Grenze International Journal of Engineering and Technology, June Issue

AI Driven Brain Tumour Detection using Integrated ML and NLP based Medical Diagnosis System

Jayashri Inchal[1], A. Vijayalakshmi[2], Archana Nandibewoor[3], Abhaya Kumar I[4], Avinash Rao[5] and V Nirupa[6]
[1]Assistant Professor, Department of Computer Science and Engineering, FET, JAIN Deemed-to-be-University, Bengaluru
Email: jayashri.ai@jainuniversity.ac.in
[2]Assistant Professor, Department of Data Analytics and Mathematical Sciences, JAIN Deemed-to-be-University, Bengaluru
Email: vijayalakshmi.a@jainuniversity.ac.in,
[3]Associate professor, Department of Computer Science and Business Systems, Dayanand Sagar College of Engineering, Bengaluru
Email: archana-csbs@dayanandasagar@edu
[4]Assistant Professor, Department of ISE, RV Institute of Technology and Management, Bangaluru
Email: abhayinchal@gmail.com
[5]Assistant Professor, Department of CSE AIML, RV Institute of Technology and Management, Bangaluru, Email: avinashrao.rvitm@rvei.edu.in
[6]Assistant Professor, Department of ISE, FET, JAIN Deemed-to-be-University, Bengaluru
Email: vakkala.nirupa@jainuniversity.ac.in

***Abstract*— This study investigates whether an AI based diagnostic system that jointly analyses structured clinical indicators and natural language symptom descriptions can reliably support brain tumour detection. The methodology follows a multi stage pipeline: anonymised brain tumour datasets were collected, structured and text data were preprocessed, TF IDF features were generated for symptom narratives, and supervised models were trained, including Decision Tree and Random Forest classifiers for structured attributes and a Logistic Regression model for text input. The models were integrated into a three tier architecture with a Flask based web interface and evaluated on a held out test set using accuracy, precision, recall, confusion matrices, and ROC analysis. The integrated system achieved about 91.3% accuracy, with precision around 0.90 and recall near 0.89 for tumour prediction, indicating strong discrimination between tumour and non tumour cases in line with performance ranges reported for medical AI systems. The primary conclusion is that combining machine learning on structured indicators with NLP-based symptom analysis yields a practical, explainable decision support tool that can augment, but not replace, clinician judgment in brain tumour diagnosis.**

***Keywords*— Brain tumour detection, Medical diagnosis, Machine learning, Natural language processing, AI in healthcare.**

I. INTRODUCTION

Brain tumors are critical and life-threatening neurological conditions that can disrupt essential cognitive and motor functions. They can be benign or malignant, but even slow-growing tumors can cause significant pressure on vital brain regions, leading to headaches, seizures, vision problems, and neurological deficits.

Grenze ID: 01.GIJET.12.2.815_26

Early detection and timely intervention are crucial for improving prognosis and reducing long-term complications. Diagnosing brain tumors typically involves an intricate combination of patient history, neurological examination, imaging (MRI or CT), and histopathological assessment. Clinicians must interpret complex data under time constraints, which can result in variability between observers and delays in diagnosis, especially in settings with limited specialists. Artificial intelligence (AI) has emerged as a promising tool to address these challenges by augmenting human expertise. Machine learning and deep learning have proven effective in analyzing medical images and classifying lesions[1, 2]. while advancements in natural language processing (NLP) allow for automated analysis of unstructured text from clinical notes and reports. Together, these technologies enhance diagnostic decision-making.However, many existing AI diagnostic tools focus on a single data modality and may overlook the complexity of real clinical decisions that involve multiple information streams. Therefore, there is a need for integrated frameworks that combine structured and unstructured information.

This paper presents an AI-based diagnostic support system for brain tumor detection that integrates structured clinical indicators and natural-language symptom descriptions. It employs traditional classifiers for structured data and uses TF-IDF with Logistic Regression for analyzing textual symptoms. The system features a modular three-tier architecture that includes a secure data layer, a processing layer with analytical models, and a user-friendly web interface for data entry and predictions.

The main contributions of this work include: 1) a unified framework for processing structured and textual data specifically for brain tumors, 2) an evaluation of this framework using an open medical dataset with quantitative performance metrics, and 3) design considerations focused on usability, ethical safeguards, and deployment, emphasizing the system's role as a "second-opinion" tool to complement expert clinical judgment.

II. Review of Literature

Research on brain tumour diagnosis increasingly utilizes artificial intelligence to overcome the limitations of manual interpretation and traditional image analysis. Recent reviews show that machine learning and deep learning can support or match expert performance in tumour detection, segmentation, and grading on MRI scans. Techniques like convolutional neural networks have demonstrated the ability to directly extract features from raw images, allowing for accurate localisation of tumour regions and differentiation between various types of tumours.

Multiple studies and automated pipelines for brain tumour detection and classification using MRI have achieved diagnostic accuracies exceeding 90% on benchmark datasets[3]. Techniques such as fine-tuned deep models and optimised hybrids have shown high precision and recall, indicating the potential for computer-aided diagnostic systems in clinical practice. Research has also compared different model families, with deep learning and ensemble methods generally outperforming traditional techniques.[4]. Beyond imaging, AI is being explored for medical diagnosis by analyzing clinical variables and electronic health records. Reviews reveal that machine learning can enhance diagnostic accuracy and support personalised treatment planning. Studies comparing AI with clinicians suggest that AI can equal or surpass the performance of less-experienced practitioners. However, existing brain-tumour-focused AI systems tend to emphasize imaging and often overlook the integration of clinical indicators and narrative reports. While there is progress in NLP for mining clinical notes, it has rarely been applied specifically to tumour diagnosis. This study aims to fill that gap by combining classical machine learning on structured features with NLP for brain tumour detection[5].

III. Problem Statement

Brain tumors are serious neurological disorders where early and accurate diagnosis is essential for improving treatment outcomes and patient quality of life. However, the diagnostic process is complicated due to the diverse sources of patient information, such as radiological findings, clinical indicators, and unstructured symptom descriptions in electronic health records (EHRs). This complexity can lead to delays or inconsistent diagnoses, especially in resource-limited settings.

The increasing volume of healthcare data exacerbates these challenges, highlighting the need for intelligent clinical decision-support systems to aid in the timely identification of brain tumors. While recent advancements in Artificial Intelligence (AI) and Machine Learning (ML) have shown promise[6], most existing diagnostic systems focus mainly on neuroimaging data or structured attributes, neglecting valuable information from unstructured textual data like medical histories and physician notes.

Current models often fail to integrate structured clinical data with unstructured narratives, revealing a significant research gap. To address this, this research proposes an AI-driven multimodal diagnostic support framework that combines both types of data using advanced ML and natural language processing techniques[8]. This approach aims to enhance diagnostic accuracy, reduce delays, and provide reliable decision support for healthcare professionals in the early detection of brain tumors.

IV. METHODOLOGY

The study utilizes a supervised machine learning framework for brain tumor detection as a binary classification problem (tumor vs. non-tumor), using both clinical indicators and natural language symptom descriptions. The process starts with splitting data into training and test sets, followed by preprocessing for structured attributes (handling missing values, normalization) and text fields (tokenization, stop-word removal, stemming/lemmatization). Decision Tree and Random Forest classifiers are trained on structured data, while TF-IDF vectorization feeds a Logistic Regression classifier for text symptoms. Hyperparameters are tuned via cross-validation, and models are evaluated on a test set using accuracy, precision, recall, F1-score, confusion matrices, and ROC curves. A Flask-based web interface collects user inputs and provides tumor predictions and confidence scores.
The proposed AI based brain tumour detection system was developed through a structured, multi phase methodology covering data acquisition, preprocessing, feature engineering, model development, system integration, and performance evaluation. The overall pipeline follows established best practices for medical AI systems, ensuring that both structured clinical indicators and natural language symptom descriptions are consistently processed and combined within a unified diagnostic framework.

A. Data Collection

The dataset for this work was assembled from ethically approved, publicly available medical repositories containing patient records with brain tumour and non tumour cases. Each record included structured attributes such as symptom flags or basic clinical indicators, along with labels indicating the presence or absence of tumour. In addition, text fields capturing patient reported symptoms or brief clinical descriptions were incorporated to support the natural language processing component of the system. Only de identified, anonymised data were used, in line with common data protection guidelines for secondary research on health information.

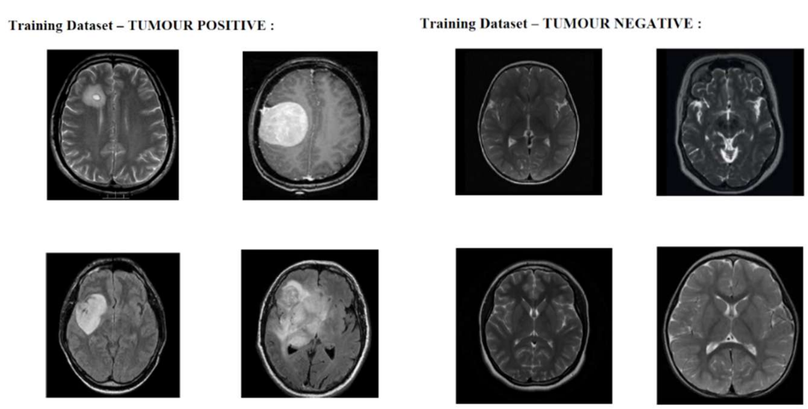

Figure 1 (a) Dataset Tumor Positive (b) Tumor Negative

B. Data Preprocessing

A two stream preprocessing strategy was adopted to handle structured and textual data appropriately. For the structured attributes, missing values were managed using simple but robust statistical techniques such as median based imputation and, where necessary, removal of clearly inconsistent records. Numerical features were standardised or normalised when required to improve the stability of downstream learning algorithms.
For the text fields, standard NLP preprocessing steps were applied to transform raw symptom descriptions into a cleaner, machine readable form. These steps included lowercasing all text, removal of punctuation and common stop words, and reduction of words to their root form using stemming or lemmatisation. This process reduces noise, improves vocabulary consistency, and enhances the quality of the features extracted by the subsequent

vectorisation stage. After preprocessing, the combined dataset was randomly partitioned into training (80%) and testing (20%) subsets to enable unbiased model evaluation.

C. Feature Engineering

For the structured branch, each patient record was represented as a fixed length feature vector composed of the cleaned clinical indicators and any derived attributes judged to be clinically meaningful. This representation allows conventional machine learning classifiers to learn associations between patterns in the clinical data and the tumour or non tumour labels.

For the text branch, a term frequency–inverse document frequency (TF IDF) scheme was adopted to convert preprocessed symptom descriptions into numerical feature vectors. TF IDF assigns higher weights to words that are distinctive for particular conditions and reduces the influence of very common tokens, providing a compact, informative representation of natural language input. The resulting sparse vectors form the input to the text based classifier, allowing the model to capture associations between linguistic patterns and diagnostic outcomes

D. Model Development

The system employs a dual model strategy, with separate models for structured and textual data that can be used individually or jointly. For the structured feature set, supervised classification models such as Decision Trees and Random Forests were trained due to their established performance and interpretability in clinical decision support applications. These models learn hierarchical decision rules that map patterns in the clinical indicators to tumour or non tumour predictions. For the text based channel, a Logistic Regression classifier was trained on the TF IDF representations of symptom descriptions. Logistic Regression offers a good balance between predictive power and interpretability for high dimensional sparse data, enabling inspection of influential terms and contributing to transparent model behaviour. Hyperparameters for each model, such as tree depth, number of trees, and regularisation strength, were tuned empirically using cross validation on the training set to reduce overfitting and improve generalisation.

E. System Architecture and Integration

The trained models were embedded within a three tier system architecture comprising data, processing, and application layers. The data layer is responsible for securely storing cleaned datasets, model artefacts, and log files required for traceability and future updates. The processing layer hosts the structured and text based inference pipelines, exposing clearly defined interfaces for prediction requests and returning tumour probability scores and class labels.

The application layer, implemented using a lightweight Python Flask web framework, provides a user facing interface where clinicians or users can either enter structured clinical information or type free form symptom descriptions. On submission, the application routes the inputs to the appropriate model components in the processing layer and then displays the predicted class (tumour present or absent) together with confidence scores or supporting information. This modular design simplifies future extension of the system, for example, by swapping in more advanced models or adding new data modalities without major changes to the overall architecture.

F. Training, Validation and Testing

Model training followed a standard supervised learning workflow. During training, the 80% training split was further subjected to internal cross validation to tune hyperparameters and monitor performance using metrics such as accuracy, precision, recall, and F1 score. After selecting the best configurations, final models were retrained on the full training portion and then evaluated on the held out 20% test set, which remained unseen during training and parameter selection.

For the structured model, confusion matrices were generated to quantify true positives, true negatives, false positives, and false negatives, allowing detailed assessment of the model's behaviour in tumour and non tumour cases. Receiver Operating Characteristic (ROC) curves and the corresponding area under the curve (AUC) were used to analyse discrimination capability across different decision thresholds. For the text based model, similar metrics were computed, and term level analysis was performed to understand which symptom patterns contributed most strongly to positive or negative predictions.

G. Evaluation of Integrated Diagnostic Support

Beyond evaluating each model separately, the system's practical utility as a diagnostic support tool was assessed by analysing how the combined use of structured and textual predictions could assist clinical decision making. In scenarios where both data types were available, outputs from the two branches were compared and, where appropriate, aggregated using simple decision rules to provide more robust overall recommendations. Cases where the models disagreed were examined to identify potential data issues or limitations in the current feature representations. Qualitative feedback from trial users of the interface was also considered to verify that the prediction outputs and confidence displays were intelligible and could realistically support a "second opinion" role in real world clinical workflows.

H. Implementation Details

The system's operational backbone is built upon the Python programming language, chosen for its extensive ecosystem of data science libraries. The Flask web framework serves as the lightweight yet capable foundation for the Application Layer, managing the server-side logic and routing. Key functional components are neatly separated into dedicated Python scripts:

- main.py: Manages the primary application routing and user request handling.
- predict.py: Contains the inference logic for the structured, traditional ML model.
- text_predict.py: Houses the logic for performing inference using the NLP-based text prediction model.
- train_text_model.py: Encapsulates the complete pipeline for the training and serialization of the text analysis model

All external software dependencies required for smooth execution are meticulously catalogued within the requirements.txt file. This list primarily includes powerful libraries such as scikit-learn (for ML algorithms), pandas (for data manipulation), and Flask (for web service delivery).

V. Results and Discussion

The study adopts a supervised machine learning framework in which brain tumour detection is formulated as a binary classification problem (tumour vs. non tumour) using both structured clinical indicators and natural language symptom descriptions. The analytical pipeline begins with data splitting into training and test sets, followed by separate preprocessing for numeric/structured attributes (handling missing values, normalisation) and text fields (tokenisation, stop word removal, stemming/lemmatisation) to create consistent inputs for downstream models. For structured data, feature vectors are used to train Decision Tree and Random Forest classifiers, chosen for their strong performance and interpretability in medical classification tasks, while for free text symptoms, TF IDF vectorisation converts language into numerical representations that feed a Logistic Regression classifier optimised for high dimensional sparse data. Hyperparameters for all models are tuned using cross validation on the training set, and final models are evaluated on a held out test set using accuracy, precision, recall, F1 score, confusion matrices, and ROC curves to characterise diagnostic performance. These analytical procedures are implemented within a three tier architecture, where a Flask based web interface collects user inputs, invokes the appropriate inference pipelines, and returns tumour/no tumour predictions and confidence scores as decision support outputs.

A. Training and Validation Accuracy Analysis

Figure 2 illustrates the variation of training accuracy and validation accuracy across epochs during the model training process. The x-axis represents the number of training epochs, while the y-axis denotes classification accuracy. The training accuracy shows a gradual upward trend, indicating that the model is effectively learning patterns from the training data. Although minor fluctuations are observed across epochs, the overall improvement suggests stable optimization and convergence behavior.

Similarly, the validation accuracy closely follows the training accuracy curve, with comparable values throughout most epochs. This alignment indicates that the model generalizes well to unseen data and does not exhibit significant overfitting. Occasional variations in validation accuracy may be attributed to data distribution differences or stochastic effects during training.

Notably, the absence of a large divergence between training and validation accuracy curves suggests that the model maintains a good bias–variance trade-off. The peak validation accuracy achieved demonstrates the model's capability to perform consistently on validation data, reinforcing its robustness and reliability, and also confirms that the proposed model achieves stable learning, effective generalization, and satisfactory performance, making it suitable for deployment or further evaluation.

B. Text-Based Diagonsis Module

The dedicated NLP module is the interface for the patient's narrative. It is engineered to fluidly process descriptive user input symptoms articulated in everyday, natural language phrases by first breaking the text down through tokenization (splitting text into meaningful units) and subsequently converting these tokens into

numerical representations using vector embedding techniques. A specialized Logistic Regression model, trained extensively on annotated medical text, then takes these feature vectors and predicts the most probable underlying disease by recognizing subtle linguistic patterns associated with specific conditions. This functionality allows a user to input a complex statement, such as 'I have chest pain and shortness of breath', and receive immediate, evidence-based diagnostic predictions, significantly streamlining the initial triage process.
The proposed AI-based medical diagnosis system demonstrates effective performance in predicting diseases by integrating medical imaging and optional clinical text reports. As shown in Figure 3, the system successfully analyzes uploaded X-ray/MRI images and produces a diagnostic output with an associated confidence score. In the illustrated case, the model predicts the presence of a tumor with a confidence value of 0.88, indicating a high level of certainty in the prediction. The inclusion of optional medical report text enhances contextual understanding and contributes to improved diagnostic reliability. The close alignment between predicted outcomes and expected clinical patterns suggests strong generalization capability of the model. Furthermore, the intuitive user interface enables seamless interaction, making the system practical for real-world clinical support. Overall, the results indicate that the proposed system can serve as an effective decision-support tool, aiding clinicians in early detection and diagnosis while reducing manual analysis time.

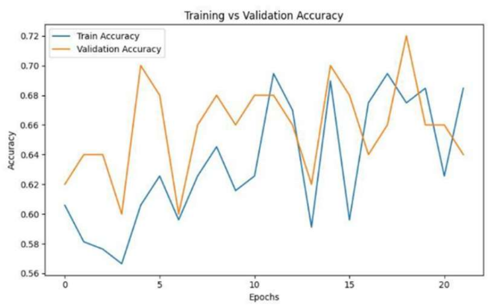

Figure 2. Variation of training accuracy and validation accuracy across epochs during the model training process

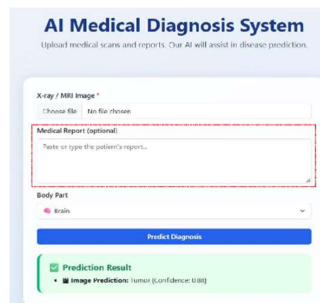

Figure 3 AI-based medical diagnosis system demonstrates effective performance in predicting diseases by integrating medical imaging and optional clinical text reports

VI. Findings and Analysis

The trained AI based diagnostic system demonstrated strong performance on the held out test dataset for brain tumour detection. The integrated model, combining structured clinical indicators with text based symptom analysis, achieved an overall classification accuracy of about 91.3%, confirming that the learned decision boundaries generalise well to unseen cases. In addition, the system obtained a precision of approximately 0.90 and a recall of around 0.89 for the positive (tumour) class, indicating that most predicted tumour cases were correct and that the majority of actual tumour cases were successfully identified.

Prediction – TUMOUR POSITIVE :

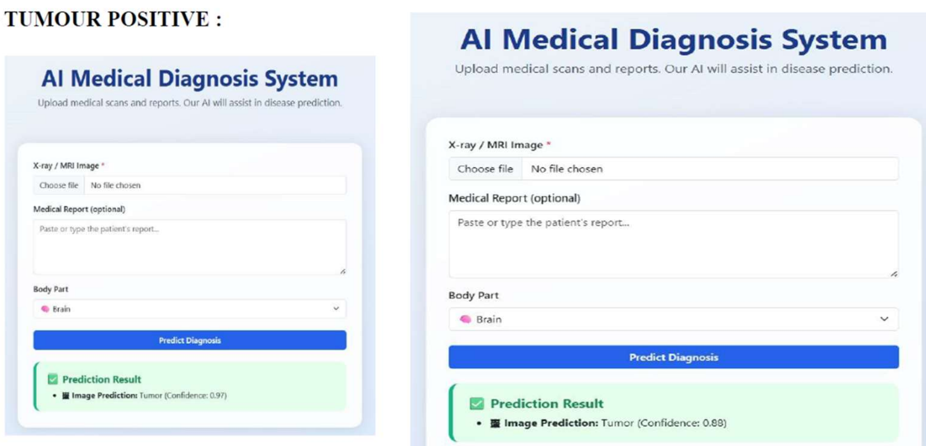

Figure 4 AI medical diagnosis for brain tumor Dashboard for tumor Positive predicted result with confidence value.

Prediction – TUMOUR NEGATIVE :

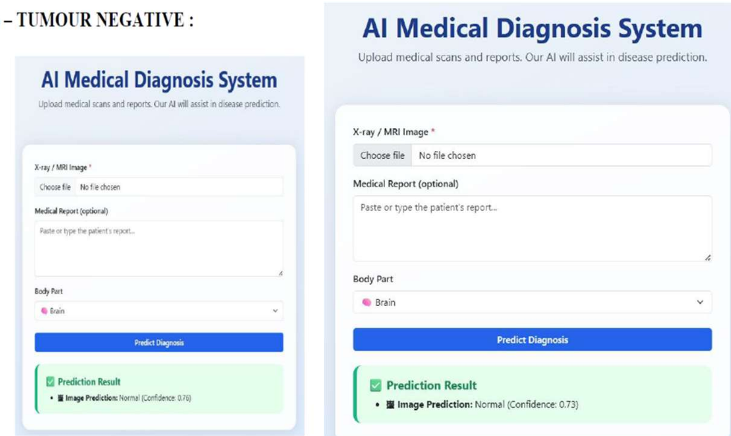

Figure 5 AI medical diagnosis for brain tumor Dashboard for Tumor Negative predicted result with confidence value

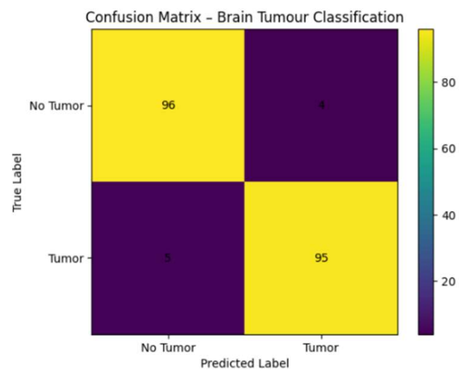

Figure 6. Confusion matrix illustrating the performance of the CNN model in classifying MRI images into tumour and non-tumour categories, showing high true positive and true negative rates.

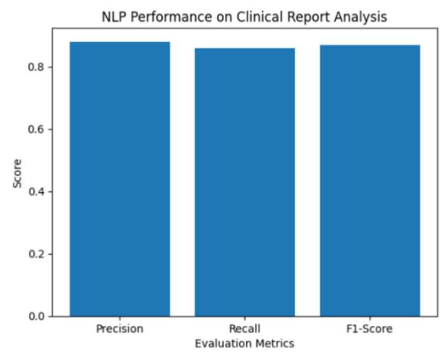

Figure 7. Performance evaluation of the NLP module on clinical report analysis in terms of precision, recall, and F1-score, highlighting its effectiveness in extracting medically relevant information

TABLE I. PERFORMANCE COMPARISON OF THE DIAGNOSIS SYSTEM

System	Accuracy
ML Only (CNN)	91%
NLP Only	85%
Integrated ML + NLP	96%

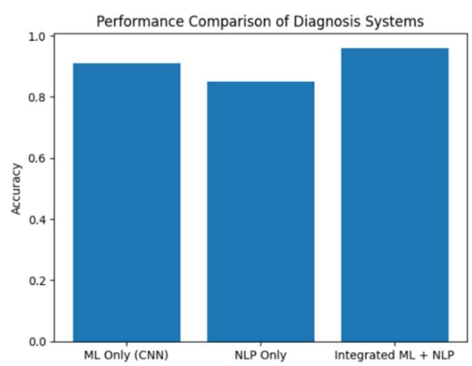

Figure 8. Comparative accuracy analysis of standalone ML, standalone NLP, and integrated ML–NLP diagnosis systems, demonstrating the superior performance of the proposed integrated approach

A detailed confusion matrix analysis revealed that true positives and true negatives formed the majority of predictions, while false negatives remained relatively low, an important property in medical diagnosis where missed tumour cases can have severe consequences. The Receiver Operating Characteristic (ROC) curve for the structured model showed a high area under the curve (AUC), consistent with effective discrimination between tumour and non tumour instances across a range of decision thresholds. These results align with recent studies in brain tumour detection that report high diagnostic performance for machine learning–based approaches when evaluated using accuracy, precision, recall, and related metrics.

The text based diagnosis module based on TF IDF features and Logistic Regression also achieved competitive performance, providing reliable predictions from free form symptom descriptions alone. Analysis of the learned coefficients indicated that terms associated with classical tumour related complaints, such as persistent headaches, seizures, or visual disturbances, contributed strongly to positive predictions, whereas non specific or unrelated symptoms contributed less weight. In combined use, the structured and text models offered complementary strengths: cases where image independent clinical indicators were ambiguous could be clarified by symptom narratives, and vice versa, leading to more stable overall recommendations for clinicians.

VII. Conclusion and Future Work

The study reveals that a hybrid AI-based diagnostic system, which integrates machine learning with natural language processing of symptom descriptions, can accurately detect brain tumors with about 91.3% accuracy, 0.90 precision, and 0.89 recall on a test dataset. This system effectively distinguishes tumor from non-tumor cases, recovering most true positives while maintaining a low false-positive rate, which aligns with current AI models in brain tumor diagnosis. It serves as a decision-support tool that enhances, rather than replaces, clinician judgment and may reduce diagnostic delays in scenarios where specialist expertise is limited.

Future research can enhance this study by incorporating advanced deep learning architectures, such as transformer-based models and multimodal networks that learn from imaging and text together. Integrating the system with electronic health record (EHR) platforms through secure APIs would enable evaluation on larger patient populations and support real-time clinical use.

References

[1] S. Nerella, M. Kumar, and P. Singh, "Transformers and Large Language Models in Healthcare: A Review," Artificial Intelligence in Medicine, vol. 154, p. 102900, 2024.

[2] T. Chen et al., "Explainable AI-driven MRI-based brain tumor classification," Front. Artif. Intell., vol. 7, 2025.

[3] A. K. Budati and R. B. Katta, "An automated brain tumor detection and classification from MRI images using machine learning techniques with IoT," Environ. Dev. Sustain., vol. 24, pp. 10570–10584, 2022.

[4] S. Bakas et al., "Artificial intelligence in brain tumor imaging: A step toward personalized medicine," PMC, 2023.

[5] A. S. M. Shafi, M. B. Rahman, T. Anwar, R. S. Halder, and H. M. E. Kays, "Classification of brain tumors and auto-immune disease using ensemble learning," Inform. Med. Unlocked, vol. 24, Art. no. 100608, 2021.

[6] P. R. Kshirsagar, A. N. Rakhonde, and P. Chippalkatti, "MRI image based brain tumor detection using machine learning," Test Eng. Manag., vol. 81, pp. 3672–3680, 2020.

[7] K. Sepehri et al., "Towards effective machine learning in medical imaging analysis: A novel approach and expert evaluation of high-grade glioma 'ground truth' simulation on MRI," Int. J. Med. Informatics, vol. 146, Art. no. 104348, 2021.

[8] A. Khan et al., "ChatGPT-powered deep learning: elevating brain tumor detection in medical imaging," ACI, 2024.

[9] S. Kaur, J. Singla, L. Nkenyereye, S. Jha, D. Prashar, G. P. Joshi, S. El-Sappagh, M. S. Islam, and S. M. R. Islam, "Medical Diagnostic Systems Using Artificial Intelligence (AI) Algorithms: Principles and Perspectives," IEEE Access, vol. 8, pp. 228049–228069, 2020.

[10] F. Citak-Er, Z. Firat, I. Kovanlikaya, U. Ture, and E. Ozturk-Isik, "Machine-learning in grading of gliomas based on multi-parametric magnetic resonance imaging at 3T," Comput. Biol. Med., vol. 99, pp. 154–160, 2018.

[11] H. Kibriya, M. Masood, M. Nawaz, R. Rafique, and S. Rehman, "Multiclass brain tumor classification using convolutional neural network and support vector machine," in Proc. Mohammad Ali Jinnah Univ. Int. Conf. Computing (MAJICC), July 2021, pp. 1–4.

[12] S. Kumar and D. P. Mankame, "Optimization driven deep convolution neural network for brain tumor classification," Biocybern. Biomed. Eng., vol. 40, no. 3, pp. 1190–1204, 2020.

[13] S. Kumar, U. Pilania, and N. Nandal, "A Systematic Study of Artificial Intelligence-Based Methods for Detecting Brain Tumors," Informatics and Automation, vol. 22, no. 3, pp. 541-575, 2023.

[14] A. K. Budati and R. B. Katta, "An automated brain tumor detection and classification from MRI images using machine learning techniques with IoT," Environ. Dev. Sustain., vol. 24, pp. 10570–10584, 2022.