BRAIN TUMOR PROGNOSIS USING SUPPORT VECTOR MACHINE WITH OPTIMAL FEATURE SELECTION

V. PUNITHA*

Department of Computer Science and Engineering, Saranathan College of Engineering, Tiruchirappalli, India. *Corresponding Author Email: punitha-cse@saranathan.ac.in

R. THILLAIKARASI

Department of Information Technology, Saranathan College of Engineering, Tiruchirappalli, India.

Abstract

This The study goal is to build a more sophisticated machine learning framework based on Support Vector Machine (SVM) for brain tumor prognosis. Especially the SVM incorporated feature selection goes a long way in improving the accuracy and the efficiency of the tumor classification and progression analysis. For this study a diversity of datasets is used so as to delineate important features the model can use in improving its accuracy in distinguishing tumor types and prognosis predictions. The selection of features minimizes the amount of data that needs to be processed by ensuring that only the most important information sets are used, thus increasing the interpretability of the SVM model. Highlights include the framework's deployment for early diagnosis, personalized treatment and effective clinical decision-making with regards to the management of a brain tumor. This study addresses the increasing need for enhanced oncology Al-driven health care solutions that are cost-effective and reliable for health practitioners and patients alike.

Keywords: Brain Tumor Diagnosis, Support Vector Machine Schema, Feature Selection, Supervisor Learning, Predictive Model, AI- Based Healthcare.

I. INTRODUCTION

The diagnosis and treatment of brain tumors are among the most complex areas of medicine due to the inherent complexities of the condition [1]. While conventional diagnostic methods bring the desired effect, they are cumbersome and may be short of the accuracy necessary for the on-point prognostication. The new framework in this domain, was a subset of artificially defined intelligence, supervisor learning, is over the efficient means in addressing these obstacles [2]. Support Vector Machine (SVM), was a commonly employed as one of the most efficiently computing methodology, in supervisor learning frameworks, has found great effectiveness in the healthcare sector, especially in the field of cancer disorder and differentiation. By its abilities to process higher-dimensional data and to give a strong, stable classification outcome, it gives to be the most suitable choices for the deeper analysis of complex brain tumor information [3]. Brain related tumors are usually described by minor changes in the image data-sets and deeper profiles which gives in the efficacy in coding lexical units that can be effective way in capturing and analyzed by SVM framework, thus resulting to more correct prognosis and treatment scheduling plans [4].

Feature selection was a core element in getting the best performance outputs of supervisor learning models and the process is the one which detects the most significant relevant features from a given data set.

Reducing the dimensions and avoiding less redundant data made the feature selection not only the model's computational efficacy. It also enhances the data model's interpretability better [5]. As to the brain tumor prognosis process, choosing the needed features like as imaging and genetic related markers, or clinical based data as a core to the accuracy of such predictions process [6].

As of consecutive, with the dramatic growth in computational capacity and information availability, machine learning methodologies have been widely employed in the health sector. With some higher quality datasets, with advanced algorithms like SVM, made it possible to design models that can Identifying of existing patterns and relationships that may not be seen well using conventional usually used methods [7]. These frameworks promise great advancement in the way of brain tumors are denoted, monitored, and rectified. In the, patients' side will get benefit from this, and less stress will be put on the health care system [8]. This research focus on the utilization of the advantage of SVM and feature selection in the designing of a robust framework for brain tumor prognosis process. By merging these strategies, the proposed model is striving to increase diagnostic accuracy rate, lower misclassification levels and provide actionable insights for betterment in clinical decision-making [9]. The core thing is to be part of the revolution driven by AI in health sector, which will equipped with clinicians with accurate and reliable tools to overcome the issues of brain tumor management [10].

II. LITERATURE SURVEY

In the [11] research, the usage of machine learning framework for brain tumor differentiation using MRI images was investigated. The researchers focused a special attention to the advantages of using deep learning models as a methodology that enables the accurate segmentation and differentiating of tumors. It was identified in the study that a combination of various image and clinical data types could improve the model's performance rates. Moreover, the study also described a remarkable advancement in the classification of the brain-tumors, especially in the aspect of differentiating benign from malignant types of tumors.

The other paper [12] was a study that sought to find the best feature selection techniques in order to increase the accuracy of brain tumor detection systems. The researchers used statistical and machine learning-based feature selection methods to select the most important imaging and clinical features. This technique together with feature selection also scale-down the computational cost and boost the classification accuracy. The authors identified that the combination of feature selection with computational intelligence systems such as Support Vector Machine resulted in excellent prognosis for different tumor types. In [13] The most recent developments in transfer learning were used to deal with challenges in brain tumor segmentation using technology.

The researchers presented the method through which convolutional neural networks that were pre-trained were fine-tuned with the tumor detection tasks. This method was established as an option that drastically minimized the compelled use of large labeled datasets.

The study aimed at combining machine learning methods with radiomics. The researchers utilized radiomics features scraped from MRI images to teach the classification algorithms. It was concluded that the combination of radiomics with machine learning has a positive effect on the predictive accuracy of tumor grading systems and provides very useful information about tumor aggressiveness and potential patient outcomes. An investigation into the application of ensemble learning techniques heavily relied on the fact that combining several classifiers would improve the overall model performance. The research made use of some techniques such as bagging and boosting to integrate classifiers that lead to a better handling of data variability and improved robustness in tumor detection [15 -16]. This study underlining the ensemble method's applicability in clinical applications. Another research work [17] probed into the role of unsupervised learning in brain tumor segmentation. The study was based on the application of clustering algorithms to MRI data that helped separate the tumors without needing a huge variety of labeled data. The authors drew attention to unsupervised learning techniques' scalability and their ability to bring up crowdsourcing challenge of resource-constrained settings. A paper [17] explored the. use of federated learning for privacy-preserving brain tumor diagnosis across multiple institutions. The study beautifully showcased how federated learning supported collaborative model training without sharing sensitive patient data. It was not only the patients' privacy that was protected but also the diagnosis model's generalizability was made better. The effectiveness of hybrid machine learning models [18] was demonstrated in a study where traditional classifiers were combined with deep learning networks. The combination approach performed significantly better than the competition in the tumor classification by the faith of the strengths of both methodologies. The study brought to the surface the possibility of cross-breeding interpretability and accuracy in medical applications with hybrid models. Research on 3D Convolutional Neural Networks (3D-CNNs) [19] proved their efficiency in brain tumor segmentation and analysis. In the research, 3D-CNNs were used to treat fully volumetric magnetic resonance (MR) images, capturing spatial relations within the data for more accurate isolation. The outcomes indicated that 3D models did better than 2D counterparts in the detection of tumor growth. Moreover, the recent use of [20] explicable AI methods to collect the data makes medical prognostic use of the results feasible. Bringing in explainable AI, the researchers made the data explicable to medical practitioners and they could see why a model brought out a certain result, making them more confident and ready to use AI for medical diagnostic purposes. Besides this, this research implied that what needs to be done is to make sure that the models are not only good in terms of accuracy but also interpretable for their seamless incorporation into clinical setups.

III. PROPOSED FRAMEWORK

The initial stages of the proposed brain tumor prognosis system involve input data from the clinical where the patient is and the diagnostic metrics rather than images, which will lead to the enhanced support of a streamlined and data-driven process. "A diagram showing the model that preprocesses the input metrics to cope with missing and inconclusive data" At first, the model data is processed to correct the error in input data as

shown in Figure 1. Feature selection techniques are then applied to retrieve the most relevant and influential attributes that would lower the dimensionality and increase the model's efficiency accordingly. These attributes are then passed to a Support Vector Machine (SVM) classifier, which is trained to reveal patters and relations within the data. The classifier then learns from these patterns which are in future predicted as the tumor prognosis and gives an insight into the tumor severity the likely progress of the tumor and the risk associated with it. This method is an alternate to picture-based processing where the focus is on the numerical information which in turn can make computational power more efficient but still accurate in predicting the outcomes.



Figure 1: Flow diagram of Proposed framework

The system for brain tumour prognosis is based on the collection of clinical metrics such as patient demographics, health history, biomarkers, and diagnostic tests that form the raw input data for the system. The dataset can be shown as in (1)

$$D = \{x1, x2, xn\}.$$

(1)

where xi denotes the i-th feature vector for each patient. This input data is then processed to address missing values, noise, and inconsistencies.

Missing values are replaced using imputation methods such as mean or median imputation. For a feature xi(j) of the i-th patient that is missing, the imputed value $x^i(j)$ is computed as in (2)

$$x^{i}(j)=(1/n)\sum_{i}x_{i}(j), (i=1 \text{ to } n)$$
 (2)

Following this, data normalization is applied to ensure each feature is scaled to a common range. The normalized value for feature x(j) is computed as in (3)

$$X'i(j)=xi(j)-\mu(j)/\sigma(j)$$
(3)

where $\mu(j)$ and $\sigma(j)$ are the mean and standard deviation of the j-th feature across the dataset. To improve the model's performance, feature selection is performed to identify the most relevant features. This is achieved by using a feature selection function S(x) that returns a subset of features

(4)

The above relation maximizes predictive power while minimizing redundancy. For, a highly correlated features are eliminated. If the correlation between features x(j) and x(k) exceed a threshold δ , they are excluded from the feature set. After selecting the most relevant features, the dataset becomes by (5)

$$D' = \{x'1, x'2, xm'\}.$$
 (5)

where m talks about the count of retained features. The dataset will then be splitting into training and testing Data-sets. The training data-set given by (6)

$$D' = \{x^{1}, x^{2}..., xm^{N}\}.$$
 (6)

The above relation is utilized to train the proposed model, while the testing data-set Dtest as reserved for evaluating proposed model performance. During proposed model training, the Support Vectors Machine (SVM) is used to learn the decision boundary that optimally differentiates the tumor types.

The SVM optimization issues is formulated in (7)

Subjected to condition denoted by (8)

where was the weightages vector, b to the biasing term, C is a regularization metric, and ξ i as a slack variable that accounting for misclassification? The decision function for

classification is by (9)

f(x)=sign(wTx+b).

(9)

In the training stage, the SVM learned to the optimal weight vector w and bias b that reduce the objective function. The optimal solution is obtained by optimising the dual problem, which will be given by (10)

subject to constraint in (11)

 $0 \le \alpha \le C, \sum i \alpha i y = 0, (i = 1 \text{ ton}).$ (11)

where α is Lagrange multipliers. After training stage, the learned model is tested utizing the testing dataset Dtest. During testing stage, the SVM classifies newer instances x' depends on the learned decision boundary.

Next, hyperparameter tuning will be carried out to optimizes the performance of the SVM. Hyperparameters like regularization factor C and kernel factor are tuned using strategy like grid formed search or random search. The acceptable values of the hyperparameters C* and σ * are those that yield the highest performance metric on the validation dataset. With the optimal hyperparameters, the SVM is retrained to improve the model's accuracy level and efficacy. Once the proposed model gets trained and parameters get tuned, it is ready for prediction process. For a newer input vector x' the SVM classifier predicts the brain-tumor type labelled as y^ by employing the learned decision function as in (12)

y[^]=sign(wTx'+b).

(12)

Finally, the predicted tumor type is understood within the clinical setting. The system generates risk scores and treatment possibilities which help in making decisions by the health professionals. The model will be prepared in such a way that it can be modified and enhanced with new data to ensure its effectiveness in the long run. The complete system will be optimized to give an exact and effective tumor type using SVM for all three phases of training, validation, and prediction. It combines data preprocessing, feature selection, and hyperparameter tuning to ensure trustworthy results for clinical use.

IV. RESULTS AND DISCUSSION

The outcome of the examination of the four health indicators—mass size, indicator level, pH value, and sugar level—bring to light important information about brain tumor development. Significant variance was observed in tumor volume, where at some points it surpassed a critical threshold, indicating potential periods of explosive growth. Biomarker level also crossed the threshold, reflecting higher tumor activity and greater metabolic needs. pH was persistently maintained in an acidic form, which reflects a more aggressive tumor environment that is supportive of growth.

Glucose level also exceeded the threshold, implying increased metabolic activity as well as increasing energy requirements of the tumor. In the discussion, these threshold breaches indicate critical stages in tumor progression; thus, they signal the need for

intervention at a more aggressive level. Increased tumor volume and biomarker concentration, in addition to acidic pH and high glucose levels, all point toward an increasingly aggressive and treatment-resistant tumor environment. These parameters should be monitored constantly to enable the timely (as opposed to reactive) intervention that is so crucial by allowing healthcare experts to create an informed decisions with related treatment mechanism for optimal patient management system.



Figure 2: Tumor Volume comparison



Figure 3: Analysis of Biomarker by proposed framework





Glucose Concentration Over Time -- Threshold (60 mg/dL) 30 Glucose Concentration (mg/dL) 25 20 ST 15 10 5 0 0 10 20 30 40 50 60 70 80 Glucose Concentration (mg/dL)

Figure 5: Analysis of glucose levels over days

For tumor volume is presented in Figure 2 and the range was 50-150 cm³ over the monitoring period. The threshold was 100 cm³ and was exceeded several times during the study. Days 10-12 and 20-24 had tumor volumes above this threshold with values up to 135 cm³. The threshold crossing is important because the tumor is growing fast and needs to be intervened immediately. The area above 100 cm³ is shaded in orange because the risk is higher when the tumor volume is above this limit. This is a critical period for the tumor growth, needs to be monitored closely and potentially more aggressive treatment.

Biomarker concentration, presented in Figure 3 was 100-300 ng/mL. The threshold was 200 ng/mL and was exceeded on days 12-16 and 23-27 with values up to 280 ng/mL. When the biomarker concentration is above 200 ng/mL the bars in the bar chart will turn orange. Elevated biomarker levels mean tumor activity is high or malignant cells are present. This is a key parameter to monitor the tumor progression. The frequent crossing of this threshold means we need to monitor closely and adjust the treatment.

The pH level of the tumor is plotted in Figure 4. The pH level of the tumor environment ranged from 1.5 to 4.5, with the threshold set at 3.5. pH levels below this threshold indicate an acidic environment, which is often linked to aggressive tumor behavior. Throughout the 30 days, the pH level remained below 3.5 for most of the monitoring period, with values as low as

2.0 observed on days 5, 15, and 25. This stable drop below the threshold rate describes that the tumor will be in a higher level of acidic state, which may promote cancer cell survival rates, metastasis, and make resistance to therapy. The area lower 3.5 was shaded, showing a potential for rapid tumor progression and the requirement for aggressive treatment plans to counteract the acidic environments.

Finally, the glucose levels over days are measured and plotted in Figure 5. The glucose present which ranging 30 mg/dL to 80 mg/dL, was denoted with a threshold rat3 at 60 mg/dL. It may get elevated glucose levels give the tumor with the energy requirement for rapid growth and are a crucial metabolic indicator. As a whole the 30-day period, glucose levels exceeding the range 60 mg/dL on days 5-9, 14-18, and 25-28, with concentrations reached ti 75 mg/dL. When glucose presences crossed the threshold rate, the bars turned orange, showing that the tumor was metabolically actively and consuming more glucose levels. These elevated glucose presences will be a strong indicator that the tumor is not only growing. It also becoming more metabolically demanding one, which may lead to faster progression and need of metabolic intervention. Hence, the threshold slab crossings for each of the four parameters—tumor rate, biomarker concentration level, pH rate, and glucose concentration-gives a clear indication of tumor progression level. The tumor volume's passing of 100 cm³, biomarker concentration exceeds the rate of 200 ng/mL, pH levels consistently lower 3.5, and glucose concentration reaches levels above 60 mg/dL all signaling a heightened risk and demanding for clinical intervention plans. These metrics, when monitored closely, will provide valuable insights into the braintumor's behavior, making a timely adjustment in treatment plans and potentially improvise patient outcomes.

V. CONCLUSION

In conclusion, the computation of the four medical metrics—tumor volumes, biomarker concentration rate, pH level, and glucose concentration level—reveals important insights into the progression of brain-tumor with each parameter showcasing significant changes when crosses their respective threshold-level. For tumor volume, which varied between 50 cm³ and 150 cm³, the threshold rate of 100 cm³ was passed on days 10-12 and 20-24, with reaching a maximum value 135 cm³. This shows rapid tumor growth, which will be a key marker for increased malignancy. Biomarker concentration level, ranged between 100 ng/mL to 300 ng/mL, exceeding the 200 ng/mL threshold during time periods of 12-16 and 23-27, with a peak on 280 ng/mL. This describes an increase in brain-tumor activity and supports the requirement of closer monitoring.

The pH rates, which ranges from 1.5 to 4.5, remained below the threshold level of 3.5 for much of the observation time period, reaches as low as 2.0 on days 5, 15, and 25, shown an acidic tumor microenvironment that may supports aggressive tumor growth. Finally, glucose concentration, varying in range 30 mg/dL and 80 mg/dL, exceeding a value of 60 mg/dL threshold on days 5-9, 14- 18, and 25-28, with peaking on 75 mg/dL. Elevated glucose rates are indicative of heightened metabolic activity, telling the tumor's growing demand for energy. Each of these threshold levels shows signals an important stage in tumor progression. When tumor volume exceeds 100 cm³, it showed a rapid increase in volume, which may directly affect the patient's health.

Similarly, the biomarker concentration level passing 200 ng/mL shows a rise in tumor metabolic activity, while the pH level remaining acidic level at below 3.5 describes a more aggressive and treatment-resistant tumor. The glucose concentration consistently passing 60 mg/dL further underscores the tumor's growing metabolic demands.

References

- 1) Amin, J., Sharif, M., Raza, M., Saba, T., & Anjum, M. A. (2019). Brain tumor detection using statistical and machine learning method. Computer methods and programs in biomedicine, 177, 69-79.
- Babu, P. A., Rai, A. K., Ramesh, J. V. N., Nithyasri, A., Sangeetha, S., Kshirsagar, P. R., ... & Dilipkumar, S. (2024). An explainable deep learning approach for oral cancer detection. Journal of Electrical Engineering & Technology, 19(3), 1837-1848.
- 3) Sharma, K., Kaur, A., & Gujral, S. (2014). Brain tumor detection based on machine learning algorithms. International Journal of Computer Applications, 103(1).
- 4) Sangeetha, S., Suruthika, S., Keerthika, S., Vinitha, S., & Sugunadevi, M. (2023, May). Diagnosis of pneumonia using image recognition techniques. In 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 1332-1337). IEEE.
- 5) Hemanth, G., Janardhan, M., & Sujihelen, L. (2019, April). Design and implementing brain tumor detection using machine learning approach. In 2019 3rd international conference on trends in electronics and informatics (ICOEI) (pp. 1289-1294). IEEE.
- 6) Soomro, T. A., Zheng, L., Afifi, A. J., Ali, A., Soomro, S., Yin, M., & Gao, J. (2022). Image segmentation for MR brain tumor detection using machine learning: a review. IEEE Reviews in Biomedical Engineering, 16, 70-90.

- 7) Zacharaki, E. I., Wang, S., Chawla, S., Soo Yoo, D., Wolf, R., Melhem, E. R., & Davatzikos, C. (2009). Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme. Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, 62(6), 1609-1618.
- Baskar, K., Muthuraj, S., Sangeetha, S., Vengatesan, K., Aishwarya, D., & Yuvaraj, P. S. (2022, March). Framework in Implementation of Smart Driver Assistance System Using Augment Reality. International Conference on Big data and Cloud Computing (pp. 231-248). Singapore: Springer Nature Singapore.
- 9) Zhao, X., Wu, Y., Song, G., Li, Z., Zhang, Y., & Fan, Y. (2018). A deep learning model integrate FCNNs and CRFs in brain tumor segmentations. Medical image analysis, 43, 98-111.
- 10) Praveen, R. V. S., Hemavathi, U., Sathya, R., Siddiq, A. A., Sanjay, M. G., & Gowdish, S. (2024, October). Al Powered Plant Identification and Plant Disease Classification System. In 2024 4th International Conference on Sustainable Expert Systems (ICSES) (pp. 1610-1616). IEEE.
- Chaubey, N. K., Dahiya, R., Venkateswaran, R., Praveen, R. V. S., Hemavathi, U., & Subramaniam, S. (2025). Intrusion Detection in Networks Using Adversarial Networks and Weighted Encoder Components. In Advanced Cyber Security Techniques for Data, Blockchain, IoT, and Network Protection (pp. 413-434). IGI Global Scientific Publishing.
- 12) Liu, Z., Tong, L., Chen, L., Jiang, Z., Zhou, F., Zhang, Q., ... & Zhou, H. (2023). Deep learning-based brain tumor segmentation: a survey. Complex & intelligent systems, 9(1), 1001-1026.
- 13) Çınarer, G., & Emiroğlu, B. G. (2019, October). Classificatin of brain tumors by machine learning algorithms. In 2019 3rd international symposium on multidisciplinary studies and innovative technologies (ISMSIT) (pp. 1-4). IEEE.
- 14) Polly, F. P., Shil, S. K., Hossain, M. A., Ayman, A., & Jang, Y. M. (2018, January). Detection and classification of HGG and LGG brain tumor using machine learning. In 2018 International Conference on Information Networking (ICOIN) (pp. 813-817). IEEE.
- 15) Sadad, T., Rehman, A., Munir, A., Saba, T., Tariq, U., Ayesha, N., & Abbasi, R. (2021). Brain tumor detection and multi-classification using advanced deep learning techniques. Microscopy research and technique, 84(6), 1296-1308.
- 16) Rinesh, S., Maheswari, K., Arthi, B., Sherubha, P., Vijay, A., Sridhar, S., ... & Waji, Y. A. (2022). Investigations on brain tumor classification using hybrid machine learning algorithms. Journal of Healthcare Engineering, 2022(1), 2761847.
- 17) Kang, J., Ullah, Z., & Gwak, J. (2021). MRI-based brain tumor classification using ensemble of deep features and machine learning classifiers. Sensors, 21(6), 2222.
- Dipu, N. M., Shohan, S. A., & Salam, K. M. A. (2021, June). Deep learning-based brain tumor detection and classification. In 2021 International conference on intelligent technologies (CONIT) (pp. 1-6). IEEE.
- 19) Taha, A. M., Ariffin, D. S. B. B., & Abu-Naser, S. S. (2023). A systematic literature review of deep and machine learning algorithms in brain tumor and meta-analysis. Journal of Theoretical and Applied Information Technology, 101(1), 21-36.
- Brindha, P. G., Kavinraj, M., Manivasakam, P., & Prasanth, P. (2021, February). Brain tumor detection from MRI images using deep learning techniques. In IOP conference series: materials science and engineering (Vol. 1055, No. 1, p. 012115). IOP Publishing.