.)
[3] Department of Pharmacognosy, Luqman College of Pharmacy, Karnataka 585102, India; ifaheemp@gmail.com (F.P.)
[4] School of Mathematics and Applied Sciences, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia; rinaldi.idroes@usk.ac.id (R.I.)

* Correspondence: rinaldi.idroes@usk.ac.id

Article History

Received 29 July 2025
Revised 17 November 2025
Accepted 24 November 2025
Available Online 29 November 2025

Keywords:
Tumor stage prediction
Machine learning
Clinical prediction
Medical machine learning

Abstract

Accurate identification of advanced tumor stages is essential for timely clinical decision-making and personalized treatment planning. This study proposes an explainable ensemble learning framework for predicting advanced tumor stage using a dataset containing 10,000 samples with 18 clinical and radiological features. Four machine learning models, namely Logistic Regression, Naïve Bayes, AdaBoost, and LightGBM, were evaluated using stratified train–test splits along with standard performance metrics. LightGBM achieved the highest performance, with an accuracy of 86.05% and an F1-score of 76.61%, outperforming linear and probabilistic classifiers. ROC–AUC and precision–recall analyses further confirmed the superior discriminative ability of ensemble methods. SHAP explainability techniques highlighted mitotic count, Ki-67 index, enhancement, and necrosis as the most influential predictors of advanced stage. The proposed framework demonstrates strong predictive capability and provides clinically interpretable insights, underscoring its potential as a decision-support tool in oncological diagnostics. Future work will involve external validation and integration of additional multimodal data to enhance generalizability.

1. Introduction

Cancer remains a leading cause of morbidity and mortality worldwide, with early diagnosis and accurate staging playing a critical role in guiding therapeutic decisions and improving patient outcomes [1, 2]. Tumor staging provides essential insights into disease progression, helping clinicians determine prognosis, select appropriate treatment strategies, and monitor therapeutic response [3, 4]. As treatment modalities become more personalized, the precision of staging becomes increasingly important for identifying patients who may benefit from targeted therapies or more intensive management.

Traditional staging processes typically rely on clinical evaluations, radiological imaging, and histopathological assessments. While these methods form the cornerstone of oncologic practice, they can be subject to inter-observer variability and may fail to fully capture complex patterns associated with disease severity [5–7]. Additionally, advanced tumor stages are often associated with more aggressive biological behavior, increased metastatic potential, and reduced survival rates, underscoring the need for timely and reliable

Table 1. Dataset features and descriptions.

Feature	Description
tumor_type	Primary tumor classification based on pathology or imaging characteristics
size	Maximum tumor diameter measured in millimeters
location	Anatomical region where the tumor is situated
edema	Presence and extent of peritumoral edema
necrosis	Degree of necrotic tissue within the tumor mass
enhancement	Contrast enhancement pattern on imaging
shape	Tumor geometric configuration (e.g., round, irregular)
margins	Tumor border definition (well-defined or infiltrative)
calcification	Presence of calcified components within the lesion
cystic_components	Proportion of cystic or fluid-filled regions
hemorrhage	Evidence of intratumoral bleeding
ki67_index	Proliferation index measured by Ki-67 immunostaining (%)
mitotic_count	Number of mitotic figures per high-power field
age	Patient age in years at time of diagnosis
gender	Biological sex of the patient
symptoms_duration	Duration of presenting symptoms before diagnosis (weeks)
neurological_deficit	Presence of neurological impairment at presentation
kps_score	Karnofsky Performance Status, assessing functional ability

identification [8–10]. Efforts to enhance staging accuracy continue to evolve, incorporating new imaging modalities, molecular markers, and computational tools to improve the detection and characterization of tumors across diverse patient populations.

Recent advancements in medical data acquisition have led to the availability of large, heterogeneous datasets comprising clinical indicators, pathological markers, and imaging-derived features [11–13]. Extracting meaningful insights from such multidimensional data is challenging using conventional statistical approaches [14, 15]. This has motivated the adoption of machine learning techniques, which offer powerful capabilities for recognizing nonlinear relationships, integrating diverse feature types, and generating accurate predictive models [16–18]. In oncology, machine learning methods have shown promise for enhancing diagnostic accuracy, risk stratification, and treatment planning by leveraging patterns that are not readily identifiable through traditional clinical evaluation.

Despite their predictive strength, many machine learning models operate as "black boxes," limiting trust and adoption in clinical environments where transparency is essential. To address this challenge, Explainable Artificial Intelligence (XAI) techniques, such as Shapley Additive exPlanations (SHAP), provide a principled way to quantify feature importance and clarify model decision pathways [19, 20]. Integrating XAI with machine learning frameworks enhances interpretability, enabling clinicians to understand not only what predictions are made but also why. This transparency is crucial for validating model reliability, improving clinical acceptance, and supporting evidence-based decision-making.

The present study aims to develop an explainable machine learning framework capable of accurately predicting advanced tumor stages using routinely collected clinical and radiological features. It evaluates the performance of four machine learning algorithms and compares their predictive capabilities using standard classification metrics. In addition, SHAP-based explainability techniques are employed to generate transparent, clinically meaningful interpretations of model behavior. By combining strong predictive performance with interpretable analytics, this study aims to enhance the reliability and clinical utility of machine learning–assisted tumor stage assessment.

2. Materials and Methods

2.1. Dataset

The dataset used in this study was sourced from a publicly available Kaggle repository and consists of 10,000 patient records, each characterized by 18 clinical and radiological features along with a target variable indicating tumor stage [21]. The original stage labels (Stage I, II, III, and IV) were restructured into two clinically meaningful categories for binary classification: non-advanced tumors (Stages I–III, 3,235 samples) and advanced tumors (Stage IV, 6,765 samples).

All features included in this study are presented in Table 1, along with their corresponding descriptions. These features represent a combination of demographic, pathological, and imaging-derived variables commonly used in tumor characterization and prognosis assessment.

2.2. Data Preprocessing

Before model development, several preprocessing procedures were implemented to enhance data quality and ensure suitability for machine learning analysis [22, 23]. All categorical features were transformed using label encoding, allowing the model to interpret non-numerical variables effectively. This step ensured that both clinical and radiological categorical attributes were represented in a numerically consistent format.

After encoding, the dataset was split into training and test sets using an 80/20 stratified split to preserve the original distribution of non-advanced and advanced tumor cases. After stratification, the training subset contained 2,593 non-advanced and 5,407 advanced samples, while the testing subset included 642 non-advanced and 1,358 advanced samples. This stratified approach ensured that performance evaluation remained unbiased and representative of real-world class proportions.

2.3. Machine Learning Models

To construct a robust, explainable framework for advanced tumor-stage prediction, four supervised machine learning algorithms were implemented: Logistic Regression, Naïve Bayes, AdaBoost, and LightGBM. These models were selected to represent a diverse spectrum of learning paradigms, including linear classification, probabilistic modeling, adaptive boosting, and gradient-boosted decision trees, respectively.

Logistic Regression was employed as a baseline linear model due to its interpretability and the ability to quantify the strength and direction of feature associations with tumor advancement [24, 25]. Naïve Bayes, which applies Bayes' theorem under conditional independence assumptions, was included for its computational efficiency and effectiveness when handling mixed numerical and categorical feature sets [26, 27].

To capture potential nonlinear relationships within the dataset and improve predictive performance, two ensemble-based algorithms were incorporated. AdaBoost (Adaptive Boosting) constructs an ensemble of weak learners by iteratively adjusting feature weights and emphasizing previously misclassified samples, thereby reducing overall classification error [28, 29]. LightGBM, a high-performance gradient boosting framework, utilizes histogram-based splitting and leaf-wise tree growth strategies to efficiently model complex feature interactions, making it well-suited for structured clinical datasets [30, 31].

All models were trained using their respective default hyperparameters to ensure a fair and unbiased comparative baseline. Additionally, a random state of 42 was applied to all algorithms and data-splitting procedures to guarantee reproducibility of results.

2.4. Performance Evaluation

The predictive performance of all machine learning models was assessed using four widely adopted classification metrics: accuracy, precision, recall, and F1-score [32]. These metrics were selected to ensure a comprehensive evaluation of model effectiveness, particularly given the class imbalance between non-advanced and advanced tumor cases and the clinical significance of correctly identifying advanced tumors. All evaluation metrics were calculated using the independent test set, which was kept entirely separate from the training and resampling procedures. This approach ensured an unbiased estimation of each model's generalization capability.

Accuracy quantifies the overall proportion of correctly classified samples across both classes. Precision measures the proportion of instances predicted as advanced tumors that are truly advanced, reflecting the model's ability to minimize false-positive predictions. Recall (sensitivity) indicates the proportion of actual advanced tumor cases correctly identified and is especially important in clinical contexts where missed detections carry substantial medical risk. Finally, the F1-score, defined as the harmonic mean of precision and recall, provides a balanced performance measure that accounts for both false positives and false negatives, making it suitable for evaluating models under imbalanced class distributions.

2.5. SHAP Explainability

To ensure transparency and interpretability within the predictive framework, SHAP was applied to the best-performing machine learning model to analyze and explain its outputs [33]. SHAP is grounded in cooperative game theory and computes Shapley values, which quantify each feature's contribution to the model's prediction. This approach enables a detailed examination of how both clinical and radiological attributes, such as tumor size, necrosis, enhancement patterns, neurological deficits, and performance status, shape the models' decision-making process.

By illustrating the direction (positive or negative influence) and magnitude of each feature's contribution, SHAP facilitates the identification of the key factors most strongly associated with advanced tumor classification. Such interpretability is crucial in clinical environments, where understanding the rationale behind a model's prediction is as important as achieving high predictive

Table 2. Model performance was evaluated on the test set.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Logistic Regression	82.85	78.48	64.17	70.61
Naïve Bayes	82.85	83.30	58.26	68.56
AdaBoost	85.50	82.23	69.94	75.59
LightGBM	86.05	84.18	69.63	76.61

accuracy. SHAP-based insights thus support both clinician trust and informed clinical decision-making.

3. Results and Discussion

The predictive performance of the four machine learning models was evaluated using accuracy, precision, recall, and F1-score on the independent test set. The results are summarized in Table 2.

The comparative results indicate that ensemble-based learning methods outperformed linear and probabilistic models in predicting advanced tumor stages. LightGBM achieved the highest overall performance across all metrics, with an accuracy of 86.05% and an F1-score of 76.61%, reflecting its ability to capture complex nonlinear patterns within the clinical and radiological features.

AdaBoost also demonstrated strong predictive capability, achieving an accuracy of 85.50% and an F1-score of 75.59%, outperforming both Logistic Regression and Naïve Bayes in recall and overall balance between precision and sensitivity. This highlights the benefit of adaptive boosting in handling misclassified samples and refining decision boundaries.

In contrast, Logistic Regression and Naïve Bayes achieved comparable accuracy (82.85%), but their recall was notably lower, particularly for Naïve Bayes (58.26%). This reduction in sensitivity suggests limitations in capturing the nonlinear interactions and heterogeneous patterns characteristic of tumor progression. Although Naïve Bayes achieved the highest precision (83.30%), its low recall indicates a tendency to miss a substantial portion of advanced-stage cases, an undesirable outcome in clinical settings where early detection of advanced disease is crucial.

To further examine the classification behavior of each model, confusion matrices were generated for the test set (Figure 1). These matrices provide valuable insight into how well each algorithm distinguishes between non-advanced and advanced tumor stages, particularly in terms of false positives and false negatives, two critical error types in clinical decision-making.

Logistic Regression demonstrated moderate performance, correctly identifying 1,245 advanced cases but misclassifying 230 non-advanced instances as advanced. The relatively high number of false negatives (113 advanced cases predicted as non-advanced) indicates limited sensitivity, consistent with the model's lower recall.

Naïve Bayes produced the fewest false negatives among all models (75 misclassified advanced cases), reflecting stronger sensitivity. However, it also generated a comparatively higher number of false positives (268 non-advanced cases labeled as advanced), which aligns with its high precision but lower recall. This behavior suggests a bias toward predicting the advanced class, likely influenced by the dataset's class imbalance.

The AdaBoost model demonstrated improved balance between sensitivity and specificity, achieving 1,261 correct advanced predictions and reducing both false positives (193) and false negatives (97) relative to the linear and probabilistic models. This balanced behavior contributes to its higher F1-score compared to Logistic Regression and Naïve Bayes.

LightGBM achieved the strongest performance, correctly identifying 1,274 advanced cases while producing the fewest false positives (195) and the fewest false negatives (84). This result aligns with LightGBM's superior precision and F1-score, demonstrating its ability to model complex feature interactions and effectively separate the two classes.

To further evaluate the discriminative capability of the models across varying classification thresholds, Receiver Operating Characteristic (ROC) curves and Area Under the Curve (AUC) values were computed for each algorithm (Figure 2). The ROC–AUC metric provides a threshold-independent assessment of model performance by capturing the trade-off between the true positive rate (sensitivity) and the false positive rate.

As shown in Figure 2, all four models achieved AUC values above 0.84, indicating strong overall classification ability. LightGBM achieved the highest AUC (0.8692), indicating its superior ability to distinguish between non-advanced and advanced tumor stages. AdaBoost also performed competitively, achieving an AUC of 0.8642, consistent with its strong recall and F1-score reported earlier.

Logistic Regression achieved an AUC of 0.8503, outperforming Naïve Bayes despite similar accuracy levels. This suggests that Logistic Regression maintains

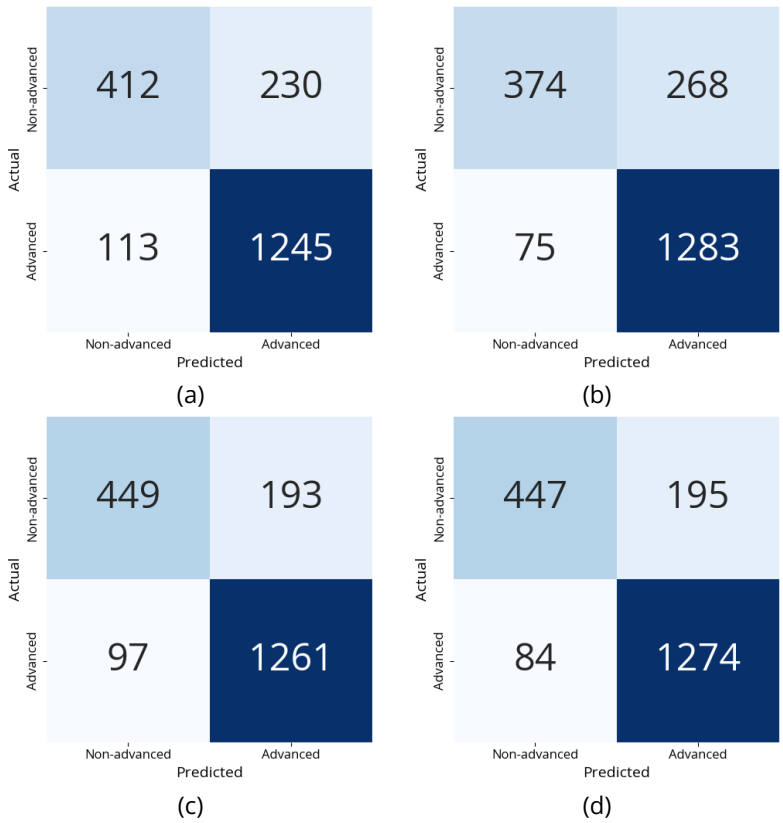

Figure 1. Confusion matrices of the four machine learning models evaluated on the test set: (a) Logistic Regression, (b) Naïve Bayes, (c) AdaBoost, (d) LightGBM.

more stable performance across different decision thresholds. Naïve Bayes, with an AUC of 0.8494, displayed the lowest score among the evaluated models, but still demonstrated good discriminative ability. Its curve shows relatively higher variability at lower false positive rates, which aligns with its observed tendency to misclassify non-advanced cases as advanced.

Figure 3 presents the precision–recall (PR) curves for the four evaluated models, offering a detailed view of performance in the context of class imbalance. Logistic Regression achieved the highest average precision (AP = 0.88), indicating strong consistency in identifying tumor-positive cases across different recall levels. AdaBoost and LightGBM displayed comparable performance (AP = 0.87 each), with both models maintaining high precision at moderate to high recall values, demonstrating reliable detection capability even when aiming to capture more positive cases. In contrast, Naïve Bayes achieved a lower AP (0.83) and exhibited sharper fluctuations in precision, suggesting greater sensitivity to changes in the decision boundary. Overall, the PR curves reaffirm the robustness of the ensemble methods and Logistic Regression in handling imbalanced tumor classification tasks.

To further assess model performance under class imbalance, Precision–Recall (PR) curves and Average Precision (AP) scores were generated for each classifier (Figure 3). Unlike ROC curves, PR curves provide a more informative evaluation when the positive class (advanced tumors) is more prevalent, and the cost of false negatives is clinically high.

As shown in Figure 3, all four models demonstrate strong and comparable performance, each achieving an Average Precision (AP) of 0.90. This consistently high AP score indicates that the models maintain a favorable balance between precision and recall across a wide range of thresholds. Notably, all models maintain precision above 0.85 across most recall levels, indicating their ability to correctly identify advanced cases while keeping false positives relatively low.

The LightGBM and AdaBoost curves exhibit greater stability across the mid-to-high recall ranges, suggesting improved robustness in identifying advanced tumors even when operating under more sensitive decision thresholds. Logistic Regression and Naïve Bayes, while performing similarly overall, show slightly more variability at lower recall levels, which aligns with earlier findings on their weaker sensitivity and limited capture of complex feature interactions.

To enhance interpretability and provide clinically meaningful insights into model behavior, SHAP analysis

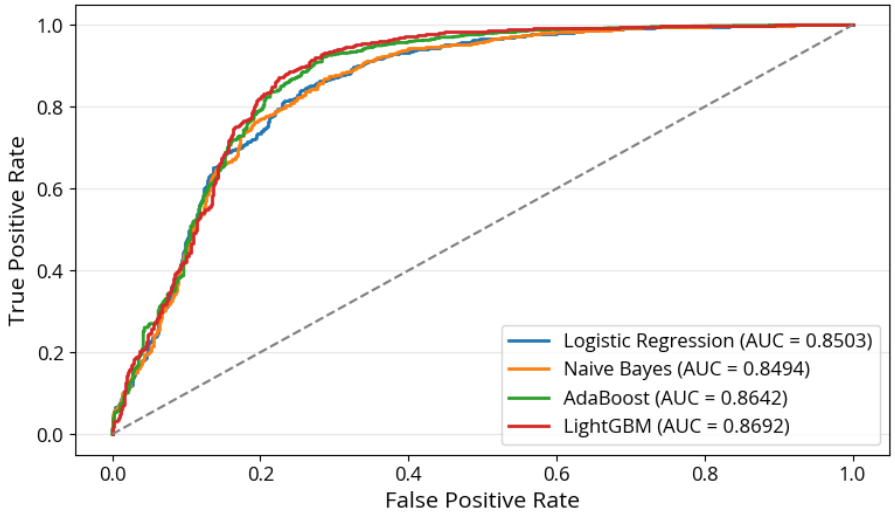

Figure 2. Receiver Operating Characteristic (ROC) curves for all machine learning models.

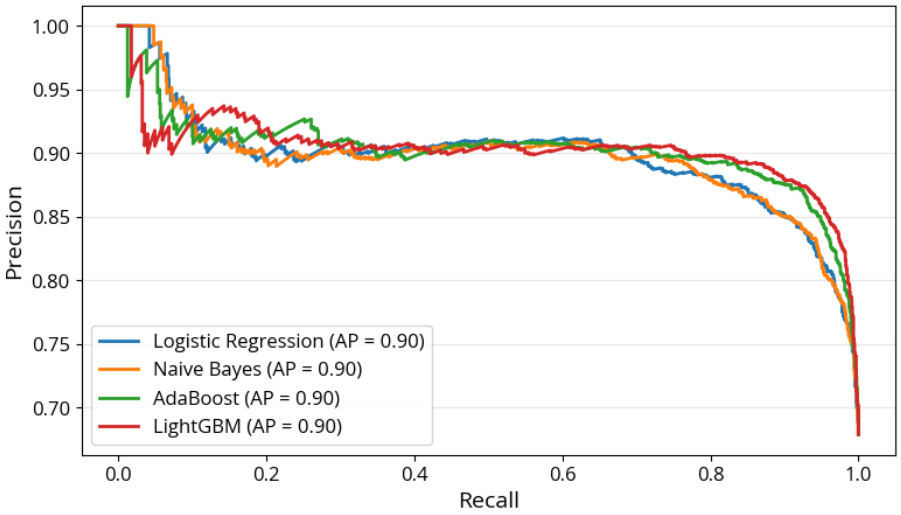

Figure 3. Precision–Recall (PR) curves for the four machine learning models.

was conducted for the best-performing classifier, LightGBM. The global feature importance rankings derived from mean absolute SHAP values are presented in Figure 4.

The SHAP summary reveals that mitotic_count is the most influential predictor of advanced tumor stage, exhibiting the highest average impact on model output. This aligns with established oncological evidence linking elevated mitotic activity to aggressive tumor behavior and poor prognosis. The ki67_index, another marker of cellular proliferation, ranks second in importance, reinforcing the role of proliferative activity as a key determinant of tumor advancement.

Radiological features also contributed significantly to the model's predictions. Attributes such as enhancement, necrosis, size, and hemorrhage were among the top contributors, reflecting their clinical relevance for identifying tumors with high metabolic activity, structural disruption, and rapid growth. These imaging-derived features are commonly associated with late-stage manifestations, explaining their strong influence within the predictive framework.

Mid-level contributors, including tumor_type, edema, location, and margins, provided additional structural and morphological context. Although less dominant individually, their combined influence supports the model's ability to distinguish nuanced variations in tumor presentation.

Demographic and clinical baseline variables such as age, kps_score, and symptoms_duration exhibited moderate importance, while neurological_deficit and gender had minimal impact on prediction outcomes. This suggests that biological and radiological markers more strongly

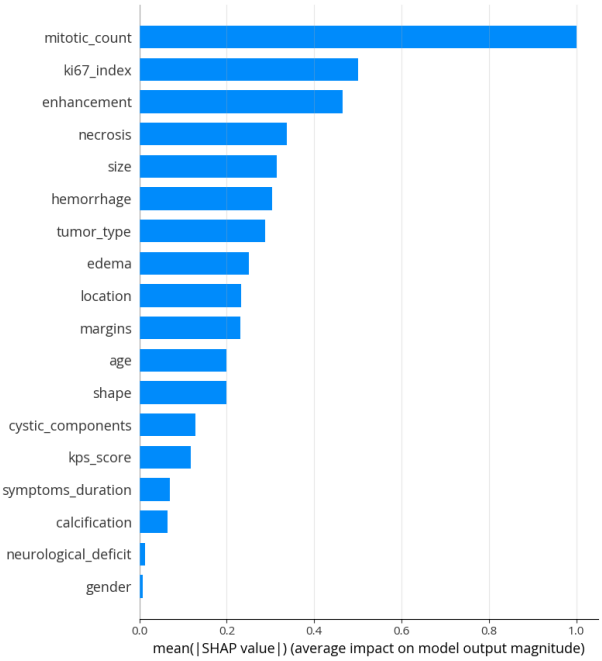

Figure 4. Global SHAP feature importance bar plot, showing the mean absolute SHAP value for each predictor.

drive advanced tumor stage in this dataset than by patient demographics or general clinical condition.

To further investigate how individual feature values influence model predictions, a SHAP beeswarm plot was generated for the LightGBM model (Figure 5). Unlike the global feature importance plot, which summarizes average impact, the beeswarm visualization illustrates the full distribution of SHAP values for each feature across all samples. This provides deeper insight into both global patterns and instance-level variability in model behavior.

The beeswarm plot confirms the dominant influence of mitotic_count and ki67_index, with high feature values (depicted in red) consistently pushing predictions toward the advanced class. This strong positive impact aligns with their biological role as markers of aggressive tumor proliferation. Conversely, lower values of these features (shown in blue) tend to shift predictions toward the non-advanced class, reflecting their association with slower-growing tumors.

Radiological features, including enhancement, necrosis, size, and hemorrhage, also demonstrate distinct SHAP value distributions. High enhancement and necrosis levels, for example, strongly drive predictions toward advanced stages, reflecting typical imaging characteristics of late-stage tumors. Variation in SHAP spread across these features indicates substantial heterogeneity in radiological presentation among patients, which the model effectively captures.

Features such as tumor_type, edema, location, and margins exhibit moderate but consistent contributions, with SHAP distributions showing both positive and negative impacts depending on specific clinical contexts. This suggests that these features influence model decisions in more nuanced ways, interacting with other attributes to refine stage prediction.

Lower-impact features, including age, shape, cystic_components, and kps_score, display narrower SHAP ranges, indicating a more modest influence on model outcomes. Neurological_deficit and gender show near-zero SHAP values across samples, underscoring their limited predictive utility within this dataset.

The findings of this study demonstrate the potential of machine learning, particularly ensemble-based approaches such as LightGBM and AdaBoost, to significantly enhance the prediction of advanced tumor stages using routinely available clinical and radiological features. The strong performance metrics, combined with the high interpretability provided by SHAP analysis, support the feasibility of deploying such models in clinical workflows to assist with the early identification of high-risk patients. By highlighting the most influential predictors, such as mitotic count, Ki-67 index, enhancement, and necrosis, the framework offers valuable decision-support insights that align with established oncological understanding. These insights could help clinicians prioritize diagnostic imaging, guide biopsy decisions, and stratify patients for timely therapeutic intervention.

Moreover, integrating explainability into predictive modeling strengthens clinician trust and promotes the safe adoption of artificial intelligence in medical environments. The model's ability to capture complex interactions between radiological, pathological, and demographic variables underscores its potential utility as a supplemental tool for precision oncology and personalized patient management.

Despite promising results, this study presents several limitations that should be acknowledged. First, the dataset was obtained from a single public source (Kaggle) and may not fully capture the heterogeneity present in real-world clinical populations. This introduces potential sampling bias and limits the generalizability of the findings across institutions with different imaging protocols, patient demographics, or diagnostic criteria.

Second, although the dataset included a broad range of clinical and radiological features, additional variables such as genomic markers, treatment histories, or longitudinal imaging data were not available. These factors may provide deeper insights into tumor

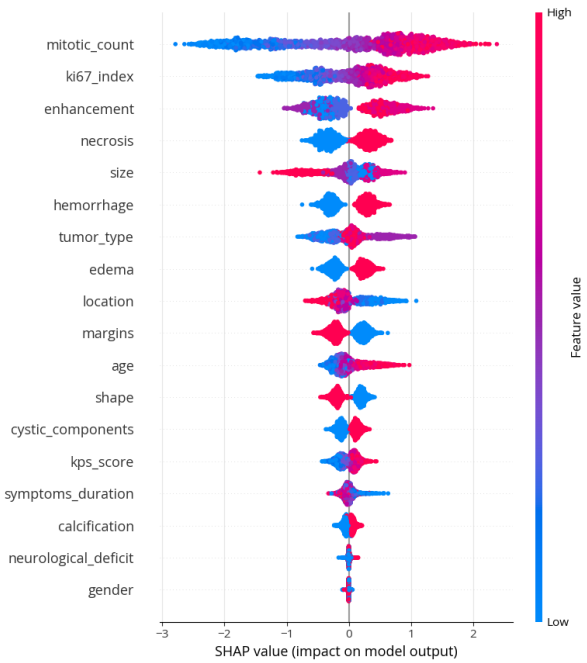

Figure 5. SHAP beeswarm plot showing the distribution of feature contributions across all samples.

progression and could further improve prediction accuracy.

Third, the study used default hyperparameters for all models to ensure consistent baseline comparisons. While this allowed fair benchmarking, more extensive hyperparameter tuning could potentially enhance model performance. Additionally, although class imbalance was addressed through stratified splitting, more advanced imbalance-handling techniques (e.g., SMOTE, cost-sensitive learning) were not explored in depth.

Finally, the lack of external validation limits the ability to assess the model's robustness in new clinical settings. Validation using multi-center datasets would provide a stronger indication of the framework's real-world applicability.

Future research should expand the dataset to include more diverse and representative patient populations, ideally by incorporating multi-institutional or longitudinal data. This would improve the robustness and generalizability of the predictive models. Integrating additional high-dimensional features, such as radiomics, genomics, or deep learning-derived histopathological embeddings, could further enhance predictive capability and capture underlying biological complexity.

Advanced hyperparameter optimization and model calibration techniques should be explored to refine performance and improve reliability across various decision thresholds. Investigating more sophisticated ensemble methods or hybrid deep learning architectures may also yield performance gains.

Finally, future studies should aim to conduct external validation and prospective clinical evaluations to determine the predictive framework's real-world impact. Incorporating clinician feedback into model refinement and developing user-friendly interfaces for clinical deployment would support translating this work into practical decision-support systems.

4. Conclusions

This study presented an explainable ensemble learning framework for predicting advanced tumor stages using clinical and radiological features. Among the evaluated models, LightGBM demonstrated the highest predictive performance, while SHAP analysis provided transparent insights into key determinants, including mitotic count, Ki-67 index, enhancement, and necrosis. The integration of accuracy-focused algorithms with interpretable machine learning strengthens the potential for clinical adoption, offering a reliable decision-support tool for early identification of high-risk patients. Although further validation across diverse, multi-center datasets is necessary, the findings highlight the promise of explainable AI for improving diagnostic precision and supporting personalized oncology care.

Author Contributions: Conceptualization, T.R.N. and R.I.; methodology, T.R.N.; software, T.R.N.; validation, M.P., F.P. and R.I.; formal analysis, T.R.N. and M.P; investigation, T.R.N. and M.P.; resources, M.P. and F.P.; data curation, F.P. and R.I.; writing—original draft preparation, T.R.N. and M.P.; writing—review and editing, F.P. and R.I.; visualization, T.R.N.; supervision, R.I.; project administration, R.I.; funding acquisition, R.I. All authors have read and agreed to the published version of the manuscript.

Funding: This study does not receive external funding.

Ethical Clearance: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset used in this study is publicly available on Kaggle and can be accessed at: https://www.kaggle.com/datasets/gayanddissanayake/brain-tumor-stages-dataset.

Conflicts of Interest: All the authors declare no conflicts of interest.

References

1. Crosby, D., Bhatia, S., Brindle, K. M., Coussens, L. M., Dive, C., Emberton, M., Esener, S., Fitzgerald, R. C., Gambhir, S. S., Kuhn, P., Rebbeck, T. R., and Balasubramanian, S. (2022). Early Detection of Cancer, *Science*, Vol. 375, No. 6586. doi:10.1126/science.aay9040.

2. Patwekar, F., Patwekar, M., and Kamal, M. A. (2025). Synergizing Phytonanotherapy and Complementary Medicine: Future

Horizons in Cancer and Diabetes Care, *Global Translational Medicine*, Vol. 4, No. 1, 16. doi:10.36922/gtm.5840.

3. Liu, B., Zhou, H., Tan, L., Siu, K. T. H., and Guan, X.-Y. (2024). Exploring Treatment Options in Cancer: Tumor Treatment Strategies, *Signal Transduction and Targeted Therapy*, Vol. 9, No. 1, 175. doi:10.1038/s41392-024-01856-7.

4. Noviandy, T. R., Alfanshury, M. H., Abidin, T. F., and Riza, H. (2023). Enhancing Glioma Grading Performance: A Comparative Study on Feature Selection Techniques and Ensemble Machine Learning, *2023 International Conference on Computer, Control, Informatics and Its Applications (IC3INA)*, IEEE, 406–411. doi:10.1109/IC3INA60834.2023.10285778.

5. Telloni, S. M. (2017). Tumor Staging and Grading: A Primer, 1–17. doi:10.1007/978-1-4939-6990-6_1.

6. Zhang, S., Xiao, X., Yi, Y., Wang, X., Zhu, L., Shen, Y., Lin, D., and Wu, C. (2024). Tumor Initiation and Early Tumorigenesis: Molecular Mechanisms and Interventional Targets, *Signal Transduction and Targeted Therapy*, Vol. 9, No. 1, 149. doi:10.1038/s41392-024-01848-7.

7. Patwekar, M., and Patwekar, F. (2025). Current Systems Biology Methods Used in Immunotoxicogenomics, *Immunotoxicogenomics*, Elsevier, 37–66. doi:10.1016/B978-0-443-18502-1.00011-0.

8. Zhou, R., Tang, X., and Wang, Y. (2024). Emerging Strategies to Investigate the Biology of Early Cancer, *Nature Reviews Cancer*, Vol. 24, No. 12, 850–866. doi:10.1038/s41568-024-00754-y.

9. Upadhyay, A. (2021). Cancer: An Unknown Territory; Rethinking before Going Ahead, *Genes & Diseases*, Vol. 8, No. 5, 655–661. doi:10.1016/j.gendis.2020.09.002.

10. Patwekar, M., and Patwekar, F. (2025). Lymphocyte Immunotherapy and Clinical Outcome in Recurrent Pregnancy Loss Patients, *Reproductive Immunogenetics: A Molecular and Clinical Overview*, Elsevier, 215–238. doi:10.1016/B978-0-443-13657-3.00015-5.

11. Sakkal, M., and Hajal, A. A. (2025). Machine Learning Predictions of Tumor Progression: How Reliable Are We?, *Computers in Biology and Medicine*, Vol. 191, 110156. doi:10.1016/j.compbiomed.2025.110156.

12. Al-Ewaidat, O. A., and Naffaa, M. M. (2025). Emerging AI- and Biomarker-Driven Precision Medicine in Autoimmune Rheumatic Diseases: From Diagnostics to Therapeutic Decision-Making, *Rheumato*, Vol. 5, No. 4, 17. doi:10.3390/rheumato5040017.

13. Kashyap, A., Rapsomaniki, M. A., Barros, V., Fomitcheva-Khartchenko, A., Martinelli, A. L., Rodriguez, A. F., Gabrani, M., Rosen-Zvi, M., and Kaigala, G. (2022). Quantification of Tumor Heterogeneity: From Data Acquisition to Metric Generation, *Trends in Biotechnology*, Vol. 40, No. 6, 647–676. doi:10.1016/j.tibtech.2021.11.006.

14. Rahnenführer, J., De Bin, R., Benner, A., Ambrogi, F., Lusa, L., Boulesteix, A.-L., Migliavacca, E., Binder, H., Michiels, S., Sauerbrei, W., and McShane, L. (2023). Statistical Analysis of High-Dimensional Biomedical Data: A Gentle Introduction to Analytical Goals, Common Approaches and Challenges, *BMC Medicine*, Vol. 21, No. 1, 182. doi:10.1186/s12916-023-02858-y.

15. Patwekar, M., Sehar, N., Patwekar, F., Medikeri, A., Ali, S., Aldossri, R. M., and Rehman, M. U. (2024). Novel Immune Checkpoint Targets: A Promising Therapy for Cancer Treatments, *International Immunopharmacology*, Vol. 126, 111186. doi:10.1016/j.intimp.2023.111186.

16. Rane, N. L., Paramesha, M., Choudhary, S. P., and Rane, J. (2024). Machine Learning and Deep Learning for Big Data Analytics: A Review of Methods and Applications, *Partners Universal International Innovation Journal*, Vol. 2, No. 3, 172–197. doi:10.5281/zenodo.12271006.

17. Noviandy, T. R., Maulana, A., Idroes, G. M., Emran, T. B., Tallei, T. E., Helwani, Z., and Idroes, R. (2023). Ensemble Machine Learning Approach for Quantitative Structure Activity Relationship Based Drug Discovery: A Review, *Infolitika Journal of Data Science*, Vol. 1, No. 1, 32–41. doi:10.60084/ijds.v1i1.91.

18. Sharma, A., Lysenko, A., Jia, S., Boroevich, K. A., and Tsunoda, T. (2024). Advances in AI and Machine Learning for Predictive Medicine, *Journal of Human Genetics*, Vol. 69, No. 10, 487–497. doi:10.1038/s10038-024-01231-y.

19. Noviandy, T. R., Idroes, G. M., and Hardi, I. (2025). Integrating Explainable Artificial Intelligence and Light Gradient Boosting Machine for Glioma Grading, *Informatics and Health*, Vol. 2, No. 1, 1–8. doi:10.1016/j.infoh.2024.12.001.

20. Noviandy, T. R., Idroes, G. M., and Hardi, I. (2025). An Interpretable Bayesian-Optimized XGBoost Framework for Neuropsychiatric Drug Candidate Classification, *Iran Journal of Computer Science*. doi:10.1007/s42044-025-00297-x.

21. Dissanayake, G. D. (2025). Brain Tumor Stages Dataset.

22. Garan, M., and Tidriri, K. (2022). A Data-Centric Machine Learning Methodology : Application, 1–21.

23. Amato, A., and Di Lecce, V. (2023). Data Preprocessing Impact on Machine Learning Algorithm Performance, *Open Computer Science*, Vol. 13, No. 1. doi:10.1515/comp-2022-0278.

24. Noviandy, T. R., Maulana, A., Idroes, G. M., Suhendra, R., Afidh, R. P. F., and Idroes, R. (2024). An Explainable Multi-Model Stacked Classifier Approach for Predicting Hepatitis C Drug Candidates, *Sci*, Vol. 6, No. 4, 81. doi:10.3390/sci6040081.

25. Zhang, T., Moro, S., and Ramos, R. F. (2022). A Data-Driven Approach to Improve Customer Churn Prediction Based on Telecom Customer Segmentation, *Future Internet*, Vol. 14, No. 3, 94. doi:10.3390/fi14030094.

26. Bai, Q., Su, C., Tang, W., and Li, Y. (2022). Machine Learning to Predict End Stage Kidney Disease in Chronic Kidney Disease, *Scientific Reports*, Vol. 12, No. 1, 8377. doi:10.1038/s41598-022-12316-z.

27. Ramzan, M., Sheng, J., Saeed, M. U., Wang, B., and Duraihem, F. Z. (2024). Revolutionizing Anemia Detection: Integrative Machine Learning Models and Advanced Attention Mechanisms, *Visual Computing for Industry, Biomedicine, and Art*, Vol. 7, No. 1, 18. doi:10.1186/s42492-024-00169-4.

28. Rahman, S., Irfan, M., Raza, M., Moyeezullah Ghori, K., Yaqoob, S., and Awais, M. (2020). Performance Analysis of Boosting Classifiers in Recognizing Activities of Daily Living, *International Journal of Environmental Research and Public Health*, Vol. 17, No. 3, 1082. doi:10.3390/ijerph17031082.

29. Noviandy, T. R., Idroes, G. M., Hardi, I., Afjal, M., and Ray, S. (2024). A Model-Agnostic Interpretability Approach to Predicting Customer Churn in the Telecommunications Industry, *Infolitika Journal of Data Science*, Vol. 2, No. 1, 34–44. doi:10.60084/ijds.v2i1.199.

30. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.-Y. (2017). Lightgbm: A Highly Efficient Gradient Boosting Decision Tree, *Advances in Neural Information Processing Systems*, Vol. 30.

31. Noviandy, T. R., Nainggolan, S. I., Raihan, R., Firmansyah, I., and Idroes, R. (2023). Maternal Health Risk Detection Using Light Gradient Boosting Machine Approach, *Infolitika Journal of Data Science*, Vol. 1, No. 2, 48–55. doi:10.60084/ijds.v1i2.123.

32. Tharwat, A. (2021). Classification Assessment Methods, *Applied Computing and Informatics*, Vol. 17, No. 1, 168–192. doi:10.1016/j.aci.2018.08.003.

33. Lundberg, S. M., and Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions, *Advances in Neural Information Processing Systems*, Vol. 30.