International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Abnormality Detection in Spinal Cord MRI Using K-Means Clustering

Padala Soumya
Student
Department of Information Technology
JBIET Autonomous,TG
Moinabad,India
padalasoumya14@gmail.com

Kurukunda Jaswanth
Student
Department of Information Technology
JBIET Autonomous,TG
Moinabad,India
jaswanthkurukunda@gmail.com

Mrs. Vallala Bhavana
M.tech
Assistant professor
Department of information and technology
JBIET Autonomous,TG
Moinabad,India
vallalabhavana4@gmail.com

G Sree Laxmi
Student
Department of Information Technology
JBIET Autonomous,TG
Moinabad,India
sreelaxmi173@gmail.com

Pebbeti Siddhartha
Student
Department of Information Technology
JBIET Autonomous,TG
Moinabad,India
siddharthapebbeti@gmail.com

***Abstract*— Early and accurate detection of spinal cord abnormalities is crucial for preventing severe neurological deficits and enabling timely medical interventions. Magnetic Resonance Imaging (MRI) provides detailed visualization of spinal anatomy, yet manual inspection is time-consuming and prone to subjective errors. This study presents an automated approach for abnormality detection in spinal cord MRI images using unsupervised clustering techniques. Histogram-based features are extracted from both DICOM and JPEG formats, followed by preprocessing and standardization to construct a robust feature space. Multiple clustering algorithms, including K-Means, BIRCH, and DBSCAN, are applied to identify abnormal patterns, with evaluation conducted using Silhouette, Davies-Bouldin, and Calinski-Harabasz indices. Results indicate that K-Means outperforms the other methods in terms of clustering cohesion and separability, with lower computational time and higher interpretability. Anomaly scores and Z-scores are computed to quantify deviations from normal patterns, enabling severity-based assessment. Additionally, a Flask-based front-end interface integrated with SQLite3 facilitates user authentication and interactive image upload, allowing clinicians to visualize clustering results and abnormality severity in real-time. This approach demonstrates the potential of combining histogram features with unsupervised clustering for efficient and scalable spinal cord abnormality detection.**

***Keywords*— spinal cord MRI, abnormality detection, histogram features, K-Means clustering, BIRCH, DBSCAN, anomaly score, unsupervised learning**

I. INTRODUCTION

Spinal cord abnormalities, including lesions, tumors, and degenerative disorders, pose significant threats to neurological health, often resulting in impaired mobility, chronic pain, or severe neurological deficits if not detected early [1]. Magnetic Resonance Imaging provides high-resolution visualization of the spinal cord and surrounding tissues, allowing detailed examination of structural and functional irregularities. Manual interpretation of these images is time-consuming, highly dependent on radiologist expertise, and prone to inconsistencies due to subjective judgment or fatigue [2]. The growing volume of imaging data further exacerbates these challenges, making timely and accurate analysis increasingly difficult in clinical settings [3].

Automated detection and analysis methods address these limitations by systematically evaluating image patterns and quantifying deviations from normal tissue. Techniques such as clustering and feature extraction have shown potential in identifying abnormalities within MRI scans, enabling objective differentiation between normal and pathological regions [4]. By extracting features from intensity distributions, textures, and spatial patterns, these systems provide numerical measures that support assessment of abnormality severity [5]. Such approaches reduce reliance on manual interpretation, improve consistency, and allow earlier identification of potentially critical conditions [6].

In addition to analysis, effective visualization and interaction tools enhance clinical utility. Web-based or software interfaces enable secure image upload, interactive visualization of detected anomalies, and real-time feedback on abnormal regions [7]. Integration with clustering methods allows grouping of similar patterns and highlights regions of concern, providing both quantitative and visual insights for clinicians [8]. The use of unsupervised learning further supports identification of abnormalities without requiring pre-labeled datasets, making detection scalable and adaptable to varying image types and resolutions [9]. Efficient automated systems also reduce workload, improve throughput, and allow consistent monitoring of patient conditions over time [10].

The primary objective is to develop a comprehensive system for spinal cord MRI analysis that extracts meaningful features, calculates anomaly scores, and applies unsupervised clustering to detect abnormal regions accurately. The system aims to provide real-time visualization of abnormality severity, facilitate objective assessment, and enhance the overall efficiency and reliability of neurological evaluations, ultimately supporting improved clinical decision-making and patient outcomes.

II. Related Work

MRI-based automatic detection and segmentation of brain and spinal abnormalities has received considerable attention due to its potential to improve diagnostic accuracy and reduce manual workload. Ray and Bandyopadhyay [11] provided a comprehensive review of MRI-based techniques for brain tumor detection, highlighting the increasing adoption of automated segmentation methods that integrate machine learning and image processing. Their study emphasizes that conventional manual interpretation is prone to observer variability, and automated methods can enhance consistency and efficiency in clinical settings.

Pravallika and Reddy [12] developed an optimized segmentation method combining K-Means++ with vectorized fuzzy membership computation, demonstrating that hybrid clustering approaches can improve boundary delineation and segmentation accuracy. Their work reinforces the significance of integrating multiple clustering strategies to capture complex intensity variations in MRI images, reducing misclassification of abnormal tissue.

Silveira et al. [13] proposed an enhanced U-Net with an inception module and dual-output mechanism for lumbar spine segmentation. The model efficiently handles complex spinal structures and provides accurate segmentation of vertebral regions, supporting quantification of abnormal areas. This approach highlights the effectiveness of convolutional neural networks in capturing hierarchical features and accommodating variations in spinal MRI scans.

Kantharimuthu and Marichamy [14] combined SVM and CNN-based methods for brain tumor segmentation and classification, demonstrating improved accuracy and reduced computational time. The hybrid framework effectively integrates feature extraction with classification, enabling identification of both normal and abnormal regions in MRI images and supporting quantitative assessment.

Liu et al. [15] explored white matter functional network abnormalities using BOLD fMRI, highlighting the importance of identifying functional deviations in neurological disorders. Their findings suggest that advanced MRI analysis combined with quantitative evaluation methods can detect subtle pathological patterns that may be overlooked in standard imaging.

Goyal, Banerjee, and Kaur [16] investigated segmentation and wavelet-based feature extraction for brain tumor detection. The use of wavelet transforms enhances texture analysis and identifies fine-grained abnormalities, complementing traditional intensity-based approaches and supporting more reliable anomaly detection.

Vajiram and Sivakumar [17] focused on quantitative assessment of brain tumor-related epilepsy using AI models. Their work illustrates the role of predictive modeling in evaluating abnormal regions and providing severity metrics, offering objective measures that can aid in clinical decision-making.

Mejri et al. [18] proposed an approach combining unsupervised domain adaptation with one-class anomaly detection to address scarcity of labeled abnormal data. This methodology is particularly relevant for spinal MRI analysis, where labeled abnormal images are limited, emphasizing the need for adaptive and unsupervised techniques to improve detection accuracy.

Jothi et al. [19] introduced a Falcon-optimized CNN model for Alzheimer’s disease detection using MRI images. The model demonstrates that optimized deep learning architectures can efficiently identify subtle abnormal patterns in complex neurological datasets, reducing false negatives and improving reliability.

Tsygankov et al. [20] explored real-time anomaly detection in healthcare imaging using intelligent technologies. Their study emphasizes integration of rapid analysis frameworks with visualization tools, highlighting the importance of interactive systems that allow clinicians to interpret detected abnormalities effectively and make timely decisions.

Overall, these studies collectively demonstrate that automated MRI analysis, integrating clustering, deep learning, feature extraction, and adaptive detection strategies, significantly improves accuracy, efficiency, and interpretability of abnormality detection. The findings underscore the importance of combining intensity, texture, and spatial features with unsupervised or hybrid learning approaches to address challenges in spinal and brain MRI segmentation, supporting scalable and clinically meaningful outcomes.

III. Materials And Methods

The system aims to provide automated detection and analysis of spinal cord abnormalities using MRI images, addressing limitations of manual interpretation and improving diagnostic efficiency. It incorporates advanced image processing techniques to extract meaningful features from both DICOM and JPEG images, capturing intensity distributions, texture patterns, and structural variations that distinguish normal from abnormal spinal regions. Preprocessing steps standardize the images and organize extracted features into a structured format, enabling systematic analysis.

Anomaly scores and statistical measures, such as Z-scores, are computed to quantify deviations from normal patterns, offering a numerical basis for assessing the severity of abnormalities. The system applies unsupervised clustering methods, including K-Means, BIRCH, and DBSCAN, to group images based on feature similarity, facilitating identification of abnormal regions without the need for labeled data. These clusters highlight areas of concern, enabling objective comparison between normal and pathological cases.

A web-based interface built with Flask provides secure user interaction, including registration, login, and image upload. Users can visualize clustering results and severity in real time, making the system practical for clinical or research applications. The approach integrates scalable and flexible clustering techniques, building upon prior work in multi-organ medical image segmentation [21], brain metabolic imaging [22], and vertebral fracture classification [23], to enhance accuracy, efficiency, and usability.

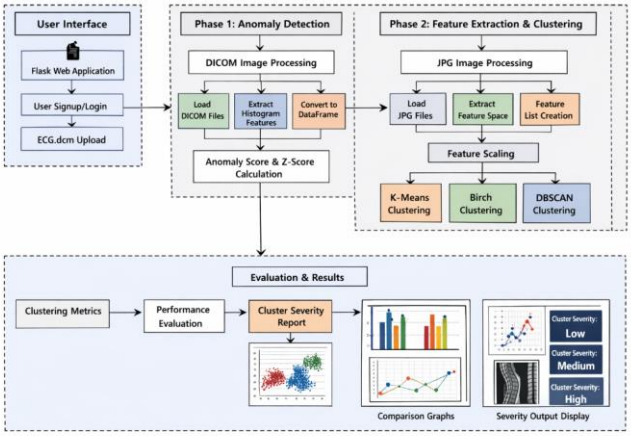

Fig.1 System Architecture

Fig. 1 illustrates a multi-phase system architecture for ECG analysis. It begins with a User Interface for DICOM file uploads via a Flask web application. Phase 1 focuses on anomaly detection through DICOM processing and Z-Score calculation, while Phase 2 handles feature extraction and scaling for K-Means, Birch, and DBSCAN clustering algorithms. Finally, the Evaluation & Results section generates performance metrics and cluster severity reports, displaying findings through comparison graphs and severity levels.

A) Dataset Collection:

The dataset used for this study consists of spinal cord MRI images, collected from publicly available sources to ensure a diverse representation of normal and abnormal spinal conditions. The dataset includes both DICOM and JPEG image formats, providing high-resolution scans that capture structural details of the spinal cord and surrounding tissues. These images encompass a range of spinal abnormalities, including lesions, degenerative changes, and other structural irregularities, allowing comprehensive analysis of varying severity levels.

Each image in the dataset is labeled with metadata, including patient demographics, imaging modality, and scan orientation, which facilitates organization and reference during analysis. The dataset contains images from multiple spinal regions, including cervical, thoracic, and lumbar segments, providing a holistic view of spinal health. The inclusion of images with different intensity distributions and tissue contrasts enables the study of subtle variations and abnormal patterns across individuals.

By using this dataset, the system can access a wide spectrum of spinal conditions, ensuring that any analysis or evaluation reflects real-world variability. The availability of both high-quality DICOM scans and more common JPEG images enhances the applicability of the system to clinical, research, and educational settings, supporting generalization and robustness in detecting and assessing spinal cord abnormalities.

B) Pre-Processing:

Pre-processing is a critical step to prepare spinal cord MRI images for systematic analysis, ensuring consistency and reliability of extracted features. In Phase 1, the focus is on handling the DICOM images, which are the standard format for medical imaging. A set of dedicated functions is defined to read and interpret DICOM files, converting the image pixel data into a structured array suitable for further analysis. In addition to DICOM handling, functions are also defined for working with JPEG images, allowing flexibility in image sources. To extract meaningful quantitative information from the images, histogram-based features are calculated. These features capture intensity distributions across the spinal cord regions, reflecting variations between normal and abnormal tissue. Once the features are extracted, they are organized into a structured dataframe, providing a uniform format that facilitates subsequent computations. Statistical measures, including anomaly scores and Z-scores, are computed to quantify deviations from normal spinal patterns. These measures help identify unusual intensity patterns and provide a numeric basis for assessing the severity of abnormalities, serving as the foundation for all further analysis.

Phase 2 focuses on extending feature extraction to the JPEG images and defining a comprehensive feature space. Pre-processing in this phase begins with loading the images and ensuring uniform dimensions and intensity scaling, which standardizes the data across the dataset. Advanced feature extraction techniques are applied to capture spatial, textural, and intensity-based characteristics of the spinal cord, allowing differentiation between normal and pathological regions. Each image's extracted features are appended to a centralized list, consolidating data for consistent access. To improve numerical stability and ensure that all features contribute equally in later stages, the feature values are scaled using standardization techniques. This scaling transforms the feature distributions to have uniform mean and variance, eliminating biases caused by differing feature ranges. These pre-processing steps ensure that the dataset is clean, standardized, and rich in relevant information, providing a robust foundation for subsequent analysis and visualization. The combined preparation from Phase 1 and Phase 2 enables the system to handle multiple image formats, accommodate variability in intensity and texture, and support accurate identification of abnormal spinal patterns.

C) Clustering:

K-Means: K-Means is an unsupervised clustering technique that partitions data into a predefined number of clusters by minimizing the sum of squared distances between data points and their cluster centroids. In spinal cord MRI analysis, it groups images or extracted features based on intensity patterns and histogram characteristics, allowing the identification of normal and abnormal regions. This method is highly effective for datasets with well-separated and spherical clusters, providing interpretable results. It quickly assigns MRI images to clusters, enabling severity-based categorization of abnormalities. The iterative approach refines centroids until convergence, ensuring that similar spinal patterns are grouped together, which facilitates the detection of subtle variations and supports quantification of anomaly scores for further clinical evaluation.

BIRCH: BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies) is a hierarchical clustering algorithm designed to handle large datasets efficiently. It incrementally builds a tree structure called the CF (Clustering Feature) tree, summarizing data points and merging similar clusters. For spinal cord MRI analysis, BIRCH organizes images based on extracted histogram features, capturing subtle variations in intensity and texture. The method reduces memory usage while maintaining cluster quality, making it suitable for extensive image datasets. By grouping similar patterns hierarchically, it enables identification of anomalies at multiple levels of severity. BIRCH is adaptive to noise and can produce clusters with varying densities, supporting comprehensive analysis of abnormal spinal regions and aiding in quantitative assessment.

DBSCAN: DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is a clustering technique that groups data points based on density and identifies points in low-density regions as outliers. In spinal cord MRI analysis, DBSCAN detects clusters of similar images based on histogram and feature space densities while labeling abnormal patterns as noise. This approach is particularly useful for identifying irregular or sparse anomalies that do not conform to predefined cluster shapes. DBSCAN does not require prior knowledge of the number of clusters, allowing flexible discovery of varying abnormality patterns. By focusing on density, it isolates unusual spinal regions, providing a quantitative assessment of severity and complementing other clustering methods in identifying subtle and diverse abnormalities.

IV. EXPERIMENTAL RESULTS

Table.1 Performance Evaluation Table

Algorithm	**Avg_Time**	**Avg_Silhouette**	**Avg_DBI**	**Avg_CHI**
K-Means	0.003802	0.309891	1.021145	6806.476363
DBSCAN	0.665623	-1.000000	-1.000000	-1.000000
BIRCH	0.012795	0.200614	1.069626	3704.100620

Table 1 presents the comparison of K-Means, DBSCAN, and BIRCH clustering performance metrics.

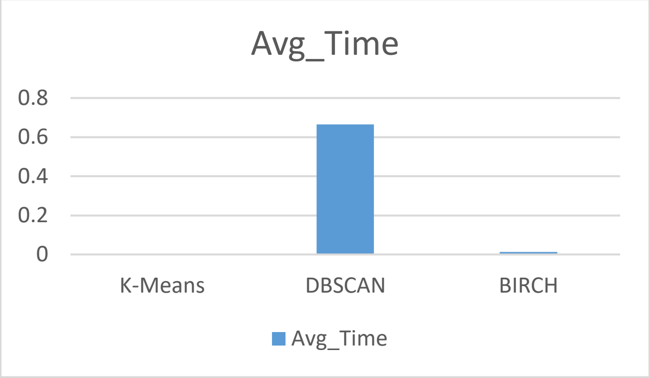

Fig.2 Avg_Time Comparison Graph

Fig. 2 displays a bar chart comparing the average execution times of K-Means, DBSCAN, and BIRCH clustering algorithms.

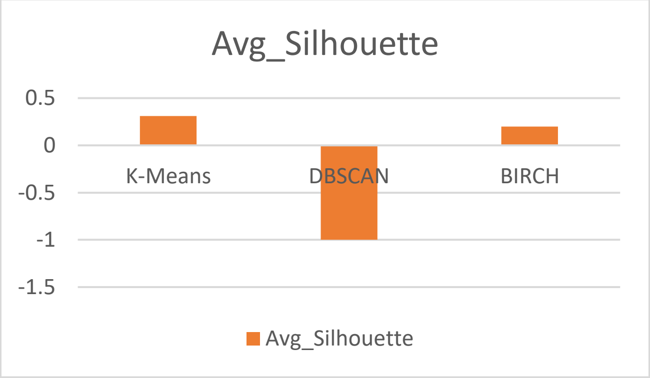

Fig.3 Avg_Silhouette Comparison Graph

Fig. 3 illustrates the average silhouette scores for K-Means, DBSCAN, and BIRCH, highlighting K-Means as the most effective clustering algorithm.

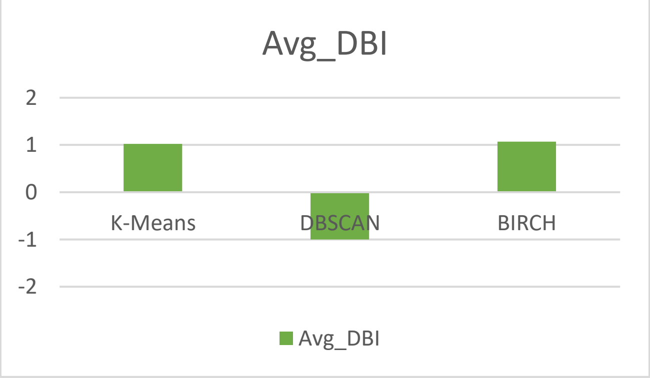

Fig.4 Avg_DBI Comparison Graph

Fig. 4 compares average Davies-Bouldin Index (DBI) scores, with K-Means and BIRCH performing similarly while DBSCAN shows an outlier value.

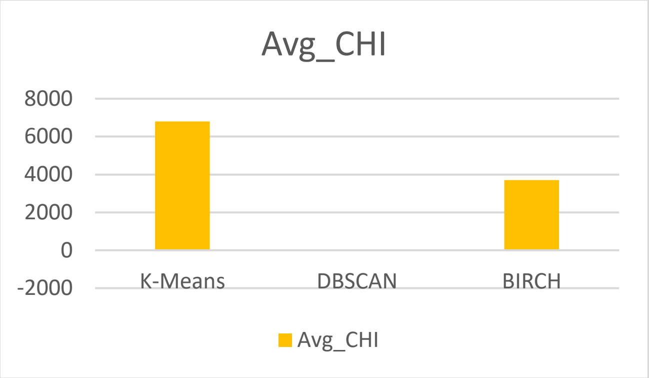

Fig.5 Avg_CHI Comparison Graph

Fig. 5 displays the Average Calinski-Harabasz Index (CHI), with K-Means achieving the highest score compared to BIRCH and DBSCAN.

V. CONCLUSION

Analysis of spinal cord MRI images using unsupervised clustering techniques demonstrates a reliable and efficient method for detecting abnormalities and quantifying severity. By extracting histogram-based features from both DICOM and JPEG formats and standardizing them for consistency, the study effectively captures intensity and texture patterns relevant to abnormal regions. Among the clustering algorithms evaluated, K-Means exhibits superior performance, achieving higher Silhouette and Calinski-Harabasz scores along with lower Davies-Bouldin indices, indicating well-separated and cohesive clusters. BIRCH shows moderate clustering capability with reasonable computational efficiency, while DBSCAN underperforms due to sensitivity to parameter selection and density variations in the feature space. Anomaly scores and Z-scores provide quantitative measures to identify deviations from normal spinal patterns, enabling prioritization of cases with higher severity. Integration with a Flask-based interface and SQLite3 database allows seamless user interaction, including secure login, image upload, and real-time visualization of abnormality clusters. These results confirm that combining histogram features with clustering approaches not only automates the detection process but also provides interpretable insights into spinal cord conditions, potentially assisting radiologists in diagnostic decision-making and reducing manual workload. The framework proves scalable and adaptable for further refinement with larger datasets or additional imaging modalities.

Future enhancements can focus on improving feature representation and clustering accuracy by incorporating advanced image descriptors, such as texture, shape, and deep learning-based embeddings, to capture more complex spinal abnormalities. Multi-modal integration of MRI sequences, including T1, T2, and diffusion-weighted images, can provide richer information for precise anomaly detection. Adaptive clustering methods and hybrid algorithms combining supervised and unsupervised learning could further improve detection performance and reduce sensitivity to parameter selection. Additionally, implementing real-time processing pipelines with optimized computational efficiency would facilitate faster clinical adoption. Expanding the dataset with diverse patient demographics and pathological cases can improve model generalization. Integration with explainable AI techniques can also provide interpretable insights for clinicians, enhancing trust and usability.

REFERENCES

[1] Lew, S. H., Jeong, Y. J., Roh, Y. R., & Kang, D. H. (2025). Exploratory Cluster-Based Radiographic Phenotyping of Degenerative Cervical Disorder: A Retrospective Study. *Medicina*, *61*(5), 916.

[2] Kumar, K., Agrawal, S., Suwalka, I., Iwendi, C., & Biamba, C. N. (2025). Early Diagnosis of Alzheimer's Disease using Adaptive Neuro K-means Clustering Technique. *IEEE Access*.

[3] Zhang, Q., Chen, X., He, Z., Wang, K., Wu, L., Shen, H., & Sun, J. (2025). U2AD: Uncertainty-based Unsupervised Anomaly Detection Framework for Detecting T2 Hyperintensity in MRI Spinal Cord. *arXiv preprint arXiv:2503.13400*.

[4] Choudhary, A. K., Mishra, K. N., Lal, R. K., & Mishra, A. (2025, November). Hybrid Machine Learning Framework for Early Detection of Fetal Brain Anomalies Using MRI and Texture Feature Analysis. In *2025 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)* (pp. 416-421). IEEE.

[5] Mac-Trevor, L., & Adu-Manu, K. S. (2025). Merged Fuzzy C-Means and K-Means Algorithm for Brain Tumour Segmentation. *AI, Computer Science and Robotics Technology*.

[6] Christodoulou, R. C., Vamvouras, G., Sarquis, M. D., Petrou, V., Papageorgiou, P. S., Rivera, L., ... & Papageorgiou, S. G. (2025). Magnetic Resonance Imaging and Cerebrospinal Fluid Biomarker Clustering Defines Biological Subtypes of Alzheimer's Disease. *Biomedicines*, *13*(11), 2632.

[7] Kiruthiga, G., Nithyasri, A., & Saravanakumar, N. M. (2025, February). A Retrospective Survey on the Recent Trends in Brain Tumors and their Detection Methodologies. In *2025 3rd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)* (pp. 2111-2124). IEEE.

[8] Lyu, C., Xia, Y., Li, Y., Que, J., Han, F., Guan, Q., ... & Gao, B. (2025). Dynamic contrast-enhanced MRI reveals glymphatic dysfunction in mice with depressive-like behavior. *Neurobiology of Disease*, 107169.

[9] Thamilarasi, V., Kanya, N., Roselin, R., Sam, D., & Babu, S. (2025). Novel Study of MRI Brain Tumor Detection and Segmentation by Digital Image Processing Techniques. *Brain Informatics Technology*, 281-303.

[10] Alsahanova, N., Yarkin, V., Spodarev, E., Bronov, O., Bychenko, V., Marinets, A., ... & Sharaev, M. (2025). Knowledge-informed randomized machine learning and data fusion for anomaly areas detection in multimodal 3D images. *Information Sciences*, *686*, 121354.

[11] Ray, A., & Bandyopadhyay, S. K. (2025). A Review of MRI Based Automatic Brain Tumor Detection and Segmentation. *Applied Healthcare Science*, *1*, 49-60.

[12] PRAVALLIKA, R. L., & REDDY, R. P. K. (2025). OPTIMIZED BRAIN MRI SEGMENTATION: K-MEANS++ AND VECTORISED FUZZY MEMBERSHIP COMPUTATION. *Journal of Theoretical and Applied Information Technology*, *103*(5).

[13] Silveira, J. T., S, G., & Kundapur, P. P. (2025). Automated lumbar spine segmentation in MRI using an enhanced U-Net with inception module and dual-output mechanism. *Scientific Reports*, *15*(1), 39215.

[14] Kantharimuthu, M., & Marichamy, M. (2025). A Hybrid Clustering Approach: Segmentation and Classification of Brain Tumour Utilizing SVM and CNN Methods. *Traitement du Signal*, *42*(4), 2281.

[15] Liu, J., Yang, W., Wen, X., Wang, J., Du, Z., Yang, L., ... & Yuan, K. (2025). Abnormalities and Plasticity of White Matter Functional Networks in Opioid Use Disorder: A BOLD fMRI Study.

[16] Goyal, R., Banerjee, K., & Kaur, J. (2025, March). Improve Detection of Brain Tumors Using Segmentation and Wavelet-Based Features.

In *International Conference on Computing and Communication Systems for Industrial Applications* (pp. 751-762). Singapore: Springer Nature Singapore.

[17] Vajiram, J., & Sivakumar, S. (2025). Quantitative Assessment of Brain Tumor Epilepsy by AI Models. *INTERNATIONAL RESEARCH JOURNAL OF MULTIDISCIPLINARY SCOPE Учредители: Iquz Galaxy Publisher*, *6*(01), 1261-1279.

[18] Mejri, N., Ghorbel, E., Kacem, A., Chernakov, P., Foteinopoulou, N., & Aouada, D. (2025). When Unsupervised Domain Adaptation meets One-class Anomaly Detection: Addressing the Two-fold Unsupervised Curse by Leveraging Anomaly Scarcity. *arXiv preprint arXiv:2502.21022*.

[19] Jothi, S. F. E., Mahalakshmi, L., Elamurugan, P., Kaliyan, N., Swapna, J., & Rajeshwari, R. (2025, March). Falcon Optimized CNN for the Detection of Alzheimer's Disease in MRI Image. In *2025 3rd International Conference on Smart Systems for applications in Electrical Sciences (ICSSES)* (pp. 1-6). IEEE.

[20] Tsygankov, V., Kataev, A., Kudrin, R., Orlov, D., & Berestneva, O. (2025, September). Healthcare Intelligent Technologies for Real-Time Brain Anomaly Detection. In *Conference on Creativity in Intelligent Technologies and Data Science* (pp. 693-709). Cham: Springer Nature Switzerland.

[21] Vali, M., & Shaterzadeh-Yazdi, Z. (2025). Extending Fast Forward Quantum Optimization Algorithm with K-Means for Multi-Organ Medical Image Segmentation. *Available at SSRN 5755542*.

[22] Zhang, J., Han, F., Wang, X., Wu, F., Song, X., Liu, Q., ... & Chen, B. T. (2026). Brain metabolic imaging with 18 F-PET-CT and machine-learning clustering analysis reveal divergent metabolic phenotypes in patients with amyotrophic lateral sclerosis. *European Journal of Nuclear Medicine and Molecular Imaging*, *53*(3), 2018-2033.

[23] Murad, H. (2025). *A Hierarchical Hounsfield Unit Clustering-Based Model for Classification of Vertebral Fracture Acuity Using CT Scans Without the Need for Bone Scans* (Master's thesis, University of Haifa (Israel)).

[24] Nidamanuri, S., MASOOD, J., Asirvatham, D., BIN_ABAS, F., VEERAPPAN, K., Budiman, F., ... & Sugumaran, V. (2025). Predictive deep learning models to identify traumatic brain injuries using MRI data. *Journal of engineering science technology*, *2*.

[25] Vyas, A., Poriya, V., Mathariya, S., Jain, N., & Jain, M. (2025, May). Brain Tumor Detection in Magnetic Resonance Images Using Genetic Algorithms With Multiple Stages. In *International Conference on Recent Advancements and Modernisations in Sustainable Intelligent Technologies and Applications (RAMSITA 2025)* (pp. 789-802). Atlantis Press.