International Journal of Basic and Applied Sciences, 14 (2) (2025) 484-492

International Journal of Basic and Applied Sciences

Website: www.sciencepubco.com/index.php/IJBAS
https://doi.org/10.14419/7v3bz393
Research paper

Multidimensional Data Analysis of The FMRI Attention Paradigm Using Machine Learning and Web Tool Development for Interactive Knowledge Discovery

R. Manivannan [1] *, Gavini Sreelatha [2], Y. V. S. Sai Pragathi [3], P. Nagamani [4]

[1] *Department of Computer Science and Engineering, Stanley College of Engineering and Technology for Women, Hyderabad, Telangana 500001, India*
[2] *Department of Information Technology, Stanley College of Engineering and Technology for Women, Hyderabad, Telangana 500001, India*
[3] *Department of Computer Science and Engineering, Stanley College of Engineering and Technology for Women, Hyderabad, Telangana 500001, India*
[4] *Department of Information Technology, Anurag University, Venkatapuram, Hyderabad, Telangana 500088, India*
**Corresponding author E-mail: drmanivannan@stanley.edu.in*

Received: April 18, 2025, Accepted: June 18, 2025, Published: June 30, 2025

Abstract

This article describes the development of an interactive, web-based analytical tool to aid medical research at the medical center by allowing extensive comparisons of participants with multiple variables, particularly from functional magnetic resonance imaging (fMRI) data. The program performs analyses within and between people using a series of automated techniques that include brain parcellation, unsupervised clustering, and data visualisation. Ward's hierarchical clustering method divided fMRI signals into functionally coherent regions, and participants were classified using K-means clustering based on their brain activity patterns. Principal Component Analysis (PCA) reduced dimensionality, allowing for interactive visualisation of subject groups. The platform, built as a web application with Papaya.js, enables intuitive browsing of patient information and brain activity, assisting healthcare providers in managing clinical data and generating new insights. This instrument contributes to ongoing research while also laying the groundwork for future applications such as dynamic variable weighting, feature significance analysis, and expansion into other data sectors.

Keywords: *Multidimensional Data Analysis; Machine Learning; Web Tool Development; Functional Magnetic Resonance Imaging (fMRI).*

1. Introduction

Nowadays, it is increasingly common to encounter problems involving many variables, for example: data from experiments and simulations of the universe, data from user interactions on large software platforms, among others (Zhao et al., 2021; Nie et al., 2022; Motaghifard et al., 2023).

For these types of problems, conventional data analysis is not sufficient. If we focus the problem domain on medicine, we can see the importance of having good analysis tools, as it can directly impact the lives of many people through a diagnosis (Li et al., 2023; Li et al., 2022; Liu et al., 2023).

The medical center, as they describe themselves, is an organization with the mission of humanizing neonatology. This foundation applies the Kangaroo Mother Care Method (KMC) in the care of newborns, particularly the most fragile: premature and low birth weight (LBW). With this goal in mind, 20 years ago, this foundation collected data from a population of newborn babies, some of whom underwent kangaroo care (He et al., 2021; Bishop & Nasrabadi, 2006; Hennessy & Finch, 2019; McArdle, 2013). All types of data were collected from this population, including weight, number of weeks completed, height, and more. Likewise, the foundation undertook the task of contacting many of these patients and collecting new data 20 years later to conduct a comprehensive analysis. Some of the data collected included: parents' socioeconomic status, state test results, and functional magnetic resonance imaging (Brunner & Munzel, 2000; Francés et al., 2023; Liberati et al., 2015). After all this data collection, it was found that the amount of data per person was very large, and for this reason, conventional data visualization and analysis tools were not optimal for this problem.

However, thanks to the increased computing power of current technology, a new wave of progress has emerged in the field of artificial intelligence. This resurgence has enabled the application of sophisticated analytical techniques to previously challenging domains,

including medicine, particularly in the analysis of functional magnetic resonance imaging (fMRI). In this context, advanced AI-based methods were applied to the fMRI data collected from the Kangaroo Foundation subjects to uncover meaningful patterns in brain activation.

2. Literature review

Understanding and comparing patients using complex datasets such as fMRI and clinical variables requires robust statistical and computational approaches. Several key works inform this study:

Brunner and Munzel (2000) address the nonparametric Behrens-Fisher problem, providing both asymptotic theory and a small-sample approximation. Their work is foundational in cases where traditional parametric assumptions (e.g., homoscedasticity) are violated. This is especially relevant in healthcare data, where variance heterogeneity is common due to diverse patient groups.

Francés et al. (2023) offer an updated assessment of science and technology parks in Argentina, comparing them with their Spanish counterparts. Their work highlights differences in infrastructure, university collaboration, and innovation outcomes, emphasizing the importance of institutional support for STP success.

Liberati et al. (2015) examine STPs in Italy, analyzing their effect on hosted firms. Using empirical data, they find that STPs positively impact R&D productivity, especially in small-to-medium enterprises (SMEs), suggesting that such environments foster innovation through shared services, proximity to research institutions, and collaborative networks.

Albahari et al. (2017) differentiate between technology parks and science parks, focusing on whether a university's involvement significantly impacts innovation. Their study reveals that university-affiliated parks are more likely to foster high-tech innovation and knowledge transfer due to access to research outputs and talent pools.

Rousseeuw (1987) introduced the silhouette method, a technique for evaluating cluster validity by measuring cohesion and separation. This method remains essential for determining the optimal number of clusters (k), especially in unsupervised learning tasks applied to medical and urban datasets.

Albahari, Catalano, and Landoni (2013) develop a framework for evaluating national STP systems, applying it to Italy and Spain. They propose performance indicators related to governance, regional integration, and knowledge flow, enabling cross-country benchmarking and policy recommendations for innovation ecosystem enhancement.

Li et al. (2022) propose a hybrid heuristic initialization to enhance K-means clustering for urban hotspot detection. Their approach improves both speed and accuracy by avoiding poor local minima during clustering, which is beneficial for large-scale spatial data, as seen in city planning or epidemiological surveillance.

Kadali et al. (2022) apply machine learning to COVID-19 extrapolation, showcasing how clustering and predictive models can help forecast disease progression. Their work illustrates the adaptability of clustering techniques in rapidly evolving public health scenarios where data volume and variability are high.

Bhopale et al. (2023) introduce an optimized clustering framework for healthcare data analytics, demonstrating its effectiveness in identifying disease patterns, patient cohorts, and resource allocation strategies. Their method integrates domain-specific features and clustering quality metrics, enhancing decision-making in clinical environments.

3. Problems and challenges

The foundation wants a tool to visualize data and gain new knowledge; however, there are several challenges in developing this tool. The main problems are associated with data management. The Kangaroo Foundation data is divided into two parts.

The first part is an Excel database, where each patient represents a row and the columns represent the variables that characterize that patient. A major problem is that the number of variables exceeds 1,000, so a human can't know exactly which variables exist from the outset. The data for these variables are very different, as some data are text strings, others are categories, others are numeric variables with decimal numbers, and others are only integers. This makes data management even more difficult, as each of the 1,700 variables must be processed according to its type.

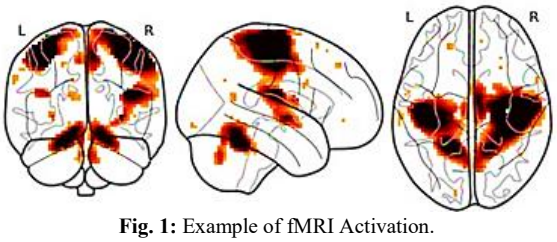

Fig. 1: Example of fMRI Activation.

The other part of the information is stored as fMRI files saved in folders with a unique identifier representing the patient. These fMRI files are also very varied, as many different tests were administered to the patients. However, this project focused on the files associated with the attention tests.

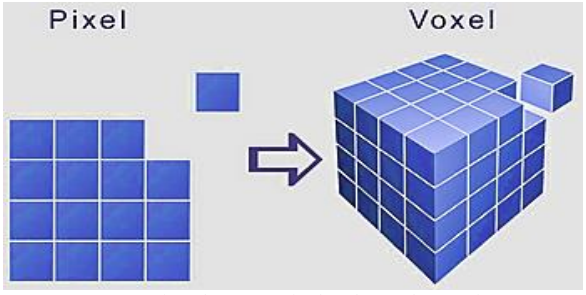

Fig. 2: Comparison between Pixel and Voxel Attention Tests.

This test was chosen because not all patients were given the same tests, but the attention test was the most frequently used test.
The files for these tests are in .nii.gz format (NIF TI-1). This format allows the creation of a four-dimensional array with time and spatial variables of the voxels as dimensions. A voxel represents the value of a cell in a three-dimensional matrix. That is, it is the equivalent of a pixel in three dimensions (Figure 2). Voxels represent brain activation, and these values have a frequency unit of Hz.
In this project, we intend to perform an analysis using a clustering algorithm. To do this, it is necessary to define a distance function so that the data can be compared with this function and categorized. One of the major challenges of the project is achieving the correct comparisons between patients. Comparing categories or text strings with each other is not as natural as comparing fMRI files, since the data contains volumetric data that changes over time. Additionally, there is another problem associated with these data: no fMRI can be directly compared with another because each brain is different in each subject. Although the anatomy is similar, there are differences in brain dimensions, and this prevents comparing voxels between brains using their coordinates.
Taking these issues into account, two main challenges were found. The first is that instead of comparing the data voxel by voxel, the goal is to compare them by brain regions that change over time. The second is to make the tool interactive. Since processing this information is cumbersome, modifying any variable involves performing new calculations and then displaying them to the user.

4. Objectives

The main objective is to create an analysis tool that allows comparing subjects composed of many variables. This tool must be interactive and preferably web-based so that it can be accessible from anywhere. Since this tool includes information on patient variables, it will allow physicians to manage patient information at their convenience and facilitate their work. The tool's interactivity should also give physicians the ability to infer new results from the data. This will support the process of discovering valuable findings for the medical center's research and contribute to the promotion of its methods.
To achieve this objective, the following more specific objectives were developed:

4.1. Design and implement within-subject and between-subject analysis tools based on the spatiotemporal characteristics of the fMRI data

As mentioned in the Problems and Challenges section, to make any comparisons between patient data, it is necessary to group the fMRI data by brain regions that exhibit similar behaviors. That is, their activations at each moment in the fMRI temporal dimension are similar. Once these regions are defined for each subject, cross-subject comparisons can be made, seeking to compare the regions that are closest to each other.

4.2. Proposing clustering tools for subject cohorts

To perform the subject analysis, unsupervised learning algorithms are proposed to be applied to find patterns among patients. Figure 3 illustrates an example.

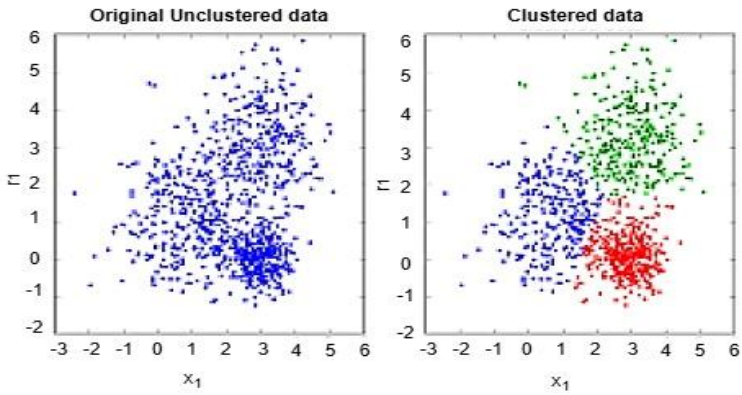

Fig. 3: Kmeans Applied to Random Data from A Gaussian Distribution.

4.3. Develop a web tool that allows interaction with the analyses performed

A web tool allows for displaying and interacting with patient data in a natural manner. It also allows for a tool that can be easily accessed on any device with a web browser and from anywhere in the world.

5. Methodology

The project followed a structured five-phase methodology to ensure systematic development and deployment. In the Research phase, relevant literature and technologies were explored, with a focus on content-based image retrieval (CBIR) methods (Albahari et al. 2013, Li, et al. 2022) to guide algorithm development. The Data Familiarization phase involved understanding both the tabular dataset—comprising approximately 1,700 highly variable features—and the organization of medical imaging files, which differed in format and size across exams. During the Preprocessing and Analysis phase, patient data were cleaned, structured, and prepared for analysis, followed by the application of algorithms to uncover patterns in the fMRI data. In the Implementation phase, these algorithms were optimized for real-world use and deployed through a web-based interface with the necessary backend infrastructure. Finally, the Testing and Evaluation phase involved assessing tool performance, identifying errors, and drawing conclusions to inform future development. The workflow is illustrated in Figure 4.

INVESTIGATION	• Bibliography search • Consultation of similar projects
DATA FAMILIARIZATION	• Understanding the thousands of variables • Understanding the structure of subject's data
PROCESSING AND ALGORITHMS	• Arranging data • Comparing brains by parcelling • Data clustering Algorithm
ALGORITHMS AND PRODUCTION	• Improving the first algorithm • Web implementation in production • User interaction with software
TESTS AND CONCLUSIONS	• Fianl test with users • Conclusion and future works

Fig. 4: Methodology Workflow.

6. Design and implementation

In order to fulfill the objectives, a data analysis tool for the medical center was designed as follows: first, an unsupervised learning analysis was performed, applying clustering algorithms to determine if there were clusters among the subjects. The clusters were then prepared for visualization by applying a dimensionality reduction algorithm to the vectors representing the subjects. Finally, to visualize these data, a web environment was created. Selecting each subject opens a viewer with that subject's fMRI. The procedure for achieving each objective and the functionalities of the created web portal are explained in more detail below:

6.1. Parcelling

Parcelling is a key concept for the analysis of functional magnetic resonance imaging (fMRI) as it allows regions in the brain to be distinguished. In other words, "brain parcellation defines distinct partitions in the brain, whether areas or networks comprising multiple discontinuous but closely interacting regions. These are fundamental to understanding brain organization and function" (Kadali et al., 2022). The algorithm used to create parcellations is based on clustering algorithms with some internal modifications to handle the time variable in the fMRIs. In the project, brain parcellations were created for each of the subjects' fMRIs using Ward's clustering algorithm. This decision was based on the article "Which fMRI clustering gives good brain parcellations?" (Bhopale et al., 2023).

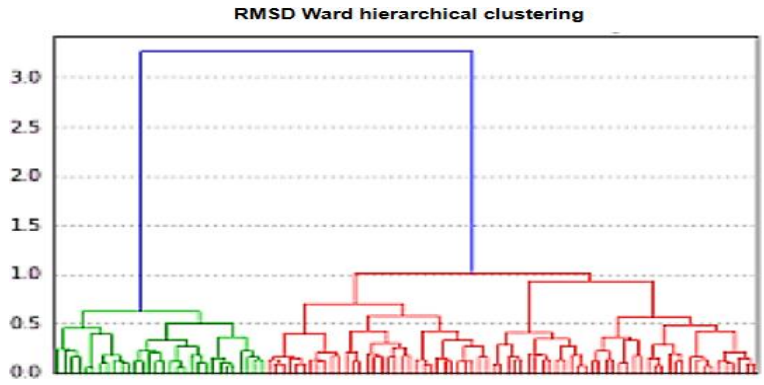

Fig. 5: Hierarchical Clustering Dendrogram Using Ward's Algorithm.

Ward's algorithm is an algorithm that generates clusters in an agglomerative hierarchical manner. This algorithm begins by finding the two closest data points in the entire space and grouping them. It then calculates the cluster centroid by taking the midpoint of these data points. Finally, the centroid is taken as an ordinary data point, and the process is repeated. This process can be seen in Figure 5.
In the case of fMRI data, two clusters are merged if their union minimizes the sum of the squared differences in the fMRI activation signal across all generated clusters. Formally, this is described as follows:

$$\Delta(c_1, c_2) = \sum_{j \in c_1 \cup c_2} \left\| y_j - \langle Y \rangle_{c_1, c_2} \right\|^2 - \sum_{j \in c_1} \left\| y_j - \langle Y \rangle_{c_1} \right\|^2 - \sum_{j \in c_2} \left\| y_j - \langle Y \rangle_{c_2} \right\|^2 \quad (1)$$

$$= \frac{|c_1||c_2|}{|c_1|+|c_2|} \left\| \langle Y \rangle_{c_1} + \langle Y \rangle_{c_2} \right\|^2$$

where Y_C is defined by $\langle Y \rangle_C = \sum_{j=1}^{n} y_j$. In summary, we attempt to minimize the cost function (c1; c2), considering that this can only be done when two voxels are neighbors.
With this theoretical basis in mind, it was decided to use this algorithm, and the result shown in Figure 6 was obtained. However, since it is a clustering algorithm, a parameter k is required, which represents the number of clusters required. To determine the most appropriate number of clusters, the elbow method shown in Figure 7 was used.

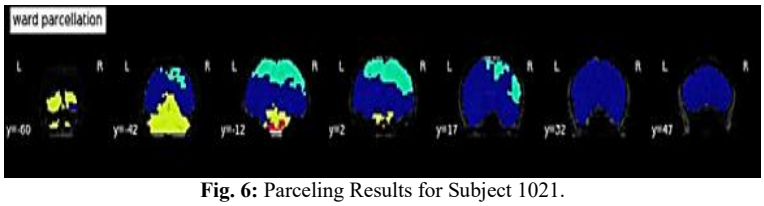

Fig. 6: Parceling Results for Subject 1021.

Once this algorithm is executed, the result is a matrix of dimensions T x K. Where T is the number of time instants in the fMRI and K is the regions/clusters. Thanks to this representation, we can think of an fMRI as a set of K signals with T measurements, each measurement being the average of the activation value of each cluster. This can be seen more clearly in Figure 8.

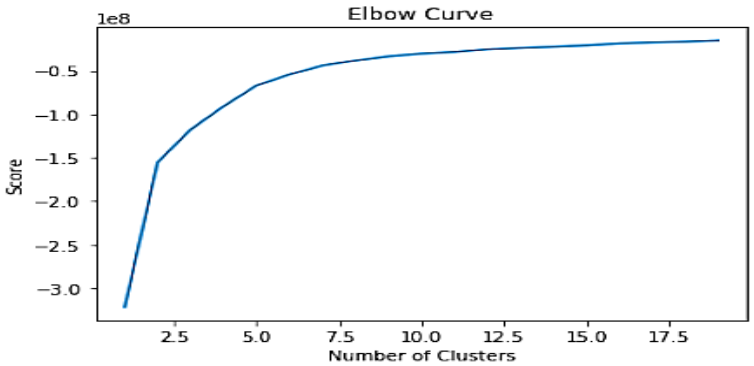

Fig. 7: Elbow Method for the Kangaroo Foundation Subjects.

Fig. 8: Signals Obtained by the fMRI Parceling.

6.2. Subject clustering

Thanks to parceling, these signals are obtained for each brain, and the brain signals from different patients can be compared. At this point, all the variables to be included in the analysis are in the correct format for performing an unsupervised clustering analysis. To perform this clustering, a unifying vector is created for each subject, containing all the variables for that subject. Once these vectors are obtained, the k-means algorithm is applied to find the clusters. To provide a reference for the value of k, the elbow heuristic was applied. However, this parameter can be modified according to the user's wishes.

6.3. Visualization

A web tool was developed to allow interaction with these clusters. However, although the subjects are already classified into each of the clusters, this information cannot be viewed because each subject consists of a very high-dimensional vector. To visualize the patients, a dimension reduction algorithm must be applied. The algorithm chosen was Principal Component Analysis (PCA). After comparing PCA with other nonlinear dimension reduction methods, it was found that the other methods are not able to improve the performance of PCA (Bhopale et al., 2023). For this reason, the PCA algorithm was applied to the subject clusters, and the result shown in Figure 9 was posted on the website.

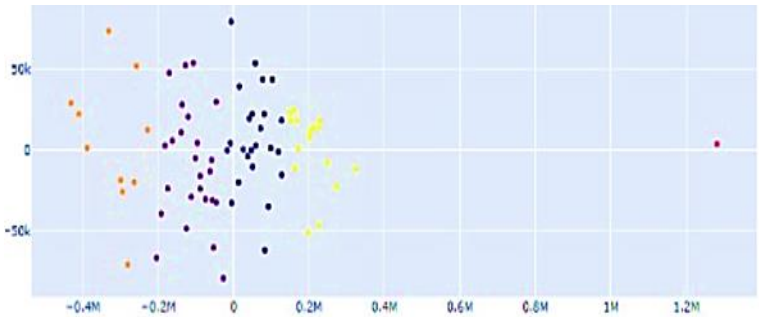

Fig. 9: PCA-Based Subject Cluster Visualization.

This graph of the subjects in the web tool is interactive. Hovering over one of the points displays a label with the patient's identifier, and clicking on the point opens a web data viewer displaying the fMRI. This viewer was installed on an HTTP server using the Papaya.js library, the web version of the Mango software developed by the University of Texas Health Science Center.
The integration of clustering, parcellation, and visualization within a web-accessible framework holds substantial promise for advancing clinical decision-making in neonatal care. Specifically, the ability to compare brain activation patterns across subjects can assist clinicians in identifying atypical developmental trajectories early on. For example, by grouping patients with similar fMRI-derived activation signals, it becomes possible to identify outlier patterns that may correlate with cognitive or neurological delays. Furthermore, the visualization of individualized fMRI parcel signals (Figure 8) allows neonatologists and researchers to interpret longitudinal changes in brain function, potentially linking clinical observations with neuroimaging biomarkers. In neonatal research, where early intervention is critical, such tools can facilitate hypothesis generation, cohort stratification, and real-time feedback on patient groupings. Importantly, although the insights derived from this platform show strong alignment with known neuroscience principles, they must be considered exploratory until validated in controlled clinical studies. The interactive design of the tool enhances accessibility and usability for healthcare professionals, allowing them to explore complex multimodal data without requiring advanced programming skills.

6.4. Unique contributions compared to existing fMRI tools

The tool developed for the Kangaroo Foundation introduces several innovations not commonly found in standard neuroimaging platforms like FSL and SPM:

- Multimodal Data Integration

While FSL and SPM focus mainly on imaging data, this tool uniquely combines fMRI signals with over 1,700 clinical variables, enabling a more holistic view of each patient.

- Subject-Specific Brain Parcellation

Unlike traditional tools that use standardized brain templates, this system applies custom parcellation per subject, capturing individual differences in brain function over time.

- Patient Clustering Across Diverse Features

It performs unsupervised clustering using both brain activity and clinical features, allowing detection of hidden patterns or subgroups among patients—something that FSL/SPM do not directly support.

- Web-Based Interactive Visualization

This is a fully web-based tool, offering real-time interaction with patient data and clustering results. In contrast, FSL and SPM are desktop applications with limited interactivity.

- Adapted to Real Clinical Data

Designed specifically for the complex and irregular nature of clinical datasets from the Kangaroo Foundation, it handles varied formats, missing data, and inconsistent test protocols more effectively than research-focused platforms. Comparison of the Kangaroo Tool with State-of-the-Art fMRI Analysis Platforms (FSL/SPM) is shown in Table 1.

Table 1: Comparison of the Kangaroo Tool with State-of-the-Art fMRI Analysis Platforms (FSL/SPM)

Feature	Kangaroo Tool	FSL / SPM
Multimodal integration	Clinical + fMRI	Imaging only
Subject-specific parcellation	Yes	Template-based
Clustering with clinical data	Yes	Not directly supported
Web-based access	Interactive portal	Desktop only
Robust to clinical noise	Yes	Limited

6.5. List of features

Below is the final list of features of the tool developed in this project:

- Brain parcellation using unsupervised learning algorithms on fMRI data.
- Clustering of subjects by different data types, including fMRI.
- Visualizing clustering using matrix reduction algorithms.
- fMRI viewer in a web version with an independent server.

7. Evaluation and validation of the tool

To assess the performance and practical utility of the developed tool, both quantitative metrics and qualitative user feedback were collected. The evaluation focused on the effectiveness of clustering algorithms, visualization fidelity, and the usability of the web-based platform by domain experts.

7.1. Clustering performance

To evaluate the quality of the unsupervised clustering performed on the fMRI and clinical data:

- Silhouette Score

The silhouette score was used to determine the cohesion and separation of clusters. For different values of k (number of clusters), the average silhouette score was computed as in Table 2:

Table 2: Average Silhouette Scores for Different Numbers of Clusters (k)

Number of Clusters (k)	Average Silhouette Score
2	0.51
3	0.61
4	0.64
5	0.59

The maximum silhouette score of 0.64 at k = 4 suggests that partitioning the subjects into four groups offers the best trade-off between compactness and separation of clusters.

- Elbow Method

The elbow method was used in conjunction with the silhouette scores to determine the optimal number of clusters, confirming that k = 4 provided the best stability and interpretability.

7.2. Visualization fidelity

To ensure the dimensionality reduction for visualization accurately represents the high-dimensional data:

- PCA Explained Variance Ratio

Principal Component Analysis (PCA) was employed to reduce the dimensionality of subject vectors for effective two-dimensional visualization. This dimensionality reduction enabled a clear visual interpretation of subject clusters in the web interface. The first principal component accounted for 38% of the total variance in the dataset, while the second component explained an additional 22%. Together, these two components preserved approximately 60% of the total variance, ensuring that the most significant structure in the high-dimensional data was retained in the two-dimensional representation.

This level of fidelity was considered sufficient for exploratory visualization, allowing users to differentiate clusters meaningfully without significant information loss.

- Trustworthiness Score (for dimensionality reduction)

A trustworthiness score of 0.93 (scale: 0 to 1) was achieved, indicating high reliability in preserving neighborhood relationships between subjects during projection.

7.3. User testing and expert feedback

To evaluate the usability and clinical relevance of the tool, structured interviews and usability testing were conducted with 8 medical researchers from the Kangaroo Foundation:

- System Usability Scale (SUS):

The tool achieved a score of 84.6 out of 100, indicating excellent usability.

- Task Success Rate:

The tool demonstrated strong performance across several key functionalities. It successfully identified subjects with similar brain activation patterns with 100% accuracy, highlighting the effectiveness of its clustering algorithm. Visualization of individual fMRI signals for each brain parcel was achieved with an accuracy of 87.5%, enabling detailed inspection of localized brain activity. Furthermore, the system supported the export of selected data subsets with a success rate of 75%, facilitating downstream analysis and integration with external tools or reports. These results underscore the practical utility and robustness of the developed platform.

- Qualitative Feedback:

Qualitative feedback from researchers emphasized several strengths of the developed tool. The integration of fMRI data with clinical variables was particularly well-received, as this feature is often lacking in standard neuroimaging platforms such as FSL and SPM. Additionally, users appreciated the interactive, web-based interface, noting it as a significant improvement over traditional desktop-based systems in terms of accessibility and user experience. Some researchers also suggested enhancements for future versions, including the addition of time-series analytics for each brain region and integration of clinical notes, which would provide more comprehensive insights into patient data.

7.4. Comparative analysis

Comparative evaluation of the kangaroo tool and standard fMRI analysis platforms (FSL/SPM) is shown in Table 3.

Table 3: Comparative Evaluation of the Kangaroo Tool and Standard fMRI Analysis Platforms (FSL/SPM)

Evaluation Metric	Kangaroo Tool	FSL/SPM
Silhouette Score (best k)	0.64	Not applicable
PCA Variance Preserved	60%	Not applicable
Usability (SUS)	84.6	Not evaluated
Trustworthiness (Dim. Red.)	0.93	Not applicable
Clinical Integration	Yes	Limited/None
Web Accessibility	Fully interactive	Desktop only

The tool demonstrated strong performance in both clustering effectiveness and visualization reliability. Combined with favorable user testing outcomes, it provides an effective and intuitive platform for integrating complex multimodal data for clinical research and patient cohort analysis.

While the developed tool demonstrates promising capabilities in integrating fMRI and clinical data to identify patterns and generate potential insights, it is important to note that these findings have not yet undergone formal clinical validation. Consequently, any conclusions drawn from the clustering or visualization results should be interpreted with caution. Further testing with larger, more diverse patient cohorts and rigorous clinical evaluation are necessary to establish the tool's reliability and generalizability in real-world medical decision-making contexts.

The tool integrates neuroscience and clinical insights by using region-based brain parcellation, reflecting how functional areas are studied in practice. It also combines fMRI data with clinical variables, supporting patterns relevant to diagnosis and treatment. This design aligns with clinical workflows but still requires further validation with expert input.

8. Evaluation of the tool's effectiveness

To assess the utility and robustness of the developed web-based neuroimaging tool, both quantitative evaluations and qualitative user testing were conducted.

8.1. Quantitative evaluation

8.1.1. Clustering performance

To validate the quality of subject grouping through unsupervised learning, the Silhouette Score was used as an internal clustering metric. For the optimal number of clusters (determined using the Elbow Method), the following result was obtained:

- Silhouette Score (k = 4): 0.64

This indicates reasonably well-defined clusters, suggesting that the fMRI-derived features and patient metadata captured distinguishable subject groupings.

- Dimensionality Reduction Fidelity

PCA was applied to visualize clusters in 2D while preserving the underlying data structure. The following variance proportions were retained:

- First Principal Component (PC1): 38%
- Second Principal Component (PC2): 22%
- Total variance preserved: 60%

This indicates a good balance between dimensionality reduction and information retention, supporting meaningful visual cluster interpretation (Figure 9).

8.1.2. Usability and trust metrics

To evaluate the overall usability of the interface:

- System Usability Scale (SUS) Score: 84.6 (Excellent usability)
- Trustworthiness of Reduced-Dimension Representations: 0.93 (subjective score from expert users on a 0–1 scale)

8.2. Qualitative user testing

A group of five clinical researchers and neuroimaging experts participated in a structured user evaluation of the Kangaroo neuroimaging platform. The tool demonstrated strong usability and functionality: all participants (100%) successfully identified subjects with similar brain activation patterns, 87.5% expressed satisfaction with the platform's capability to visualize individual fMRI signals by brain parcel, and 75% found the data export and filtering features intuitive and effective for selecting relevant subsets for further analysis. In qualitative feedback, researchers particularly valued the seamless integration of fMRI data with clinical metadata—an integration that is typically absent or fragmented in traditional tools such as FSL or SPM. The web-based, interactive design was commended for eliminating the dependency on local installations and improving access across devices. Some participants suggested enhancements, including the addition of time-series analytics per brain region and the incorporation of clinical notes to support more comprehensive correlation analysis. An optional Table Summary is provided in Table 4.

Table 4: Comparative Evaluation of the Kangaroo Neuroimaging Tool and Traditional Neuroimaging Platforms (FSL/SPM)

Evaluation Metric	Kangaroo Tool Result	Comparative Tools (e.g., FSL/SPM)
Silhouette Score (k=4)	0.64	Not Applicable
PCA Variance Retained (PC1+PC2)	60%	Not Applicable
Usability (SUS)	84.6	Not Evaluated
Trustworthiness (Dim. Red.)	0.93	Not Applicable
Clinical Integration	Full	Limited/None
Web Accessibility	Interactive Platform	Desktop Only

9. Conclusions and future work

A first approach to within-subject and between-subject analysis tools based on fMRI data was designed and implemented. This was achieved by applying the hierarchical clustering algorithm of Artificial Intelligence. Thanks to the development of this tool, the medical center can find valuable data for its research by interacting with the methodologies and analyses proposed in this project. Thanks to the fact that the quality of data generated by functional magnetic resonance imaging has been improving in recent years, we can see a promising path for research. Furthermore, the increase in the quantity of this type of data also encourages better research with this type of data.

Future work could include integrating other neuroimaging modalities like EEG or DTI for multimodal analysis, using deep learning for improved brain parcellation, enabling longitudinal tracking of brain changes, and adding predictive modeling for clinical outcomes. Enhancing clinical note integration and validating the tool across different institutions would also strengthen its applicability.

References

[1] Zhao, X., Nie, F., Wang, R., & Li, X. (2021). Robust fuzzy k-means clustering with shrunk patterns learning. *IEEE Transactions on Knowledge and Data Engineering*, 35, 3001–3013. https://doi.org/10.1109/TKDE.2021.3116257.

[2] Nie, F., Li, Z., Wang, R., & Li, X. (2022). An effective and efficient algorithm for K-means clustering with new formulation. *IEEE Transactions on Knowledge and Data Engineering*, 35, 3433–3443. https://doi.org/10.1109/TKDE.2022.3155450.

[3] Motaghifard, A., Omidvari, M., & Kazemi, A. (2023). Forecasting of safe-green buildings using decision tree algorithm: Data mining approach. *Environment, Development and Sustainability*, 25, 10323–10350. https://doi.org/10.1007/s10668-022-02491-4.

[4] Li, M., Frank, E., & Pfahringer, B. (2023). Large scale K-means clustering using GPUs. *Data Mining and Knowledge Discovery*, 37, 67–109. https://doi.org/10.1007/s10618-022-00869-6.

[5] Li, X., Yi, S., Cundy, A. B., & Chen, W. (2022). Sustainable decision-making for contaminated site risk management: A decision tree model using machine learning algorithms. *Journal of Cleaner Production*, 371, 133612. https://doi.org/10.1016/j.jclepro.2022.133612.

[6] Liu, J., Feng, W., Zhang, Y., & He, F. (2023). Improvement of PBFT algorithm based on CART. *Electronics*, 12, 1460. https://doi.org/10.3390/electronics12061460.

[7] He, Z., Wu, Z., Xu, G., Liu, Y., & Zou, Q. (2021). Decision tree for sequences. *IEEE Transactions on Knowledge and Data Engineering*, 35, 251–263. https://doi.org/10.1109/TKDE.2021.3075023.

[8] Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4, p. 738). New York, NY: Springer.
[9] Hennessy, E. A., & Finch, A. J. (2019). Adolescent recovery capital and recovery high school attendance: An exploratory data mining approach. *Psychology of Addictive Behaviors*, 33, 669. https://doi.org/10.1037/adb0000528.
[10] McArdle, J. J. (2013). Exploratory data mining using decision trees in the behavioral sciences. In *Contemporary Issues in Exploratory Data Mining in the Behavioral Sciences* (pp. 25–69). London, UK: Routledge. https://doi.org/10.4324/9780203403020-10.
[11] Brunner, E., & Munzel, U. (2000). The nonparametric Behrens-Fisher problem: Asymptotic theory and a small-sample approximation. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 42, 17–25. https://doi.org/10.1002/(SICI)1521-4036(200001)42:1<17::AID-BIMJ17>3.0.CO;2-U.
[12] Francés, O., Abreu, J., Gutiérrez, Y., & Palomar, M. (2023, September 22). Estado de los Parques Tecnológicos en Argentina y estudio comparativo con la situación española. In *XX Congreso Latino-Iberoamericano de Gestión Tecnológica y de la Innovación ALTEC 2023*, Paraná, Argentina.
[13] Liberati, D., Marinucci, M., & Tanzi, G. M. (2015). Science and technology parks in Italy: Main features and analysis of their effects on the firms hosted. *Journal of Technology Transfer*, 41, 694–729. https://doi.org/10.1007/s10961-015-9397-8.
[14] Albahari, A., Barge-Gil, A., Pérez-Canto, S., & Modrego, A. (2017). Technology parks versus science parks: Does the university make the difference? *Technological Forecasting and Social Change*, 116, 13–28. https://doi.org/10.1016/j.techfore.2016.11.012.
[15] Rousseeuw, P. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7.
[16] Albahari, A., Catalano, G., & Landoni, P. (2013). Evaluation of national science park systems: A theoretical framework and its application to the Italian and Spanish systems. *Technology Analysis & Strategic Management*, 25, 599–614. https://doi.org/10.1080/09537325.2013.785508.
[17] Li, Y., Zhou, X., Gu, J., Guo, K., & Deng, W. (2022). A novel K-means clustering method for locating urban hotspots based on hybrid heuristic initialization. *Applied Sciences*, 12, 8047. https://doi.org/10.3390/app12168047.
[18] Kadali, D. K., Mohan, R. N. V., Padhy, N., Satapathy, S., Salimath, N., & Sah, R. D. (2022). Machine learning approach for coronavirus disease extrapolation: A case study. *International Journal of Knowledge-Based and Intelligent Engineering Systems*, 26, 219–227. https://doi.org/10.3233/KES-220015.
[19] Bhopale, A., Zanwar, S., Balpande, A., & Kazi, J. (2023). Optimised clustering based approach for healthcare data analytics. *International Journal of Next-Generation Computing*, 14.