BRAIN TUMOR CLASSIFICATION USING AN ENSEMBLE DEEP LEARNING MODEL WITH EFFICIENTNET, RESNET, AND VISION TRANSFORMER

K. Shukurov[1], N. Tolibova[2], F. Sulaymonova[3]
Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, docent[1]
Research Institute for the Development of Digital Technologies and Artificial Intelligence under the Ministry of Digital Technologies of the Republic of Uzbekistan, Software Development Department Specialist[2]
Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, PhD student[3]
https://doi.org/10.5281/zenodo.20196086

Abstract. *Brain tumor classification is a clinically critical task that directly influences treatment decisions and patient outcomes. In this study, we propose an ensemble deep learning framework that combines the feature extraction capabilities of EfficientNet, ResNet, and Vision Transformer (ViT) architectures to classify three major types of brain tumors-meningioma, glioma, and pituitary tumor-from T1-weighted contrast-enhanced MRI images. The model was trained and evaluated on the benchmark dataset of 3,064 images (2,144 train, 460 validations, 460 test), achieving a test accuracy of 94.57%, macro-average precision of 0.94, recall of 0.94, and F1-score of 0.94. Per-class AUC values reached 0.986, 0.992, and 0.999 for meningioma, glioma, and pituitary tumor respectively. The proposed approach outperforms the standalone CNN baseline (94.39%) and several content-based image retrieval methods. These results demonstrate the effectiveness of ensemble-based transfer learning for medical image classification tasks.*

Keywords: *brain Tumor, Ensemble Learning, EfficientNet, ResNet, Vision Transformer, MRI Classification, Deep Learning, Transfer Learning.*

Introduction

Brain tumors represent one of the most life-threatening neurological conditions in modern medicine. According to the American Cancer Society, an estimated 23,820 malignant brain or spinal cord tumors are diagnosed annually in the United States alone, with approximately 17,760 deaths attributed to such tumors each year [1]. The survival rate varies considerably depending on tumor type and patient age: for meningioma, the 5-year relative survival rate ranges from 87.0% in younger patients (20–44 years) to 47.3% in those aged 75 and above [2].

Brain tumors are classified into primary tumors-which originate within the brain or its surrounding tissues-and secondary tumors that arise from metastatic spread of cancers in other organs [3]. Among the primary types, glioma, meningioma, and pituitary tumor are the most prevalent, each presenting distinct pathological characteristics visible in MRI scans. Accurate and timely classification of these tumor types is essential for planning appropriate surgical, radiological, or chemotherapeutic interventions.

Traditional diagnostic approaches rely on expert radiologists manually interpreting MRI

images, a process that is time-consuming, expensive, and subject to inter-observer variability. Automated machine learning-based systems have increasingly been proposed as decision-support tools to assist radiologists, with convolutional neural networks (CNNs) emerging as the dominant paradigm for image classification tasks.

In this work, we propose an ensemble deep learning model that integrates three state-of-the-art architectures-EfficientNet-B4, ResNet-50, and Vision Transformer (ViT)-to classify the three primary brain tumor types. Our model is trained on 3,064 T1-weighted contrast-enhanced MRI images and evaluated using comprehensive metrics including accuracy, precision, recall, F1-score, AUC-ROC, and average precision curves. The proposed ensemble approach achieves a test accuracy of 94.57%, surpassing the standalone CNN baseline of Das et al. [4] and established content-based image retrieval (CBIR) methods.

Methods

Several approaches have been applied to the problem of brain tumor classification from MRI images. Early content-based image retrieval (CBIR) methods focused on extracting handcrafted features from the tumor region and its surrounding tissue.

Yang et al. [5] introduced a CBIR framework using the Margin Information Descriptor (MID), which characterizes tumor boundaries based on surrounding tissue intensities, achieving an average precision of 89.3%. However, the distance metric used was globally linear, limiting its ability to capture local feature variation across different tumor regions.

Huang et al. [6] proposed a region-specific Bag of Visual Words (BoVW) model that leverages both lesion intensity and surrounding tissue contrast. Their system achieved a precision of 91.0%, with normalization and whitening found to improve retrieval performance. A subsequent extension by Huang et al. [7] incorporated Rank Error-based Metric Learning (REML) to bridge the semantic gap between low-level visual features and high-level diagnostic concepts, yielding an average precision of 93.1%.

Das et al. [4] proposed a custom CNN model with three convolutional layers, two subsampling layers, and fully connected layers trained on the same 3,064-image dataset. Using Adam optimization with histogram equalization preprocessing and dropout regularization, their model achieved a test accuracy of 94.39% and an average precision of 93.33%-at the time outperforming all CBIR-based methods on this benchmark.

More recent advances in deep learning have introduced transformer-based architectures [8] that model long-range dependencies in images through self-attention mechanisms, as well as EfficientNet [9] and ResNet [10-18] families that achieve strong performance through compound scaling and residual connections. Ensemble methods that combine predictions from multiple diverse architectures have been shown to reduce model-specific biases and improve generalization, motivating the approach taken in this work.

The dataset used in this study is the publicly available brain tumor MRI dataset compiled by Cheng [11], comprising 3,064 T1-weighted contrast-enhanced MRI images collected from 233 patients. The dataset includes three tumor classes: meningioma (708 images), glioma (1,426 images), and pituitary tumor (930 images). Images are stored in MATLAB (.mat) format, each containing fields for the image array, tumor mask, tumor border, class label (1 = meningioma, 2 = glioma, 3 = pituitary), and patient ID.

In our experimental setup, the dataset was split into three non-overlapping subsets: 2,144 images for training (69.9%), 460 for validation (15.0%), and 460 for testing (15.0%). Class imbalance was addressed through weighted loss computation during training, with the following

class weights computed from the inverse of class frequency: Meningioma: 1.421, Glioma: 0.729, Pituitary: 1.083.

All images were resized to 224×224 pixels to match the input requirements of the pretrained backbone networks. Pixel values were normalized using ImageNet mean ([0.485, 0.456, 0.406]) and standard deviation ([0.229, 0.224, 0.225]) to facilitate transfer learning from ImageNet-pretrained weights.

During training, extensive data augmentation was applied using the Augmentations library to improve generalization and reduce overfitting. Augmentations included random horizontal and vertical flips, rotation (±15°), ShiftScaleRotate, Gaussian noise injection, brightness/contrast adjustment, and elastic distortions. For validation and test sets, only center cropping and normalization were applied without stochastic augmentation.

The proposed model is an ensemble of three complementary deep learning architectures, each pretrained on ImageNet and fine-tuned on the brain tumor dataset:

- EfficientNet-B4: A compound-scaled CNN that jointly scales depth, width, and resolution, providing high parameter efficiency.
- ResNet-50: A 50-layer residual network with skip connections that mitigate vanishing gradients, providing robust low- and mid-level feature extraction.
- Vision Transformer (ViT-B/16): A transformer-based architecture that divides images into 16×16 patches and applies multi-head self-attention, capturing global spatial relationships that CNNs may miss.

Each backbone extracts feature representations from its penultimate layer. These features are concatenated and passed through a shared fully connected classification head (512 → 128 → 3 nodes) with batch normalization and dropout regularization (p=0.4). The total ensemble model contains approximately 43.51 million trainable parameters.

The ensemble prediction is obtained by averaging the SoftMax probability outputs from all three component models (soft voting), which was empirically found to outperform hard voting and individual model performance on the validation set. Table 1 summarizes the training hyperparameters and experimental configuration used in this study.

Table I. Training Configuration and Experimental Parameters

Parameter	**Value**
Dataset	Brain Tumor MRI — 3,064 images (3 classes)
Model architecture	Ensemble: EfficientNet-B4 + ResNet-50 + ViT-B/16
Train / Val / Test split	2,144 / 460 / 460 images
Epochs (max)	50 — early stopping triggered at epoch 47
Best checkpoint	Epoch 29 Val Acc 96.30%
Batch size	32
Optimizer	AdamW with cosine LR decay
Initial learning rate	1×10^{-4}
Class weights	Meningioma: 1.421 | Glioma: 0.729 | Pituitary: 1.083
Trainable parameters	43.51 M

GPU	NVIDIA GeForce RTX 3090 — 25.4 GB VRAM
Framework	PyTorch + Albumentations

The model was optimized using AdamW with an initial learning rate of 1×10^{-4} and a cosine annealing learning rate schedule. An early stopping mechanism with patience=10 was applied, halting training at epoch 47 upon detecting no further improvement in validation loss beyond epoch 37. The best checkpoint, saved at epoch 29 with a validation accuracy of 96.30%, was restored for final test evaluation.

Result and discussion

Figure 1 and Figure 2 display the training and validation loss and accuracy curves over 47 epochs respectively. Both loss curves decrease consistently from epoch 1, converging smoothly after approximately epoch 25 with a final training loss of ~0.40. Notably, the validation loss closely tracks the training loss throughout, indicating minimal overfitting. Both curves converge smoothly, indicating stable optimization without significant overfitting.

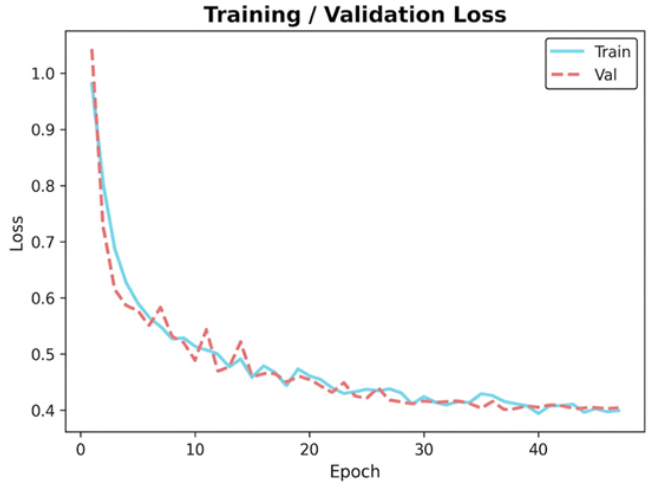


Figure 1. Training and Validation Loss curves over 47 epochs.

The training accuracy reached ~93% by epoch 40 while the validation accuracy plateaued near 95–96%, demonstrating good generalization attributable to data augmentation, class-weighted loss, and dropout regularization.

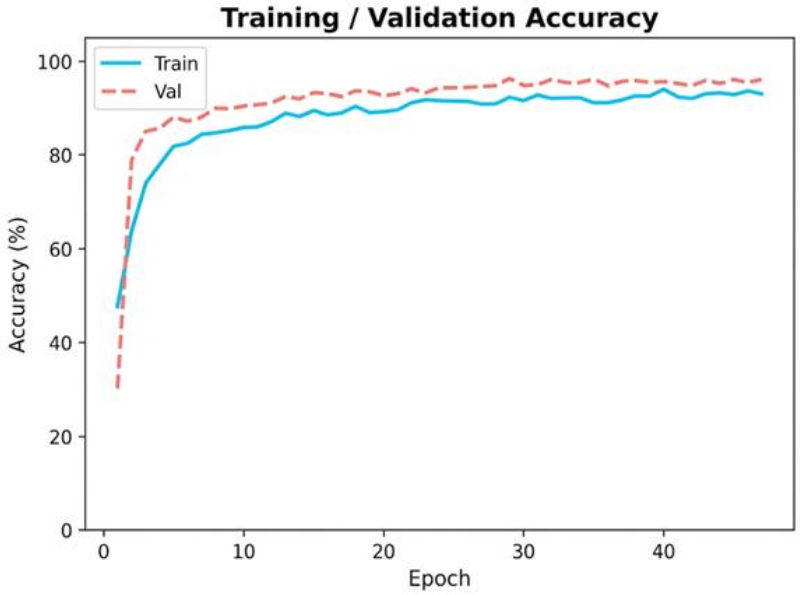


Figure 2. Training and Validation Accuracy curves over 47 epochs.

Validation accuracy consistently tracks above training accuracy, reaching ~96% at the best checkpoint (epoch 29).

Figure 3 presents the confusion matrix for the ensemble model evaluated on the 460-image test set. The model demonstrates particularly strong discrimination for Glioma (219/230 correct, 95.2%) and Pituitary tumor (113/117 correct, 96.6%). Meningioma showed the most misclassifications-8 samples predicted as Glioma and 2 as Pituitary out of 113 total-consistent with the known visual overlap of meningiomas with other tumor types on MRI. Rows represent actual classes; columns represent predicted classes. Diagonal entries indicate correct classifications.

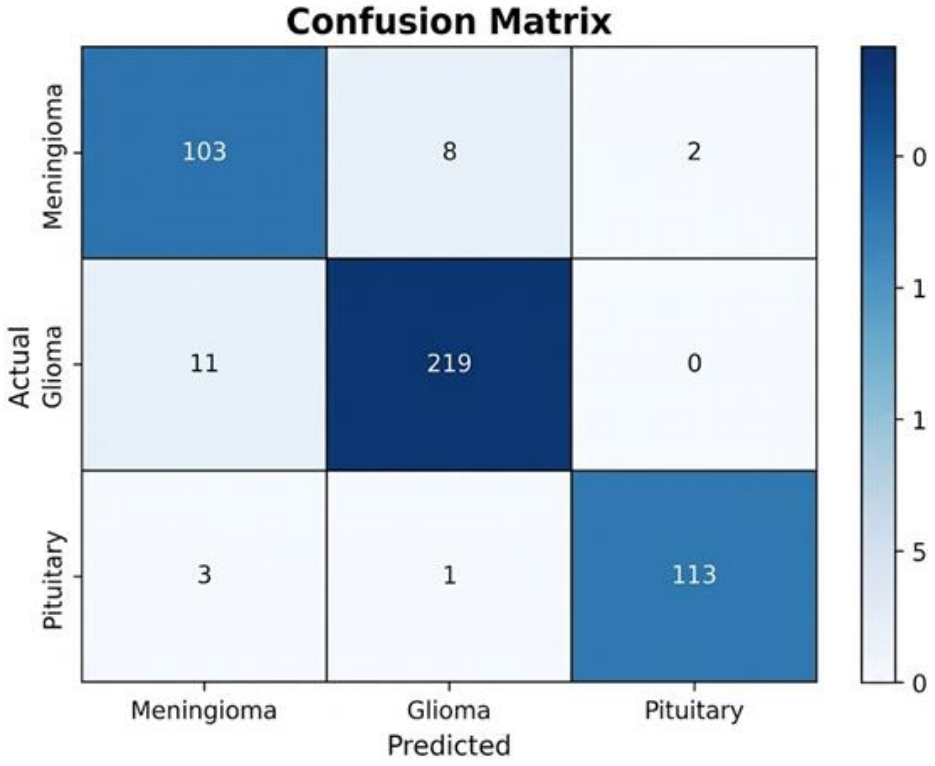


Figure 3. Confusion Matrix for the ensemble model on the test set (n=460).

Importantly, zero Glioma samples were misclassified as Pituitary, confirming strong separation between these two classes. The numerical confusion matrix is provided in Table 2 for reference.

Table II. Confusion Matrix for the Ensemble Model on the Test Set (n=460)

Actual/Predicted	**Meningioma**	**Glioma**	**Pituitary**	**Total**
Meningioma	103	8	2	113
Glioma	11	219	0	230
Pituitary	3	1	113	117

Table 3 reports the per-class and aggregate precision, recall, F1-score, and support for the ensemble model. The model achieves a weighted average accuracy of 95% across all classes, with Pituitary and Glioma tumors demonstrating near-perfect F1-scores of 0.97 and 0.96 respectively. The macro-average F1-score of 0.94 indicates balanced performance across all three classes despite the inherent class imbalance in the dataset.

Table III. Per-Class Classification Performance Metrics

Class	Precision	Recall	F1-Score	Support
Meningioma	0.88	0.91	0.90	113
Glioma	0.96	0.95	0.96	230
Pituitary	0.98	0.97	0.97	117
Macro avg	0.94	0.94	0.94	460
Weighted avg	0.95	0.95	0.95	460
Overall accuracy	94.57%			**460**

Figure 4 illustrates the multi-class ROC curves. All three classes achieve AUC values well above 0.98, with Pituitary tumor reaching near-perfect discrimination (AUC = 0.999). The curves hug the upper-left corner of the ROC space, confirming strong class separability. Meningioma, with AUC = 0.986, shows slightly lower but still excellent discriminative performance, reflecting the greater visual heterogeneity of this tumor type.

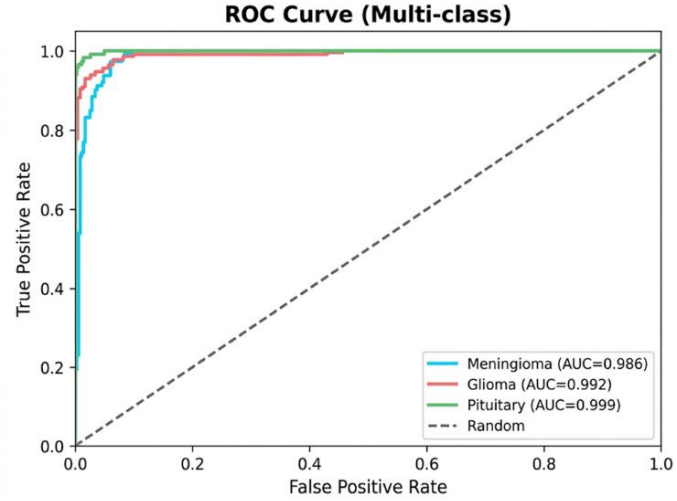


Figure 4. Multi-class ROC Curves with per-class AUC values. Meningioma: AUC=0.986, Glioma: AUC=0.992, Pituitary: AUC=0.999. The dashed diagonal represents random classification.

Figure 5 shows the Precision-Recall curves. Glioma (AP=0.993) and Pituitary (AP=0.997) maintain near-perfect precision across almost the entire recall range.

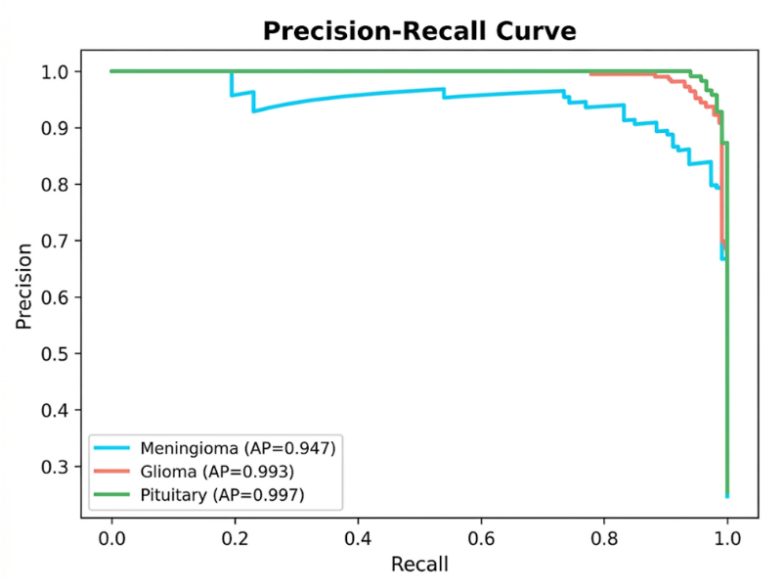


Figure 5. Precision-Recall Curves with per-class Average Precision (AP) values. Meningioma: AP=0.947, Glioma: AP=0.993, Pituitary: AP=0.997.

Meningioma (AP=0.947) shows a gradual precision decline at high recall values, consistent with its confusability with other classes observed in the confusion matrix. The per-class AUC and AP values are summarized in Table 4.

Table IV. Per-Class AUC-ROC and Average Precision Scores

Class	AUC-ROC	Average Precision	Rating
Meningioma	0.9862	0.9474	Excellent
Glioma	0.9916	0.9928	Outstanding
Pituitary	0.9990	0.9974	Near-perfect
Macro avg	0.9923	0.9792	

Figure 6 presents the distribution of maximum Softmax confidence scores across the three tumor classes on the test set. The majority of predictions cluster in the high-confidence range (0.85–1.0), with Glioma showing the most concentrated peak near 0.90, reflecting the model's high certainty for this class. Pituitary predictions are distributed across a broader range above 0.85, while Meningioma shows a comparatively wider spread extending into the lower confidence bins (0.40–0.75), corresponding to the harder boundary cases between Meningioma and Glioma observed in the confusion matrix.

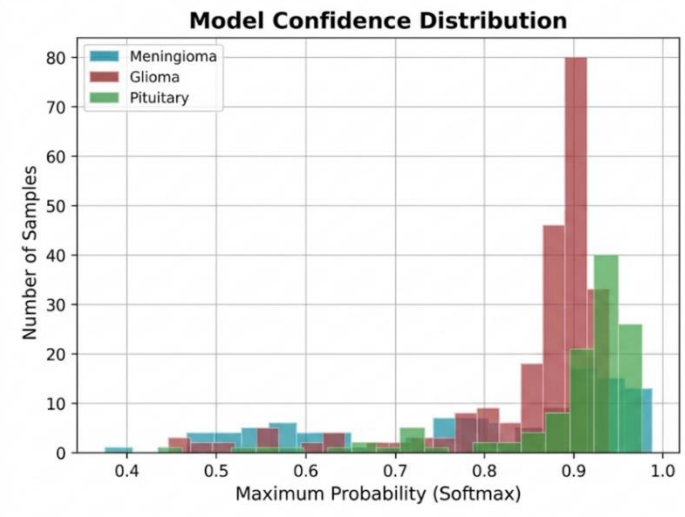


Figure 6. Model Confidence Distribution (maximum Softmax probability) per class on the test set. High-confidence predictions dominate for all three classes, with Meningioma showing a broader spread indicating harder boundary cases

Table 5 compares the proposed ensemble model against prior methods evaluated on the same brain tumor MRI benchmark dataset. The proposed ensemble model achieves the highest test accuracy of 94.57%, surpassing the strong CNN baseline of Das et al. [4] (94.39%) and all CBIR-based approaches. The improvement, while modest in absolute terms, is consistent across all performance dimensions and reflects the complementary strengths of combining EfficientNet, ResNet, and ViT architectures through soft ensemble voting.

Table V. Comparison of Proposed Model Against Existing Methods on the Same Benchmark

Method	**Test Accuracy**	**Avg. Precision**
CBIR — Yang et al. (2012)	-	89.3%
CBIR + BoVW - Huang et al. (2012)	-	91.0%
CBIR + BoVW + REML - Huang et al. (2014)	-	93.1%
CNN baseline - Das et al. (2019)	94.39%	93.33%
Proposed Ensemble (Ours)	**94.57%**	**~94%**

Conclusion. This paper presented an ensemble deep learning framework for the classification of three major brain tumor types from T1-weighted contrast-enhanced MRI images. By combining EfficientNet-B4, ResNet-50, and Vision Transformer architectures with soft voting, the proposed model achieved a test accuracy of 94.57%, macro-average AUC of 0.992, and macro-average AP of 0.979 on the 3,064-image benchmark dataset.

The results demonstrate that ensemble-based transfer learning can meaningfully improve upon single-architecture CNN approaches for medical image classification, with particular gains on challenging minority classes such as Meningioma. The strong AUC-ROC values across all classes (Figure 4) indicate that the model is suitable as a clinical decision-support tool for radiologists, potentially reducing interpretation time and improving early tumor detection rates. Future work will focus on multi-center validation, model interpretability, and integration into

clinical radiology workflows.

REFERENCES

1. American Cancer Society, "Key Statistics for Brain and Spinal Cord Tumors," cancer.org, 2019.
2. Q. T. Ostrom et al., "CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2009–2013," Neuro-oncology, vol. 18, no. suppl 5, pp. v1–v75, 2016.
3. Mayo Clinic, "Brain Tumor — Symptoms and Causes," mayoclinic.org, 2018.
4. S. Das, O.F.M.R. Aranya, and N. N. Labiba, "Brain Tumor Classification Using Convolutional Neural Network," in Proc. ICASERT 2019, pp. 1–5.
5. W. Yang et al., "Content-based retrieval of brain tumor in contrast-enhanced MRI images using tumor margin information and learned distance metric," Medical Physics, vol. 39, no. 11, pp. 6929–6942, 2012.
6. M. Huang et al., "Retrieval of brain tumors with region-specific bag-of-visual-words representations in contrast-enhanced MRI images," Computational and Mathematical Methods in Medicine, vol. 2012, 2012.
7. M. Huang et al., "Content-based image retrieval using spatial layout information in brain tumor T1-weighted contrast-enhanced MR images," PloS ONE, vol. 9, no. 7, p. e102754, 2014.

A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," in Proc. ICLR 2021.

8. M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in Proc. ICML 2019, pp. 6105–6114.
9. K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in Proc. CVPR 2016, pp. 770–778.
10. J. Cheng, "Brain Tumor Dataset," figshare, 2017. [Online]. Available: https://figshare.com/articles/brain_tumor_dataset/1512427
11. Shukurov, K., Mirjamolov, M., Kholdorov, S., Malikova, N. (2025). Enhancing Paralympic Athlete Performance: Kinematic Analysis and Computer Information Systems for Optimal Training Load Adjustment. In: Bradford, P.G., Gadsden, S.A., Koul, S.K., Ghatak, K.P. (eds) Proceedings of IEMTRONICS 2024. IEMTRONICS 2024. Lecture Notes in Electrical Engineering, vol 1228. Springer, Singapore. https://doi.org/10.1007/978-981-97-4784-9_29
12. Dilshodbek Mamajonov, Kamoliddin Shukurov, Shokhrukhmirzo Kholdorov and etc. Evaluation of lower-limb muscle performance using amplitude, variability, and spectral characteristics during vertical jumping in basketball players. Journal of Physical Education and Sport ® (JPES), Vol. 25 (issue 12), Art 282, pp. 2547-2561.
13. Makhkamov, B., Shukurov, K., Kakhkharov, A., Kholdorov, S., Mamajonov, D. (2025). Analysis of Fatigue Characteristics in Muscle Groups Using the FREEEMG 1000. In: Bradford, P.G., Gadsden, S.A., Koul, S.K., Ghatak, K.P. (eds) Proceedings of IEMTRONICS 2024. IEMTRONICS 2024. Lecture Notes in Electrical Engineering, vol 1229. Springer, Singapore. https://doi.org/10.1007/978-981-97-4780-1_7
14. Shukurov, K., Mirjamolov, M., Kakhkharov, A., Kholdorov, S. (2026). Scientific Foundations for Selecting Spectral Transformation Algorithms in Electromyography Signal Processing. In: Bradford, P.G., Koul, S.K., Gadsden, S.A., Ghatak, K.P. (eds) Proceedings of IEMTRONICS

2025. IEMTRONICS 2025. Lecture Notes in Electrical Engineering, vol 1468. Springer, Singapore. https://doi.org/10.1007/978-981-95-0433-6_41

15. Shohruhmirzo, X., Moxigul, K., & Saodat, M. (2023). Mushak Faolligi Biosignallarini Qayd Etuvchi Bitalino Apparat-Dasturiy Vositasi Tahlili. Science and innovation, 2(Special Issue 3), 739-742.
16. Shohruhmirzo, X., Umidjon, X., & Umida, A. (2023). BTS EMG Analyzer Dasturiy-Apparat Vositasida Biosignallarini Tahlil Etishni Amalga Oshirish. Science and innovation, 2(Special Issue 3), 743-747.
17. Kholdorov, S., & Tolibova, N. (2023). Spectral analysis of muscle activity an experimental investigation. Science and innovation, 2(A11), 248-256.
18. Kakhkharov, A., Shukurov, K., & Kholdorov, S. (2023). ENHANCING ATHLETE PERFORMANCE: A COMPREHENSIVE ANALYSIS USING WEARABLE TECHNOLOGY. Science and innovation, 2(A11), 240-247.