IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 3 October 2023, accepted 26 November 2023, date of publication 28 November 2023,
date of current version 5 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3337363

RESEARCH ARTICLE

Brain Tumor Detection Using 3D-UNet Segmentation Features and Hybrid Machine Learning Model

BHARGAV MALLAMPATI[1], ABID ISHAQ[2], FURQAN RUSTAM[3], VENU KUTHALA[4], SULTAN ALFARHOOD[5], AND IMRAN ASHRAF[6]

[1]Department of Engineering, University of North Texas, Denton, TX 76205, USA
[2]Department of Computer Science and Information Technology, The Islamia University of Bahawalpur, Bahawalpur, Punjab 63100, Pakistan
[3]School of Computer Science, University College Dublin, Dublin 4, D04 V1W8 Ireland
[4]Master Card Technologies LLC, Blvd, O'Fallon, MO 63368, USA
[5]Department of Computer Science, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia
[6]Department of Information and Communication Engineering, Yeungnam University, Gyeongsan-si 38541, South Korea

Corresponding authors: Imran Ashraf (ashrafimran@live.com) and Sultan Alfarhood (sultanf@ksu.edu.sa)

This work was supported by King Saud University, Riyadh, Saudi Arabia, through the Researchers Supporting Project RSPD2023R890.

ABSTRACT Machine learning has significantly improved disease diagnosis, enhancing the efficiency and accuracy of the healthcare system. One critical area where it proves beneficial is diagnosing brain tumors, a life-threatening disease, where early and accurate predictions can save lives. This study focuses on deploying a machine learning-based approach for brain tumor detection, utilizing Magnetic Resonance Imaging (MRI) features. We train the proposed model using 3D-UNet and 2D-UNet segmentation features extracted from MRI, encompassing shape, statistics, gray level size zone matrix, gray level dependence matrix, gray level co-occurrence matrix, and gray level run length matrix values. To improve performance, we propose a hybrid model that combines the strengths of two machine learning models, K-nearest neighbor (KNN) and gradient boosting classifier (GBC), using soft voting criteria. We combine them because, in cases where KNN exhibits poor performance for certain data points, GBC demonstrates significant performance, and vice versa, where GBC shows poor results, KNN performs significantly better. With 2D-UNet segmentation features, the model achieves a 64% accuracy. By training it on 3D-UNet segmentation features, we achieve a significant accuracy of 71% which surpasses existing state-of-the-art models that utilize 3D-UNet segmentation features.

INDEX TERMS Brain tumor detection, MRI features, machine learning, ensemble learning, segmentation.

I. INTRODUCTION

Recently a GLOBOCAN survey across 185 countries reported a staggering 250,000 deaths attributed to brain tumors in 2020, with an estimated 300,000 new brain tumor cases [1]. Glioblastoma Multiforme (GBM) is the most lethal type of brain tumor with limited treatment options and low survival chances. In the United States (US), there are over 13,000 reported GBM cases annually, resulting in more than 10,000 deaths [2]. GBM has been classified as grade five cancer by the World Health Organization (WHO), the highest grade for any brain tumor, known for its aggressive nature. GBM has a post-diagnosis survival rate of only 5% up to 5 years, with an average survival duration of just 15 months [3]. Almost all GBM patients experience relapse despite aggressive therapies like temozolomide, maximal safe surgical resection, chemotherapy, and radiotherapy. The complexity of GBM tumors, including factors like temporal and spatial intra-tumor heterogeneity, location, and extent, presents significant challenges for resection and sometimes renders complete removal impractical. The poor prognosis and treatment outcomes are attributed to difficulties in delivering drugs to the brain and the inability to achieve complete surgical tumor resection [4].

The associate editor coordinating the review of this manuscript and approving it for publication was Jinhua Sheng.

Certain biological factors such as specific gene mutations like IDH1, 1p/19q deletion, and the MGMT promoter methylation status, have been found to partially explain the variations observed in GBM. Additionally, factors like gender and age of the patient also play a significant role. However, the diverse clinical behaviors shown by GBM suggest that considering it as a single disease is too simplistic. Researchers have made various attempts to classify GBM into distinct subcategories based on molecular subtyping, aiming to strengthen our understanding [5]. Yet, due to the expensive and complex nature of the tests, implementing this approach in routine clinical practice remains a challenge. Moreover, GBM tumors can change and evolve between different subtypes during their progression, making the task of predicting patient survival very challenging [6].

MRI (Magnetic Resonance Imaging) is a crucial tool in neuro-oncology, providing valuable support for initial diagnosis and treatment response assessment. It serves as a non-intrusive prognostic tool, with several studies linking MRI features to the survival of high-grade glioma patients [7], [8], [9]. The uniqueness of MRI lies in its ability to predict outcomes independent of pathological and clinical data. Moreover, it has expanded its applications to create non-invasive imaging biomarkers based on cellular and molecular characteristics. By linking imaging phenotypes with genomic signatures, MRI can serve as a non-invasive biomarker of cellular gene expression [10], [11]. This capability enables the connection between molecular and imaging diagnostics [12], offering the potential to refine diagnosis, prognostication, and identify specific therapeutic targets [13].

MGMT functions as a DNA repair enzyme. It plays a crucial role in safeguarding tumor cells against damage caused by alkylating agents, consequently contributing to their resistance to chemotherapy involving such agents. Additionally, MGMT serves as a protective shield for tumor cells, protecting them from the potentially lethal impacts of alkylating agent-based chemotherapies like dacarbazine (DTIC) or temozolomide (TMZ), which are extensively employed in treating melanoma and glioblastoma. Determining the MGMT promoter methylation status through invasive surgeries is a time-consuming and iterative process, requiring brain tissue biopsies and weeks of genetic analysis, which can have potential side effects. However, a non-invasive alternative exists by applying computer vision techniques to analyze MRI data. Recent advancements in deep learning have shown great success in various fields, making it promising for detecting MGMT promoter methylation by extracting biomarkers and patterns from MRI scans [14]. This paper introduces an ensemble machine learning architecture designed to be lightweight yet shows significantly improved performance for brain tumor classification Class 0 (representing No Brain Tumor) and Class 1 (representing Brain Tumor). The key contributions of this study are as follows

- This study presents an efficient approach utilizing state-of-the-art machine learning techniques to classify MRI scans as having a tumor or not.
- In this research, we propose a hybrid model for classification, selecting the best individual classifiers and combining them under soft voting criteria to address each other's weaknesses. The final prediction by the hybrid model is based on prediction probabilities.
- The study focuses on 2D-UNet and 3D-UNet MRI segmentation features, conducting an extensive analysis of machine learning models on these features to predict the MGMT values in the dataset. We illustrate the significance of adopting a hybrid approach as it yields significant results for brain tumor classification, leveraging the 3D-UNet segmentation features.
- We perform an in-depth comparison between machine learning and deep learning architectures for brain tumor classification using various evaluation metrics.

II. RELATED WORK

In recent years, there has been a surge in the development of machine learning models aimed at the diagnosis of GBM. These models are involved in predicting brain tumor disease using the MRI features and they can also extract multiple radiomic features from MRI images, such as area, volume, texture, and Euler characteristics, derived from different regions within the tumor. In this section, we explore the latest research in the field of brain tumor prediction, which serves as the foundation for the current study.

The study [15] introduced a brain tumor classification framework utilizing a hierarchical deep-learning neural network (HieDNN) classifier. By applying preprocessing techniques and feature extraction methods, the proposed approach achieves significantly improved accuracy compared to existing methods. For instance, it attains higher accuracy of 31.14%, 16.09%, and 11.48% for benign tumors, 35.18%, 19.17%, and 22.80% for malignant tumors, and 44.20%, 29.97%, and 20.44% for normal brain tumors. Along the same lines, [16] addresses the classification of gliomas, a type of brain tumor, into high-grade and low-grade categories to improve treatment planning. Using MRI images from the Brats dataset, it extracts essential features via pyradiomics and wavelet filters. Various classifiers are employed, with the SVM and factor analysis reduction demonstrating superior stability and an impressive mean accuracy of approximately 97%.

MRI data offers multiple modalities that effectively train deep neural network (DNN) models. Researchers have explored various techniques to achieve better predictive performance. For example, Myronenko [17] utilized an AutoEncoder to combine inputs from different modalities, resulting in improved 3-dimensional magnetic resonance imaging brain tumor segmentation. Tseng et al. [18] proposed a deep encoder-decoder design with cross-modality convolution layers for 3D image segmentation. Shachor et al. [19] introduced an ensemble network that incorporates multiple

data sources for sorting tasks. The network consists of three modality-specific encoders, capturing features at different levels, which can be extended to various views and scans. Nie et al. [20] used fully convolutional networks (FCNs) to segment isointense phase brain MRI images. The authors trained one network for each modality image and fused their high-layer features for segmentation. Kamnitsas et al. [21] directly fused input modality-wise information using a dual pathway, 3-dimensional, 11-layer deep convolutional neural network (CNN) for brain lesion segmentation. They also developed an efficient and effective dense training scheme, processing adjacent image patches in one pass. Another study [22] focused on medical image segmentation and used different datasets. The authors introduced a new, deeper, and more compact split-attention UNet model called DCSAU-Net for image analysis. The proposed DCSAU-Net effectively used high and low-level information based on two models: the primary features conservation framework and the compact-attention block. The study's findings demonstrated that DCSAU-Net outperformed other methods in terms of mean crossing over dice and union coefficient, showing its superior performance in medical image segmentation.

The study [23] introduced a network called SPP-Net for brain tumor segmentation, which uses a combination of spatial pyramid pooling (SPP) and attention blocks instead of residual connections. This change allows the network to gather information from different down-sampling blocks, making the reconstruction process more comprehensive. The attention blocks also provide important context by combining global dependencies with local reconstruction. The recommended model, SOO-UNet, achieved an accuracy score of 0.883 on the Brats 2021 dataset. Another model called BiTr-Unet was developed by [24] for brain tumor segmentation using multi-modal MRI data. BiTr-Unet combines CNN and transformer architecture. On the BraTS 2021 validation dataset, BiTr-Unet showed impressive performance with mean Dice scores of 0.9076, 0.8392, and 0.8231 for the whole tumor, tumor core, and enhancing tumor, respectively. Additionally, it achieved mean Hausdorff distances of 4.5322, 13.4592, and 14.9963 for the same tumor regions.

Along the same lines, Elmezain et al. [25] introduced an ensemble model for automatic brain tumor segmentation, which combines a latent-dynamic conditional random field with a deep capsule network. The approach involves three distinct procedures: image pre-processing, image segmentation, and image post-processing. By utilizing this ensemble model, the authors achieved improved performance in segmenting brain tumors. In the domain of brain tumor radiogenomic classification, Qu and Xiao [26] proposed an attention-based multimodal CNN. The model integrates various techniques including multimodal feature aggregation, separable embedding, model-wise shortcut connections, and a lightweight attention mechanism, all designed to enhance overall performance. The proposed system surpassed the state-of-the-art systems in terms of accuracy value, establishing its superiority in brain tumor radiogenomic classification. In another study, Nauman et al. [27] explored the relationship between MRI images and the MGMT promoter state. They used various deep learning models including 2D and 3D CNNs, and vision transformers. Interestingly, the authors used the same dataset as the current study. The results showed that the proposed ResNet10-based model, applied to resized images of $256 \times 256 \times 64$, achieved an area under the curve (AUC) value of 0.58 during validation.

For the brain tumor classification Alturki et al. [9] proposed a machine learning-based system. They have used a variety of machine learning models such as RF, k-NN, SGD, ETC, DT, SVM, LR, GBC, and VC(LR+SGD) on the dataset obtained from the Kaggle. A customized CNN model is used for feature extraction. Results show that the proposed ensemble model VC(LR+SGD) achieved the highest performance. For the prognosis of brain tumors using the brain MRI images, Shah et al. [28] proposed a refined EfficientNet-B0 for brain tumor prediction and also employed data augmentation techniques to obtain higher-quality images. The proposed Efficient-B0 system achieved an accuracy of 98.87%. For the efficient classification of MRI images Daz-Pernas et al. [29] proposed a multi-scale CNN system for the efficient classification of brain tumors. The authors skipped the pre-processing step in this study. Multi-scale CNN achieved a tumor classification accuracy of 97.3%.

The above-discussed literature reports excellent results for brain tumor classification, however, presents several limitations and gaps in the field of brain tumor prediction using machine learning models. While various techniques have been explored, including AutoEncoder-based fusion of modalities, deep encoder-decoder designs, and individual state-of-the-art learning models, the results still lack consistent accuracy. Some studies have achieved promising results, but they often focus on specific modalities or use different datasets, making direct comparisons challenging. In order to address these limitations and fill the existing gaps, this study considers a comprehensive ensemble approach that combines various techniques and modalities to enhance the overall performance and accuracy of brain tumor prediction models. Additionally, efforts to standardize datasets and evaluation metrics can facilitate more reliable comparisons between different models and studies. To provide a comprehensive review of the discussed literature, Table 1 presents a critical summary of these works.

III. MATERIALS AND METHODS

In this section, we offer a comprehensive explanation of the brain tumor dataset and the methodology proposed for brain tumor classification. We present the architecture and components of the hybrid model designed for efficient classification. To implement the proposed approach, we utilize a machine with a 12th Gen Intel(R) Core(TM) i7-1265U 1.80 GHz processor running on the Windows operating system. The implementation is carried out on Jupyter Notebook using

TABLE 1. Critical overview of the discussed literature.

Ref	Classifier	Dataset	Performance	Aims	Limitation
[15]	HieDNN, CNN-BTC, 3D CNN-BTC, and DAEN-BTC	BraTS (Brain Tumor Segmentation)	Benign tumors: 31.14% Malignant tumors: 35.18% Normal brain tumors: 44.20%	Efficient framework for brain tumor classification using GLMC features.	The reported accuracy rates, while improved, may still leave room for further enhancement in the classification of benign and malignant tumors.
[16]	SVM, NB, DT, Bagging classifier, KNN, and LR	BraTS	97%	The research aims to classify gliomas into high-grade and low-grade categories. It uses radiomic features and machine learning techniques on MRI brain images to improve the efficiency of glioma grading	The choice of classifiers and feature reduction techniques may not be universally optimal, and further research might explore alternative approaches.
[22]	DCSAU-Net (U-Net + PFC + CSA)	RSNA-MICCAI Brain Tumor Dataset	0.788 value of DSC and 0.703 value of mloU	Hybrid deep learning model used to classify brain tumor using image features	During segmentation proposed model fails to separate the target from images due to the nuclei with tiny size, or high similarity between the foreground and background.
[23]	SPP-U-Net	RSNA-MICCAI Brain Tumor Dataset	0.883 Dice score and 7.84 Hausdorff distance	Deep ensemble model that employs UNet segmentation for the brain tumor classification.	Mechanism is only adapted to one particular model and cannot be adaptable to various other architectures
[24]	BiTr-Unet	RSNA-MICCAI Brain Tumor Dataset	Mean Dice Scores of 0.9076, 0.8392, and 0.8231; Mean Hausdorff Distances of 4.5322, 13.4592, and 14.9963	Hybrid deep ensemble model used that employs image segmentation techniques.	Time-consuming and computationally complex
[25]	CapsNet, CapsNet + LDCRF, CapsNet + LDCRF + Post-processing	RSNA-MICCAI Brain Tumor Dataset	0.91 Dice score using CapsNet + LDCRF +Post-processing	Deep ensemble model is used with some data processing techniques using the image features for brain tumor detection.	Slices of images are used
[26]	Attention-based multimodal CNN	RSNA-MICCAI Brain Tumor Dataset	61.09%	Deep learning model CNN used on the image features for the classification of brain tumor.	This study only considers the temporal association between same-modality data, while the relationship between different-modality data is not examined.
[27]	ResNet34, ResNet50, ResNet10, DenseNet121, 3D-EfficientNet, Custom ResNet, ROI Custom ResNet	RSNA-MICCAI Brain Tumor Dataset	AUC 0.58 ResNet10	Its objective is to use transfer learning models on the image features to predict mgmt promoter methylation.	Models with this kind of performance cannot be used in clinical settings.
[9]	RF, k-NN, SGD, ETC, DT, SVM, LR, GBC and VC(LR+SGD)	Kaggle	99.9% VC(LR+SGD)	CNN-based features are used from the dataset for the classification of brain tumor.	Only CNN features were used.
[28]	VGG16, InceptionV3, Xception, ResNet50, EfficientNet-B0, and InceptionResNetV2	Self-collected	98.87% EfficientNet-B0	Implement transfer learning model for the brain tumor detection.	Due to data preparation differences, they do not employ machine learning models
[29]	MultiScale CNN	Self-collected	97.3%	Brain tumor classification and segmentation	Skull and vertebral column parts are not removed and the variability of the three tumor types, which caused false positives in some images.

Python 3.9, Pandas 1.4.4, sklearn 1.0.2, and Tensorflow for conducting the experiments.

Figure 1 shows the proposed methodology for brain tumor prediction and the architecture of the hybrid Model. In the proposed approach, we initially obtained the 2D-UNet and 3D-UNet features dataset from Kaggle [30]. To ensure consistent feature values, we performed data scaling, which normalized the dataset values within specific ranges. This data scaling technique is particularly beneficial for distance-based models like KNN, as it enables them to enhance their performance after learning from standardized data. In our study, data scaling was applied uniformly to both the training and test datasets using the StandardScaler from the scikit-learn library [31]. In our study, we utilized the StandardScaler with its default settings. Therefore, the 'specific ranges' for normalization, in this case, refer to achieving a mean-centered distribution with a standard deviation of 1, as per the default behavior of the StandardScaler. Following the initial dataset preparation, we proceeded to split it into training and testing sets using a ratio of 90:10. Specifically, 90% of the data was dedicated to training the proposed hybrid model.

A. PROPOSED HYBRID MODEL

The hybrid model (HM) leverages the strengths of both KNN and gradient boosting machine (GBM) through soft voting

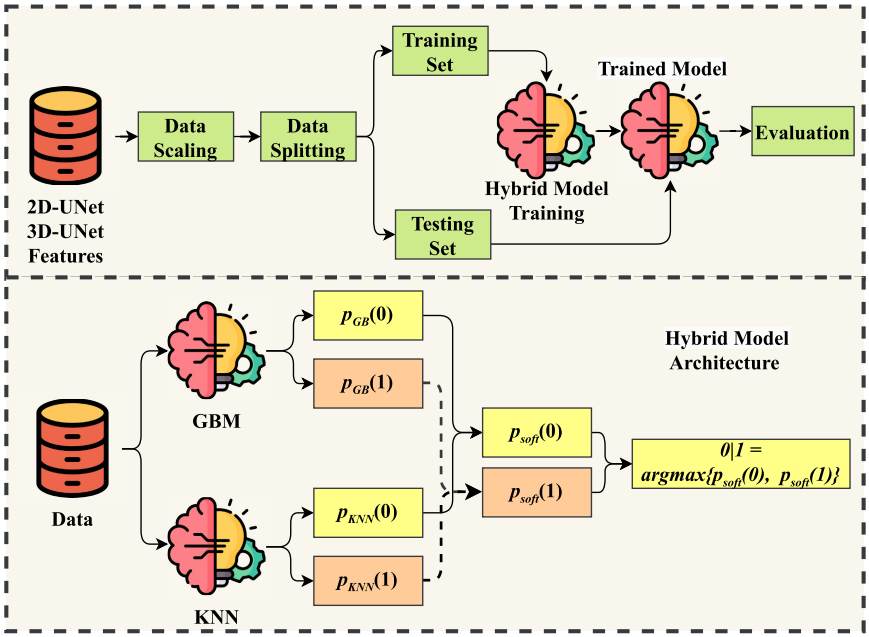

FIGURE 1. Architecture of the proposed methodology.

criteria. By analyzing the individual performance of these models, we discovered that KNN tended to produce more false negative predictions, while GBC gave high false positive predictions and performed well with false positives. Hence, by combining these models, we capitalize on their respective strengths, compensating for their individual weaknesses. The diversity in their predictions further enhances the efficacy of the model under the soft voting criteria, ultimately yielding significant results.

KNN and GBC are distinct models with different approaches. KNN operates on a distance-based principle, making predictions for a test point by measuring its distance to samples in the training data. The Euclidean distance matrix is commonly utilized in KNN as follows

$$EcludianMatrix = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{1}$$

In the proposed approach, we employ KNN with a five-neighbors strategy, where the five nearest neighbors based on the Euclidean distance are selected to make predictions. By utilizing the soft voting criteria, KNN generates prediction probabilities for both Class 0 (representing Not Brain Tumor) and Class 1 (representing Brain Tumor). For instance, if among the k nearest neighbors, 3 belongs to Class 1 and 2 belongs to Class 0, then the prediction probability for Class 1 would be 3/5 = 0.6, and for Class 0, it would be 2/5 = 0.4. This approach allows us to obtain prediction probabilities for each class, providing valuable information for making decisions.

GBM, on the other hand, is an ensemble learning method that sequentially combines weak learners (e.g., decision trees) to build a powerful predictive model through iterative error correction. In this study, we utilized GBC for classification, even though it can also be employed for regression tasks. The key distinction between regression and classification in GBC lies in the choice of the loss function. For the specific brain tumor problem, we employed GBC as a classifier by utilizing the log-likelihood loss function. The log-likelihood loss function is defined as

$$\mathcal{L}(\theta) = -\frac{1}{N}\sum_{i=1}^{N}(y_i \log(p_i) + (1 - y_i)\log(1 - p_i)) \tag{2}$$

where $\mathcal{L}(\theta)$ represents the log-likelihood loss, N is the total number of samples in the dataset, y_i is the true class label for the i-th sample (0 or 1), and p_i is the predicted probability of the positive class for the i-th sample.

By utilizing this log-likelihood loss function, GBC can effectively serve as a classifier, enabling us to tackle the brain tumor classification problem with its iterative boosting approach.

Soft Voting: We combine both models using soft voting which aggregates the probability scores generated by each model for each target class instead of relying solely on their final class predictions. This allows us to leverage the strength of both models and improve the overall classification performance.

Let's explore the concept of soft voting in the context of brain tumor classification. In our case, ''Class 1'' corresponds to the presence of a brain tumor, while ''Class 0'' indicates the absence of a tumor.

- **Probability Scores for GBC:** GBC provides probability scores for both classes (0 and 1) based on the

log-likelihood loss function as follows

$$p_{\mathrm{GB}}(0) = \mathrm{GBM}\{dataset_{\mathrm{Sample}}\}$$
$$p_{\mathrm{GB}}(1) = \mathrm{GBM}\{dataset_{\mathrm{Sample}}\}$$
$$p_{\mathrm{GB}}(0) \quad \text{and} \quad p_{\mathrm{GB}}(1) \tag{3}$$

- **Probability Scores for KNN:** KNN produces probability estimates for the two classes (0 and 1) based on its nearest neighbor search and label assignment

$$p_{\mathrm{KNN}}(0) = \mathrm{KNN}\{dataset_{\mathrm{Sample}}\}$$
$$p_{\mathrm{KNN}}(1) = \mathrm{KNN}\{dataset_{\mathrm{Sample}}\}$$
$$pp_{\mathrm{KNN}}(0) \quad \text{and} \quad p_{\mathrm{KNN}}(1) \tag{4}$$

- **Soft Voting Combination:** Soft voting combines the probability scores from both models (GBM and KNN) to form the ensemble probability scores for each class. This is done by taking the average of the probabilities as shown below

$$p_{\mathrm{soft}}(0) = \frac{p_{\mathrm{GB}}(0) + p_{\mathrm{KNN}}(0)}{2} \tag{5}$$
$$p_{\mathrm{soft}}(1) = \frac{p_{\mathrm{GB}}(1) + p_{\mathrm{KNN}}(1)}{2} \tag{6}$$

- **Final Classification Decision:** The final prediction is made by assigning the class with the highest combined probability score (p_{soft}). For example:
  - f $p_{\mathrm{soft}}(1) > p_{\mathrm{soft}}(0)$, the final prediction is "Class 1", indicating the presence of a brain tumor.
  - Otherwise, the final prediction is "Class 0", indicating the absence of a tumor.

We adopt this hybrid approach as it yields significant results for brain tumor classification, leveraging the 3D-UNet segmentation features.

The reason for the selection of KNN and GBC in the proposed HM is the individual performance of these models and the diversity in their predictions. As mentioned above, we discovered that both KNN and GBC showed strong performance, especially because they exhibited different strengths when we combined them. For instance, for those samples where KNN give correct prediction with high probabilities, GBC made incorrect predictions, and vice versa. We illustrate it through a scenario where samples from the dataset used in this study are taken.

1) Sample 1: Actual class ⟶ Class 0
   - Class 0 = GBC probability ⟶ 0.9
   - Class 0 = KNN probability ⟶ 0.2
   - Class 1 = GBC probability ⟶ 0.1
   - Class 1 = KNN probability ⟶ 0.4
   - Predicted probability:
   - Class 0 ⟶ (GBC probability + KNN probability)/ 2 = (0.9+0.2)/2= 0.55
   - Class 1 ⟶ (GBC probability + KNN probability)/ 2 = (0.1+0.4)/2= 0.25
   - Prediction: Class 0 ⟵ argmax(0.55,0.25)
2) Sample 2: Actual class ⟶ Class 1
   - Class 0 = GBC probability ⟶ 0.3
   - Class 0 = KNN probability ⟶ 0.8
   - Class 1 = GBC probability ⟶ 0.4
   - Class 1 = KNN probability ⟶ 0.9
   - Predicted probability:
   - Class 0 ⟶ (GBC probability + KNN probability) /2 = (0.3+0.8)/2= 0.55
   - Class 1 ⟶ (GBC probability + KNN probability) /2 = (0.4+0.9)/2= 0.65
   - Prediction: Class 1 ⟶ argmax(0.55,0.65)
   - Prediction: Class 0 ⟵ argmax(0.55,0.25)

In the above example, for sample 1 GBC was good and KNN was poor so GBC helped to achieve the correct prediction while for sample 2, KNN showed good performance while GBC was poor. Combining these two models helped to achieve better performance.

B. DATASET

In this study, we obtained the brain tumor detection dataset from the renowned dataset repository, Kaggle. The dataset is "RSNA-MICCAI Brain Tumor Classification" [30] which consists of two types of feature files such as 3D-UNet segmentation features and 2D-UNet segmentation features. We take these extracted features from Kaggle and use them in our study. We describe 2D and 3D UNet used by [30] below:

- **2D U-Net Model for Segmentation**: The 2D U-Net model is designed for image segmentation tasks, where it classifies pixels in a 2D input image into predefined categories [32]. It features an encoder-decoder architecture with skip connections. The encoder downsamples the image through convolutional and pooling layers, while the decoder upscales the features with up-convolutional layers, maintaining contextual information through skip connections, making it suitable for 2D image segmentation tasks.
- **3D U-Net Model for Segmentation**: In contrast, the 3D U-Net extends the 2D model to process volumetric (3D) images, such as brain MRI scans [33]. It retains the U-shaped architecture with skip connections but operates in three dimensions. Convolutional layers in the 3D U-Net handle the depth, width, and height of the input volume. This model is ideal for comprehensive segmentation of medical images, capturing intricate 3D structures and details, thus enabling precise analysis and diagnosis in volumetric data.

The 2D-UNet segmentation generates 54 features using MRI images, while the 3D-UNet segmentation generates 111 features. 3D-UNet segmentation provides more features compared to the 2D-UNet segmentation data because it is an extended version of 2D-UNet. Segmentation is valuable as it captures essential information related to the shape, texture, and grayscale statistics of the brain MRI images. This information plays a crucial role in the accurate classification of medical images. While 3D segmentation is more significant

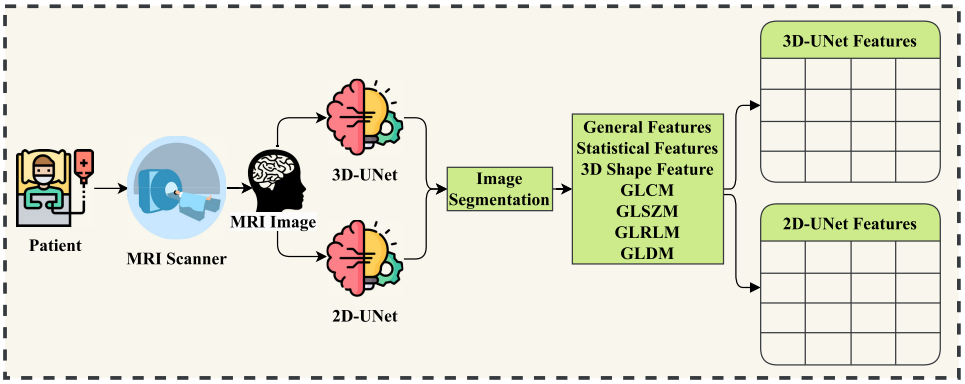

FIGURE 2. Dataset Building Approach.

TABLE 2. Sample dataset obtained from 3D-UNet Feature Extraction.

MGMT_value	MeshVolume	VoxelVolume	SurfaceArea	...	Sphericity
1	1912.083	1971	4067.785	...	0.183146
1	1337.583	1394	2907.736	...	0.201903
0	410	590	1391.187	...	0.191847

TABLE 3. Sample dataset obtained from 2D-UNet Feature Extraction.

MGMT_value	MeshSurface	PixelSurface	Perimeter	...	PerimeterSurfaceRatio
1	1970	1971	275.664	...	0.139931
1	1393	1394	263.1787	...	0.188929
0	551	590	733.6367	...	1.331464

as it can extract various features from MRI images including shape (3D), first-order statistics features, gray level size zone matrix (GLSZM), gray level dependence matrix (GLDM), gray level co-occurrence matrix (GLCM), and gray level run length matrix (GLRLM) features [30]. These diverse features contribute to a comprehensive understanding of the MRI data and enhance the accuracy of the classification process. Figure 2 shows how the dataset is built using both 3D-UNet and 2D-UNet features for this study.

In this study, we used both 2D-UNET segmentation features and 3D-UNET segmentation features for experiments using machine learning deep learning algorithms. The sample of the dataset from 3D-UNet and 2D-UNet segmentation is shown in Tables 2 and 3, respectively. The dataset consists of 585 samples 307 belong to class 1 while 278 belong to class 0.

TABLE 4. Turning range of hyperparameters for machine learning models.

Model	Tuning Range
RF	n_estimators={50 to 200}, max_depth= {2 to 8}, random_state=42
SVM	kernel={'linear','rbf', 'sigmoid'}, C={1.0 to 5.0}, gamma='scale', random_state=42
LR	C={1.0 to 5.0}, random_state=42
KNN	n_neighbors={1 to 10}
DT	max_depth={2 to 8}, random_state=42
NB	Default settings
MLP	hidden_layer_sizes={50 to 200}, max_iter={500 to 2000}, random_state=42
GBM	n_estimators={50 to 200}, learning_rate={0.01 to 1}, max_depth={2 to 8}, random_state=42
XGB	n_estimators={50 to 200}, max_depth={2 to 8}, learning_rate={0.01 to 1}, random_state=42
LGBM	n_estimators={50 to 200}, max_depth={2 to 8}, learning_rate={0.01 to 1}, random_state=42

C. MACHINE LEARNING MODELS

Machine learning (ML) models have applications across various domains including text analysis, IoT, computer vision, image processing, security, and medical analysis. In this study, these models are employed for brain tumor classification.

Specifically, the study employed ten ML models in comparison with the proposed model, each applied with their best hyper-parameter settings to ensure optimal performance. This hyperparameter setting we achieved by tuning the model between a specific range of values as shown in Table 4.

The description of ML models and hyperparameters setting is given in Table 5. Logistic Regression (LR), Decision Trees (DT), Random Forests (RF), KNN, Extreme gradient boosting (XGB), Light GBM (LGBM), Naive Bayes (NB), Support Vector Machines (SVM), Multi-Layer Perceptron (MLP) and GBC are used in this study. For the implementation of these models, the Scikit-learn library we utilize.

D. EVALUATION PARAMETERS

In this study, we use several evaluation metrics to assess the performance of the classifiers. These metrics include

TABLE 5. Machine Learning Models' Description and Hyperparameters Setting.

Model	Description	Hyperparamters
RF	RF is an ensemble learning technique that builds a number of decision trees and combines their outputs through voting, resulting in a reliable and accurate classification model. RF's capability to handle high-dimensional numeric features and its resistance to overfitting make it well-suited for medical data analysis tasks, such as brain tumor classification. Its ability to handle both categorical and continuous data allows for straightforward integration of different types of features commonly extracted from brain MRI scans [34].	n_estimators=100, max_depth=5, random_state=42
SVM	SVM can classify tumor and non-tumor instances in high-dimensional numeric feature space using an optimal hyperplane, providing effective and interpretable results for medical data analysis. SVM's ability to handle complex relationships in high-dimensional data, like those extracted from brain MRI scans, makes it a valuable choice for brain tumor classification, particularly when dealing with smaller datasets or when data is linearly separable. [35].	kernel='rbf', C=1.0, gamma='scale', random_state=42
LR	LR is a linear classifier used for binary classification tasks such as tumor and non-tumor classification, predicting the probability of an instance belonging to a particular class based on input features. It uses the logistic function (sigmoid) to map predicted values to probabilities, with values above 0.5 classified as tumor and below 0.5 as non-tumor. [36].	C=1.0, random_state=42
KNN	KNN is a non-parametric algorithm that relies on the distances between data points to make predictions. It classifies an instance based on the majority class of its k-nearest neighbors, where k is a user-specified parameter controlling the number of neighbors considered. KNN is straightforward to implement, and suitable for small to medium-sized datasets, but may struggle with high-dimensional data and could require proper feature scaling and optimal k-value selection. [37].	n_neighbors=5
DT	DT is a data-driven algorithm that constructs a tree structure based on relevant features extracted from brain MRI images, enabling accurate tumor categorization by recursively partitioning the data into distinct classes. Its interpretable nature allows clinicians to understand the decision-making process and potentially identify crucial imaging biomarkers for diagnosis [38].	max_depth=5, random_state=42
NB	NB is a straightforward probabilistic classifier for brain tumor classification, using the assumption of feature independence to efficiently analyze brain MRI data and make predictions based on likelihoods. While NB is computationally efficient and useful with limited data, its naive assumption may lead to insignificant results when dealing with highly correlated features or intricate tumor patterns. [39].	Default settings
MLP	MLP, short for Multi-Layer Perceptron, is a suitable choice for brain tumor classification, as it can effectively handle numerical data and learn complex patterns from the numeric features extracted from brain MRI images, offering accurate predictions based on these MRI features. [40].	hidden_layer_sizes=100, max_iter=1000, random_state=42
GBC	It is a technique in ensemble learning that combines several weak learners to create a strong one by optimizing the loss function, typically employing the exponential loss function or deviance [41].	n_estimators=100, learning_rate=0.1, max_depth=3, random_state=42
XGB	XGB is an advanced ensemble learning algorithm that combines multiple weak learners (decision trees) to form a strong predictive model, effectively capturing complex relationships in numeric data from brain MRI scans. XGB employs gradient boosting techniques to iteratively correct errors and optimize the model's performance, making it particularly powerful for handling high-dimensional datasets with numeric features and achieving state-of-the-art results in brain tumor classification tasks. [42].	n_estimators=100, max_depth=3, learning_rate=0.1, random_state=42
LGBM	The LGBM (Light Gradient Boosting Machine) is a gradient boosting framework that utilizes a histogram-based approach to construct decision trees, leading to faster training and higher efficiency for numeric data extracted from brain MRI scans. LGBM is significant in handling large-scale datasets and high-dimensional features, making it a suitable choice for medical data analysis tasks like brain tumor classification, where interpretable and accurate results are required [43].	n_estimators=100, max_depth=3, learning_rate=0.1, random_state=42

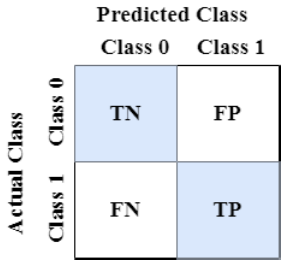

FIGURE 3. Confusion matrix representation.

precision, accuracy, recall, and F1-score, which are calculated based on the True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). These parameters are shown in Figure 3

- Accuracy: Overall correctness of the model is known as accuracy and it can be calculated as the total number of correctly classified samples divided by the total number of predictions.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

- Precision: It evaluates the proportion of true positive predictions among all the predicted positive samples.

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

- Recall: It is also known as sensitivity or true positive rate and it measures the ability of a classifier to correctly identify positive samples from the actual positive class. It is calculated as follows

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

- F1 score: It is a harmonic mean of precision and recall and is particularly useful when dealing with imbalanced

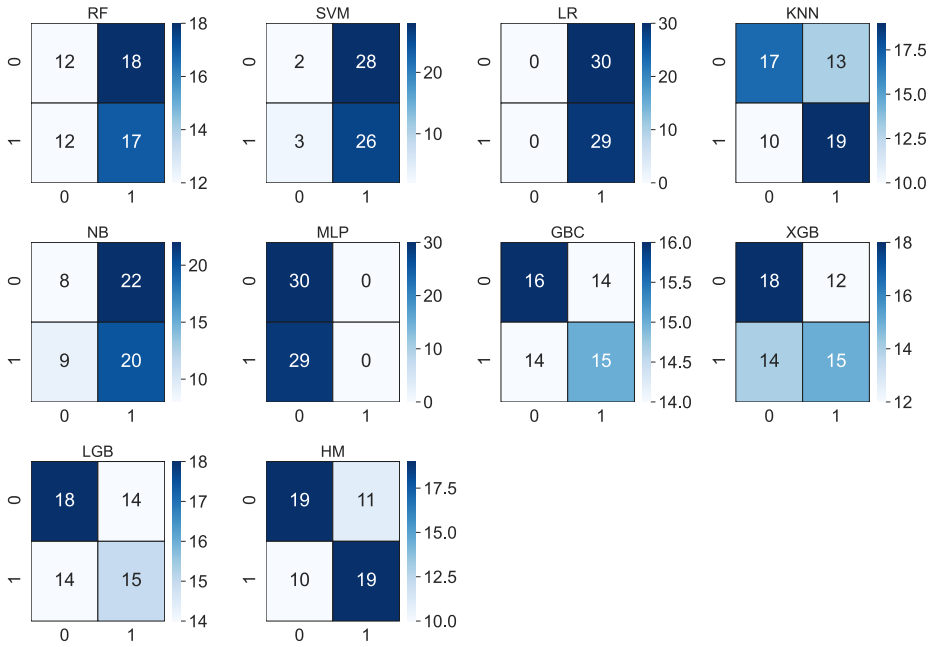

FIGURE 4. Confusion matrix using 2D-UNet segmentation features.

data. It balances the trade-off between precision and recall and is calculated as follows

$$F1\ score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{10}$$

IV. RESULTS AND DISCUSSION

In this section, we present the outcomes achieved by machine learning and deep learning models for brain tumor prediction. We apply these models to both 2D-UNet segmented features and 3D-UNet segmented features. The performance evaluation of all models encompasses overall accuracy, precision, recall, F1 score, per-class results, and confusion matrices.

A. MACHINE LEARNING RESULTS USING 2D-UNET SEGMENTATION FEATURES

In this section, we present the results of multiple machine learning classifiers applied to 2D-UNet segmentation features for brain tumor classification. The classifiers exhibit varying levels of performance, as shown in Table 6. The proposed HM achieves the highest accuracy score of 0.64, combining GBC and KNN for the most accurate predictions, while SVM, LR, and NB show the lowest accuracy of 0.47.

In terms of precision, HM achieves the highest score of 0.64, indicating that it makes the most accurate predictions for class 0. In contrast, SVM and LR have the lowest precision of 0.44 and 0.25, respectively, resulting in a higher number of false positives. In terms of recall, which measures the ability to correctly identify class 0 instances, HM exhibits the highest recall of 0.64, implying its effectiveness in detecting positive cases. On the other hand, SVM has the lowest recall of 0.48, showing a higher number of wrong predictions for class 1. The F1 score, which considers both precision and recall, reveals that HM achieves the highest value of 0.64, signifying its strong overall performance in achieving a balanced score. In contrast, SVM has the lowest F1 score of 0.37, indicating a lower balance between precision and recall scores.

TABLE 6. Performance of ML models for 2D-UNet segmentation features.

Classifier	Accuracy	Precision	Recall	F1 score
RF	0.49	0.49	0.49	0.49
SVM	0.47	0.44	0.48	0.37
LR	0.49	0.25	0.50	0.33
KNN	0.61	0.61	0.61	0.61
DT	0.51	0.51	0.51	0.49
NB	0.47	0.47	0.48	0.45
MLP	0.51	0.25	0.50	0.34
GBM	0.53	0.53	0.53	0.53
XGB	0.56	0.56	0.56	0.56
LGBM	0.54	0.54	0.54	0.54
HM(GBM+KNN)	0.64	0.64	0.64	0.64

According to the confusion matrix shown in Figure 4, the HM achieves the highest ratio of correct predictions, with 38 correct predictions out of a total of 59 test predictions. On the other hand, both SVM and NB show poorer results, with only a total of 31 correct predictions out of 59 test predictions. Table 7, shows the per-class accuracy for class 0 and class 1 using each machine learning model.

B. RESULTS USING 3D-UNET SEGMENTATION FEATURES

In this section, we present the results of machine learning classifiers using 3D-UNet segmentation features. From the results in Table 8, it is evident that the classifiers have significant performance as compared to 2D-UNet features.

TABLE 7. Class-wise performance of ML models For 2D-UNet segmentation features.

Class	RF			SVM			LR		
	Precision	Recall	F1 Score	Precision	Recall	F1 Score	Precision	Recall	F1 Score
0	0.50	0.40	0.44	0.40	0.07	0.11	0.00	0.00	0.00
1	0.49	0.59	0.53	0.48	0.90	0.63	0.49	1.00	0.66
	KNN			DT			NB		
0	0.63	0.57	0.60	0.53	0.30	0.38	0.47	0.27	0.34
1	0.59	0.66	0.62	0.50	0.72	0.59	0.48	0.69	0.56
	MLP			GBC			XGB		
0	0.51	1.00	0.67	0.53	0.53	0.53	0.56	0.60	0.58
1	0.00	0.00	0.00	0.52	0.52	0.52	0.56	0.52	0.54
Class	LGBM			HM (GBM+KNN)					
	Precision	Recall	F1 Score	Precision		Recall		F1 Score	
0	0.55	0.60	0.57	0.66		0.63		0.64	
1	0.54	0.48	0.51	0.63		0.66		0.64	

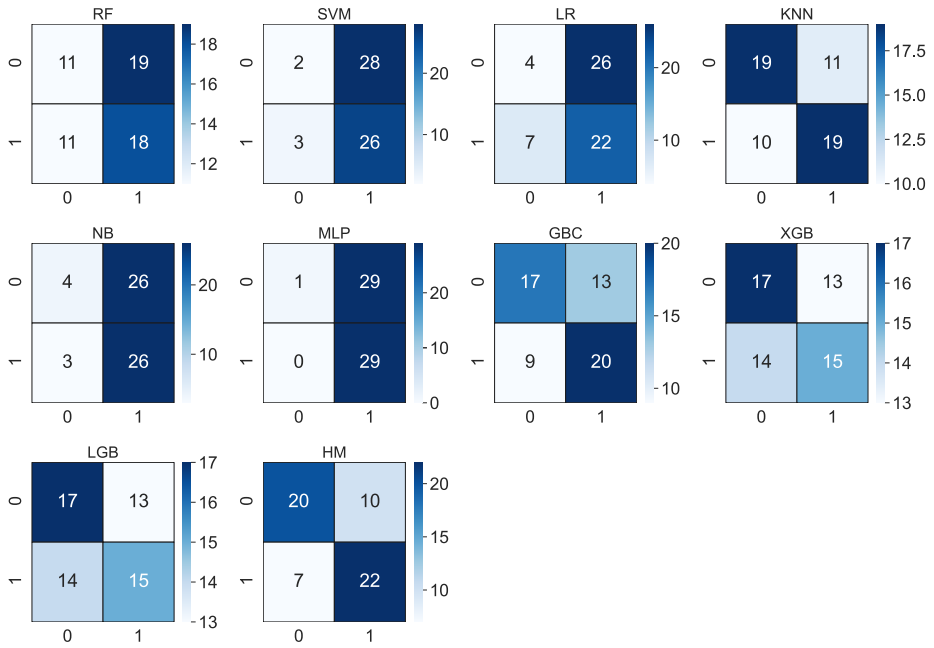

FIGURE 5. Confusion matrix using 3D-UNet segmentation features.

The highest accuracy score of 0.71 is demonstrated by the ensemble model HM. On the other hand, RF and SVM demonstrate the lowest accuracy scores of 0.49 and 0.47, respectively. The highest accuracy with the 2D-UNet feature was 0.64 while with 3D-UNet features, the accuracy score increased to 0.71.

According to the confusion matrix in Figure 5, the HM achieves the highest ratio of correct predictions, with 42 correct predictions out of a total of 59 test predictions. On the other hand, the SVM shows poor results, with only a total of 28 correct predictions out of 59 test predictions. The confusion matrices provide valuable insights for selecting individual models to build the HM. We can observe that certain models like KNN and GBC, exhibit diversity in making correct and wrong predictions. KNN gives 11 wrong predictions for class 0, while GBC gives 13 wrong predictions for the same class. Similarly, KNN has 10 wrong predictions for class 1, whereas GBC has 9 wrong predictions for class 1. This indicates that each model has its strengths and weaknesses.

TABLE 8. Performance of ML models For 3D-UNet segmentation features.

Classifier	Accuracy	Precision	Recall	F1 score
RF	0.49	0.49	0.49	0.49
SVM	0.47	0.44	0.48	0.37
LR	0.44	0.41	0.45	0.38
KNN	0.64	0.64	0.64	0.64
DT	0.51	0.52	0.51	0.47
NB	0.51	0.54	0.51	0.43
MLP	0.51	0.75	0.52	0.37
GBC	0.63	0.63	0.63	0.63
XGB	0.54	0.54	0.54	0.54
LGBM	0.54	0.54	0.54	0.54
VC(GBM+KNN)	0.71	0.71	0.71	0.71

In terms of precision, the HM again attains a significant score of 0.71, indicating that it is most successful in making

TABLE 9. Class-wise performance of ML models For 3D-UNet segmentation features.

Class	RF			SVM			LR		
	Precision	Recall	F1 Score	Precision	Recall	F1 Score	Precision	Recall	F1 Score
0	0.50	0.37	0.42	0.40	0.07	0.11	0.36	0.13	0.20
1	0.49	0.62	0.55	0.48	0.90	0.63	0.46	0.76	0.57
	KNN			DT			NB		
0	0.66	0.63	0.64	0.54	0.23	0.33	0.57	0.13	0.22
1	0.63	0.66	0.64	0.50	0.79	0.61	0.50	0.90	0.64
	MLP			GBC			XGB		
0	1.00	0.03	0.06	0.65	0.57	0.61	0.55	0.57	0.56
1	0.50	1.00	0.67	0.61	0.69	0.65	0.54	0.52	0.53
Class	LGBM			HM (GBM+KNN)					
	Precision	Recall	F1 Score	Precision		Recall		F1 Score	
0	0.55	0.57	0.56	0.74		0.67		0.70	
1	0.54	0.52	0.53	0.69		0.76		0.72	

accurate class 0 predictions. Conversely, MLP has the highest precision of 0.75, while LR and NB have relatively lower precision values. Regarding recall, which measures the classifiers' ability to correctly identify positive instances, the HM exhibits the highest recall of 0.71, implying its effectiveness in detecting positive cases. GBC also shows a high recall of 0.63, while SVM and RF have lower recall values. The F1-score indicates that the HM achieves the highest value of 0.71, signifying its strong overall performance in achieving a balance between precision and recall. The other models demonstrate varying degrees of F1 scores, with MLP showing the lowest F1 score of 0.37 and GBC, XGB, and LGBM having an F1 score of 0.63.

TABLE 10. Models evaluation in terms of ROC AUC using 2D-UNet and 3D-UNet features.

Model	2D-UNet	3D-UNet
RF	0.5046	0.5356
SVM	0.5253	0.5276
LR	0.5402	0.4885
KNN	0.6115	0.6270
DT	0.5862	0.5954
NB	0.5218	0.5575
MLP	0.4667	0.5167
GBM	0.5552	0.7138
XGB	0.5667	0.6230
LGBM	0.5644	0.5701
HM	0.6368	0.7437

By combining these models in the HM, their diverse performance helps to improve overall accuracy. The areas where one model may have weaknesses are compensated by the strengths of the other model. As a result, the hybrid model achieves better accuracy due to the collaborative efforts of its individual components. Table 9, shows the per-class accuracy for class 0 and class 1 using each machine learning model. We also presented model results in terms of ROC AUC as shown in Table 10.

C. DEEP LEARNING MODEL RESULTS USING 3D-UNET SEGMENTATION FEATURES

The performance of machine learning models is significant with 3D-UNet features as compared to 2D-UNet features that's why we only explore 3D-UNet features for further experiments. In this section, four deep learning architectures are applied for brain tumor classification using 3D-UNet segmentation features. We deploy those deep learning models for experiments that have shown promising results in medical data analysis in existing studies. The models utilized in this study include CNN [44], Long Short-Term Memory (LSTM) [45], Gated Recurrent Unit (GRU), and DNN. The architecture of these used models is shown in Table 11.

TABLE 11. Architecture of deep learning models.

Model	Architecture
CNN	Sequential() Conv1D(filters=32, kernel_size=3, activation='relu',input_shape=) Dense(units=1, activation='sigmoid')
LSTM	Sequential() LSTM(units=64, activation='relu', input_shape=) Dense(units=1, activation='sigmoid')
ANN	Sequential() Dense(units=64, activation='relu', input_shape=) Dropout(0.5) Dense(units=1, activation='sigmoid')
GRU	Sequential() GRU(units=64, activation='relu', input_shape=) Dense(units=1, activation='sigmoid')
compile(loss='binary_crossentropy', optimizer='adam', metrics=['accuracy']) fit(X_train_scaled, y_train, epochs=10, batch_size=32)	

Table 12 illustrates the performance of four deep learning models using 3D-UNet segmentation features for brain tumor classification. CNN achieved an accuracy score of 0.49 with moderate recall but low precision, leading to a poor F1 score. Similarly, LSTM achieved an accuracy score of 0.47 with an imbalanced recall and precision, resulting in a lower F1 score. GRU attained an accuracy of 0.49 with balanced precision and recall, still a lower F1 score. DNN performed poorly with an accuracy of 0.39 and balanced precision and recall. The models show mixed results, indicating the potential for improvement in accuracy and F1-score.

Deep learning models, especially those with a large number of parameters, often require a substantial amount of data to generalize well and achieve high performance. When the dataset is small, deep learning models may suffer from overfitting, meaning they memorize the training data

TABLE 12. Performance of deep learning models using 3D-UNet segmentation features.

Classifier	Accuracy	Precision	Recall	F1 Score
CNN	0.49	0.25	0.50	0.33
LSTM	0.47	0.41	0.48	0.35
GRU	0.49	0.50	0.50	0.36
DNN	0.39	0.39	0.39	0.39

and fail to generalize to new, unseen data as in this case. In contrast, traditional machine learning models, like RF, NB, or KNN, can perform well with smaller datasets. These models are generally less prone to overfitting and may be more robust in cases of limited data. They are also less computationally demanding and can be trained faster on small datasets. Table 13 shows the per-class accuracy for class 0 and class 1 using each deep learning model.

TABLE 13. Class-wise performance of deep learning models using 3D-UNet segmentation features.

CNN			
Class	Precision	Recall	F1 Score
0	0.00	0.00	0.00
1	0.49	1.00	0.66
LSTM			
0	0.33	0.03	0.06
1	0.48	0.93	0.64
GRU			
0	0.50	0.03	0.06
1	0.49	0.97	0.65
DNN			
0	0.38	0.30	0.33
1	0.40	0.48	0.44

D. PERFORMANCE COMPARISON WITH PREVIOUS STATE-OF-THE-ART APPROACHES

To demonstrate the significance of the proposed approach, we conducted a comparison with other studies that have previously worked on the same dataset. The accuracy achieved in those previous studies exhibited a notable gap, suggesting that there is considerable room for improvement in terms of accuracy. Most of the previous studies focused on directly using the RSNA-MICCAI image dataset and employed transfer learning techniques. In contrast, this study utilized a 3D-UNet extracted features from the RSNA-MICCAI image dataset. Leveraging this feature set, the proposed hybrid models achieved highly significant results, as shown in Table 14. This demonstrates the effectiveness of the proposed approach in surpassing the performance of previous studies and highlights the potential for further advancements in this field.

Additionally, we also carried out k-fold cross-validation to validate the model. A 10-fold cross-validation is used to validate the significance of the proposed approach. Validation results indicate that the proposed model achieved a mean accuracy of 0.655 and a standard deviation of 0.03, as shown in Table 15.

TABLE 14. Comparison with recent studies.

Ref.	Year	Model	Accuracy
[27]	2022	ResNet10	58.0%
[26]	2022	ResNet34	53.1%
		EfficientNet	54.8%
		3D CNN + ResNet10 (SOTA)	60.7%
		multi-modality + attention	56.7%
		multi-modality + attention	59.0%
		multi-modality + attention	61.1%
		multi-modality + attention + shortcut	63.71%
[46]	2022	T1w	70.0%
This study	2022	HM+3D-UNet Segmentation Features	71.1%

TABLE 15. Results of k-fold cross-validation.

Fold	Accuracy
Fold 1	0.6440678
Fold 2	0.6101694
Fold 3	0.6101694
Fold 4	0.7118644
Fold 5	0.6949152
Fold 6	0.6949152
Fold 7	0.5932203
Fold 8	0.6440678
Fold 9	0.6779661
Fold 10	0.6779661
Mean Accuracy	0.65593220
Standard Deviation	0.03942272

E. DISCUSSION

In this study, we proposed an approach for brain tumor prediction by deploying various machine learning and deep learning models to predict brain tumor instances using 2D-UNet and 3D-UNet segmented features. The analysis of the machine learning models using 2D-UNet features revealed diverse performance levels, with the proposed model combining GBC and KNN demonstrating the highest accuracy and F1 score. However, SVM, LR, and NB exhibited relatively lower accuracy and precision, indicating room for improvement. The incorporation of 3D-UNet features resulted in significant enhancements in performance, with the proposed model achieving the highest accuracy of 0.71, surpassing the best accuracy achieved with 2D-UNet features, as shown in Figure 6. Deep learning models, however, exhibited poor performance, because of smaller dataset size and feature set for good generalization. Overall, the study highlights the importance of feature extraction techniques and the potential benefits of combining multiple models for improved brain tumor classification accuracy.

F. STATISTICAL SIGNIFICANCE TEST

To show the significance of the proposed approach, we conducted a statistical T-test for the proposed approach. This test allowed us to compare the results from the new approach with those from other methods. The T-test works by creating two hypotheses the 'null hypothesis' and the 'alternative hypothesis.' If the P-value is smaller than the T-stat value, it indicates that the T-test rejects the null hypothesis and supports the alternative hypothesis. In simpler terms, this indicates that there is a significant statistical difference

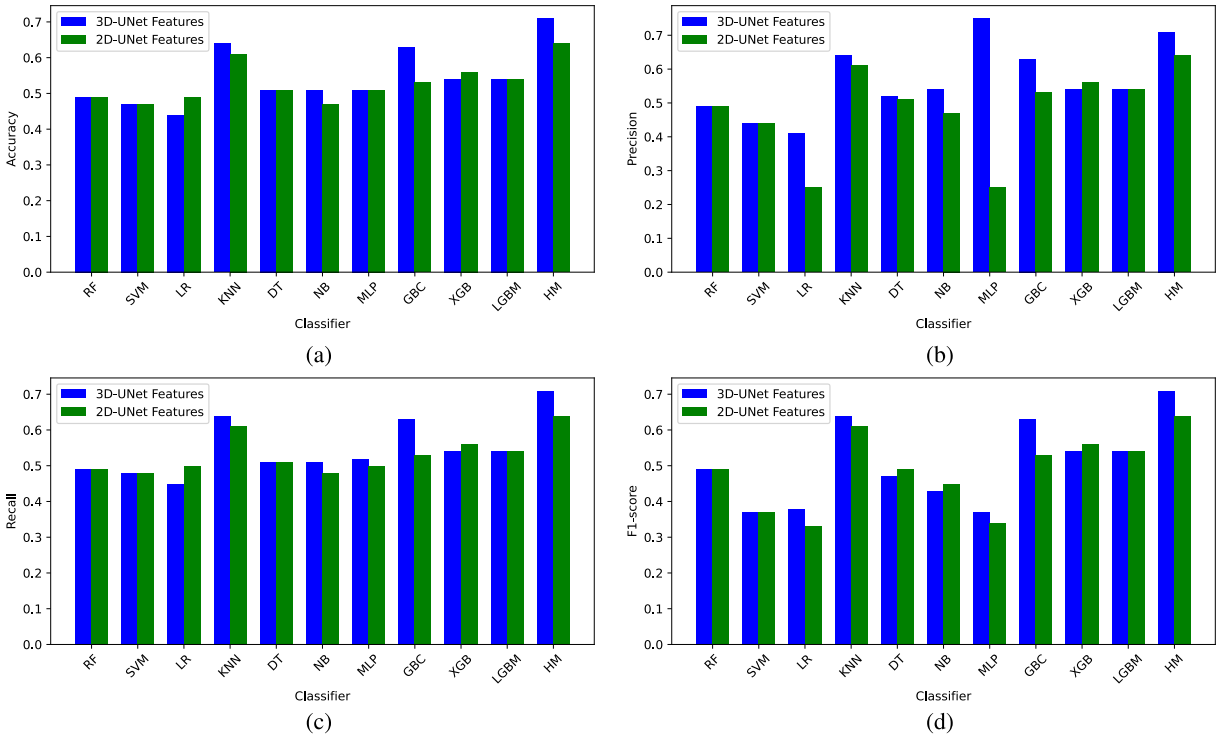

FIGURE 6. Performance scores comparison (a) Accuracy, (b) Precision, (c) Recall, and (d) F1 score.

TABLE 16. Results of T-test.

Model	T-Stat	P-Value	Statistically Significant
HM vs RF	inf	0	Yes
HM vs SVM	10.8328	3.66424e-05	Yes
HM vs LR	18.3412	1.69269e-06	Yes
HM vs KNN	∞	0	Yes
HM vs DT	18.716	1.50197e-06	Yes
HM vs NB	9.00998	0.000104616	Yes
HM vs MLP	2.19259	0.0708211	No
HM vs GBC	∞	0	Yes
HM vs XGB	∞	0	Yes
HM vs LGBM	∞	0	Yes

between the compared results. On the other hand, if the T-test accepts the null hypothesis, it suggests that there is no significant statistical difference between the results.

Table 16 shows the results of the T-test. Results indicate that the proposed HM rejects the null hypothesis for every instance, indicating that the performance of the proposed model has a strong statistical significance over other approaches.

V. CONCLUSION

The brain tumor is a life-threatening disease, with a slim probability of survival among all tumors. Automated diagnosis using MRI images can help increase diagnosis accuracy and reduce time. In this study, we investigate the application of machine learning techniques for brain tumor detection using a hybrid model on 3D-UNet MRI segmentation features. The use of GBC and KNN through soft voting criteria in the proposed model proved to be advantageous, capitalizing on the strengths and weaknesses of each method. The overall accuracy of 71% for brain tumor detection shows the effectiveness of the proposed approach where existing models show poor results on these features. Experimental findings reveal that 3D-UNet segmentation features hold greater significance compared to 2D-UNet features. The 3D-UNet generates a larger feature set, facilitating more efficient learning for the models. Moreover, the combination of KNN and GBC in an ensemble learning framework was crucial as the diversity in their individual performances contributed to the overall significant performance. While KNN shows poor performance for certain data points, GBC shows significant improvements, and vice versa.

A. LIMITATIONS & FUTURE WORKS

However, the study did identify some limitations that should be addressed in future work. Firstly, the available feature set size and the number of samples are insufficient to train deep learning models effectively. A larger dataset is required to obtain optimal parameter weights for improved performance. Additionally, though the achieved accuracy was significantly better than existing approaches, the medical application demands even higher precision. Hence, efforts to enhance accuracy should be a focus of future research. Furthermore, the study's work on a relatively small dataset may have impacted the models' performance, so we will increase the size of the data in future work. Incorporating more data into the existing dataset can also lead to enhanced model performance.

ACKNOWLEDGMENT

The authors extend their appreciation to King Saud University for funding this research through Researchers Supporting Project Number (RSPD2023R890), King Saud University, Riyadh, Saudi Arabia.

REFERENCES

[1] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA, Cancer J. Clinicians*, vol. 71, no. 3, pp. 209–249, May 2021.

[2] Q. T. Ostrom, N. Patil, G. Cioffi, K. Waite, C. Kruchko, and J. S. Barnholtz-Sloan, "Cbtrus statistical report: Primary brain and other central nervous system tumors diagnosed in the United States in 2013–2017," *Neuro-Oncol.*, vol. 22, no. 1, pp. iv1–iv96, 2020.

[3] K. Chang, B. Zhang, X. Guo, M. Zong, R. Rahman, D. Sanchez, N. Winder, D. A. Reardon, B. Zhao, P. Y. Wen, and R. Y. Huang, "Multimodal imaging patterns predict survival in recurrent glioblastoma patients treated with bevacizumab," *Neuro-Oncol.*, vol. 18, no. 12, pp. 1680–1687, Dec. 2016.

[4] A. Mahajan, A. V. Moiyadi, R. Jalali, and E. Sridhar, "Radiogenomics of glioblastoma: A window into its imaging and molecular variability," *Cancer Imag.*, vol. 15, no. S1, p. 1, Dec. 2015.

[5] R. G. W. Verhaak, "Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1," *Cancer Cell*, vol. 17, no. 1, pp. 98–110, Jan. 2010.

[6] E. Lee, R. L. Yong, P. Paddison, and J. Zhu, "Comparison of glioblastoma (GBM) molecular classification methods," *Seminars Cancer Biol.*, vol. 53, pp. 201–211, Dec. 2018.

[7] H. Akbari, L. Macyszyn, X. Da, R. L. Wolf, M. Bilello, R. Verma, D. M. O'Rourke, and C. Davatzikos, "Pattern analysis of dynamic susceptibility contrast-enhanced MR imaging demonstrates peritumoral tissue heterogeneity," *Radiology*, vol. 273, no. 2, pp. 502–510, Nov. 2014.

[8] M. Law, R. J. Young, J. S. Babb, N. Peccerelli, S. Chheang, M. L. Gruber, D. C. Miller, J. G. Golfinos, D. Zagzag, and G. Johnson, "Gliomas: Predicting time to progression or survival with cerebral blood volume measurements at dynamic susceptibility-weighted contrast-enhanced perfusion MR imaging," *Radiology*, vol. 247, no. 2, pp. 490–498, May 2008.

[9] N. Alturki, M. Umer, A. Ishaq, N. Abuzinadah, K. Alnowaiser, A. Mohamed, O. Saidani, and I. Ashraf, "Combining CNN features with voting classifiers for optimizing performance of brain tumor classification," *Cancers*, vol. 15, no. 6, p. 1767, Mar. 2023.

[10] M. Diehn, C. Nardini, D. S. Wang, S. McGovern, M. Jayaraman, Y. Liang, K. Aldape, S. Cha, and M. D. Kuo, "Identification of noninvasive imaging surrogates for brain tumor gene-expression modules," *Proc. Nat. Acad. Sci. USA*, vol. 105, no. 13, pp. 5213–5218, Apr. 2008.

[11] Y. Li, J. M. Lupo, M.-Y. Polley, J. C. Crane, W. Bian, S. Cha, S. Chang, and S. J. Nelson, "Serial analysis of imaging parameters in patients with newly diagnosed glioblastoma multiforme," *Neuro-Oncol.*, vol. 13, no. 5, pp. 546–557, May 2011.

[12] A. M. Rutman and M. D. Kuo, "Radiogenomics: Creating a link between molecular diagnostics and diagnostic imaging," *Eur. J. Radiol.*, vol. 70, no. 2, pp. 232–241, May 2009.

[13] P. O. Zinn, B. Majadan, P. Sathyan, S. K. Singh, S. Majumder, F. A. Jolesz, and R. R. Colen, "Radiogenomic mapping of edema/cellular invasion MRI-phenotypes in glioblastoma multiforme," *PLoS ONE*, vol. 6, no. 10, Oct. 2011, Art. no. e25451.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[15] F. H. Shajin, S. P, P. Rajesh, and V. K. Nagoji Rao, "Efficient framework for brain tumour classification using hierarchical deep learning neural network classifier," *Comput. Methods Biomechanics Biomed. Eng., Imag. Visualizat.*, vol. 11, no. 3, pp. 750–757, May 2023.

[16] N. E. Varghese, A. John, and U. D. A. C, "Classification of glioma by exploring wavelet-based radiomic features and machine learning techniques using brats dataset," in *Proc. 3rd Int. Conf. Adv. Electr., Comput., Commun. Sustain. Technol. (ICAECT)*, Jan. 2023, pp. 1–7.

[17] A. Myronenko, "3D MRI brain tumor segmentation using autoencoder regularization," in *Proc. MICCAI*, Sep. 2018, pp. 311–320.

[18] K.-L. Tseng, Y.-L. Lin, W. Hsu, and C.-Y. Huang, "Joint sequence learning and cross-modality convolution for 3D biomedical segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3739–3746.

[19] Y. Shachor, H. Greenspan, and J. Goldberger, "A mixture of views network with applications to multi-view medical imaging," *Neurocomputing*, vol. 374, pp. 1–9, Jan. 2020.

[20] D. Nie, L. Wang, Y. Gao, and D. Shen, "Fully convolutional networks for multi-modality isointense infant brain image segmentation," in *Proc. IEEE 13th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2016, pp. 1342–1345.

[21] K. Kamnitsas, C. Ledig, V. F. J. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Med. Image Anal.*, vol. 36, pp. 61–78, Feb. 2017.

[22] Q. Xu, Z. Ma, N. He, and W. Duan, "DCSAU-Net: A deeper and more compact split-attention U-net for medical image segmentation," *Comput. Biol. Med.*, vol. 154, Mar. 2023, Art. no. 106626.

[23] S. Vijay, T. Guhan, K. Srinivasan, P. M. D. R. Vincent, and C.-Y. Chang, "MRI brain tumor segmentation using residual spatial pyramid pooling-powered 3D U-Net," *Frontiers Public Health*, vol. 11, Feb. 2023, Art. no. 1091850.

[24] Q. Jia and H. Shu, "BiTr-UNet: A CNN-transformer combined network for MRI brain tumor segmentation," in *Proc. Int. MICCAI Brainlesion Workshop*. Cham, Switzerland: Springer, 2021, pp. 3–14.

[25] M. Elmezain, A. Mahmoud, D. T. Mosa, and W. Said, "Brain tumor segmentation using deep capsule network and latent-dynamic conditional random fields," *J. Imag.*, vol. 8, no. 7, p. 190, Jul. 2022.

[26] R. Qu and Z. Xiao, "An attentive multi-modal CNN for brain tumor radiogenomic classification," *Information*, vol. 13, no. 3, p. 124, Mar. 2022.

[27] N. Saeed, S. Hardan, K. Abutalip, and M. Yaqub, "Is it possible to predict MGMT promoter methylation from brain tumor MRI scans using deep learning models?" in *Proc. Int. Conf. Med. Imag. With Deep Learn.*, 2022, pp. 1005–1018.

[28] H. A. Shah, F. Saeed, S. Yun, J.-H. Park, A. Paul, and J.-M. Kang, "A robust approach for brain tumor detection in magnetic resonance images using finetuned EfficientNet," *IEEE Access*, vol. 10, pp. 65426–65438, 2022.

[29] F. J. Díaz-Pernas, M. Martínez-Zarzuela, M. Antón-Rodríguez, and D. González-Ortega, "A deep learning approach for brain tumor classification and segmentation using a multiscale convolutional neural network," *Healthcare*, vol. 9, no. 2, p. 153, Feb. 2021.

[30] (2021)*RSNA-MICCAI Brain Tumor Classification*. Accessed: Jul. 22, 2023. [Online]. Available: https://www.kaggle.com/code/yannicksteph/rsna-miccai-brain-tumor-classification

[31] O. Kramer and O. Kramer, "Scikit-learn," in *Machine Learning for Evolution Strategies*. New York, NY, USA: Springer, 2016, pp. 45–53.

[32] S. Rajagopal, T. Thanarajan, Y. Alotaibi, and S. Alghamdi, "Brain tumor: Hybrid feature extraction based on UNet and 3DCNN," *Comput. Syst. Sci. Eng.*, vol. 45, no. 2, pp. 2093–2109, 2023.

[33] J. Marti Asenjo and A. Martinez-Larraz Solís, "MRI brain tumor segmentation using a 2D-3D U-Net ensemble," in *Proc. Int. MICCAI Brainlesion Workshop*, Lima, Peru, Oct. 2020, pp. 354–366.

[34] M. Manzoor, M. Umer, S. Sadiq, A. Ishaq, S. Ullah, H. A. Madni, and C. Bisogni, "RFCNN: Traffic accident severity prediction based on decision level fusion of machine and deep learning model," *IEEE Access*, vol. 9, pp. 128359–128371, 2021.

[35] S. Huang, N. Cai, P. P. Pacheco, S. Narrandes, Y. Wang, and W. Xu, "Applications of support vector machine (SVM) learning in cancer genomics," *Cancer Genomics Proteomics*, vol. 15, no. 1, pp. 41–51, 2018.

[36] S. Menard, *Applied Logistic Regression Analysis*. Newbury Park, CA, USA: Sage, 2002, no. 106.

[37] M.-L. Zhang and Z.-H. Zhou, "ML-KNN: A lazy learning approach to multi-label learning," *Pattern Recognit.*, vol. 40, no. 7, pp. 2038–2048, Jul. 2007.

[38] Y. Y. Song and Y. Lu, "Decision tree methods: Applications for classification and prediction," *Shanghai Arch. Psychiatry*, vol. 27, no. 2, p. 130, Apr. 2015.

[39] K. M. Leung, "Naive Bayesian classifier," *Polytech. Univ. Dept. Comput. Sci./Finance Risk Eng.*, vol. 2007, pp. 123–156, Nov. 2007.

[40] H. Taud and J. Mas, "Multilayer perceptron (MLP)," in *Geomatic Approaches for Modeling Land Change Scenarios*. Cham, Switzerland: Springer, 2018, pp. 451–455.

[41] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, "A comparative analysis of gradient boosting algorithms," *Artif. Intell. Rev.*, vol. 54, no. 3, pp. 1937–1967, Mar. 2021.

[42] Q. Zhang, P. Liu, X. Wang, Y. Zhang, Y. Han, and B. Yu, "StackPDB: Predicting DNA-binding proteins based on XGB-RFE feature optimization and stacked ensemble classifier," *Appl. Soft Comput.*, vol. 99, Feb. 2021, Art. no. 106921.
[43] R. M. Aziz, M. F. Baluch, S. Patel, and A. H. Ganie, "LGBM: A machine learning approach for Ethereum fraud detection," *Int. J. Inf. Technol.*, vol. 14, no. 7, pp. 3321–3331, Dec. 2022.
[44] F. Rustam, A. Ishaq, K. Munir, M. Almutairi, N. Aslam, and I. Ashraf, "Incorporating CNN features for optimizing performance of ensemble classifier for cardiovascular disease prediction," *Diagnostics*, vol. 12, no. 6, p. 1474, Jun. 2022.
[45] J. Amin, M. Sharif, M. Raza, T. Saba, R. Sial, and S. A. Shad, "Brain tumor detection: A long short-term memory (LSTM)-based learning model," *Neural Comput. Appl.*, vol. 32, no. 20, pp. 15965–15973, Oct. 2020.
[46] K. R. Spoorthy, A. R Mahdev, B. Vaishnav, and M. L.Shruthi, "Deep learning approach for radiogenomic classification of brain tumor," in *Proc. IEEE 19th India Council Int. Conf. (INDICON)*, Nov. 2022, pp. 1–6.

BHARGAV MALLAMPATI received the master's degree from the Department of Electrical Engineering, University of North Texas. His recent research interests include data mining, machine learning, and artificial intelligence, mainly working on creative computing and supervised machine learning.

ABID ISHAQ received the M.S. degree in computer science from the Department of Computer Science, Khwaja Fareed University of Engineering and Information Technology (KFUEIT), Rahim Yar Khan, Pakistan. His research interests include text mining, data mining, and machine learning and deep learning-based IoT.

FURQAN RUSTAM received the M.C.S. degree from the Department of Computer Science, The Islamia University of Bahawalpur, Pakistan, in October 2017, and the Master of Computer Science degree from the Department of Computer Science, Khwaja Fareed University of Engineering and Information Technology (KFUEIT), Rahim Yar Khan, Pakistan. He is currently pursuing the Ph.D. degree in computer science with University College Dublin, Ireland. He was a Research Assistant with the Fareed Computing & Research Center, KFUEIT. His recent research interests include data mining, machine learning, and artificial intelligence, mainly working on creative computing and supervised machine learning.

VENU KUTHALA is currently with Master Card Technologies, USA. His research interests include data mining, machine learning, and artificial intelligence.

SULTAN ALFARHOOD received the Ph.D. degree in computer science from the University of Arkansas. He is currently an Assistant Professor with the Department of Computer Science, King Saud University (KSU). Since joining KSU, in 2007, he has made several contributions to the field of computer science through his research and publications. He has published several research papers on cutting-edge topics, such as machine learning, recommender systems, linked open data, and text mining. His work includes proposing innovative approaches and techniques to enhance the accuracy and effectiveness of recommender systems and sentiment analysis.

IMRAN ASHRAF received the Ph.D. degree in information and communication engineering from Yeungnam University, South Korea, in 2018, and the M.S. degree in computer science from the Blekinge Institute of Technology, Karlskrona, Sweden, in 2010. He was a Postdoctoral Fellow with Yeungnam University, Gyeongsan-si, South Korea, where he is currently an Assistant Professor with the Information and Communication Engineering Department. His research interests include indoor positioning and localization in 5G and beyond, indoor location-based services in wireless communication, smart sensors for smart cars, and data analytics.