



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078.

Privacy preserving federated machine learning and blockchaining for reduced cyber risks in a world of distributed healthcare



Deliverable D5.3
“Federated multi-OMICS system-medicine based clustering software available for download”

Work Package WP5
“Unsupervised Federated Machine Learning”

Disclaimer

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078. Any dissemination of results reflects only the author's view and the European Commission is not responsible for any use that may be made of the information it contains.

Copyright message

© FeatureCloud Consortium, 2023

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both. Reproduction is authorised provided the source is acknowledged.

Document information

Grant Agreement Number: 826078			Acronym: FeatureCloud	
Full title	Privacy preserving federated machine learning and blockchaining for reduced cyber risks in a world of distributed healthcare			
Topic	Toolkit for assessing and reducing cyber risks in hospitals and care centres to protect privacy/data/infrastructures			
Funding scheme	RIA - Research and Innovation action			
Start Date	1 January 2019	Duration	60 months	
Project URL	https://featurecloud.eu/			
EU Project Officer	Christos Maramis, Health and Digital Executive Agency (HaDEA)			
Project Coordinator	Jan Baumbach, University of Hamburg (UHAM)			
Deliverable	D5.3 - Federated multi-OMICS system-medicine based clustering software available for download			
Work Package	WP5 - Unsupervised Federated Machine Learning			
Date of Delivery	Contractual	30/06/2023 (M54)		Actual31/07/2023 (M55)
Nature	Report	Dissemination Level	Public	
Lead Beneficiary	04 SDU			
Responsible Author(s)	Richard Röttger, Lucas Alexander Damberg Torp Dyssel, Tomás Bordoy Garcia-Carpintero, Nicolai Dinh Khang Truong (all from SDU)			
Keywords	Multi-OMICS, Multiple Factor Analysis, Principal Component Analysis, Clustering, k-means			

Table of Contents

1	Table of acronyms and definitions.....	5
2	Objectives of the deliverable based on the Description of Action (DoA)	6
3	Executive Summary	7
4	Introduction (Challenge).....	8
4.1	Multi-OMICS analysis	8
4.2	The FeatureCloud Setting.....	8
4.3	Approaches to Multi-OMICS	8
5	Methodology	9
5.1	Multiple Factor Analysis	9
5.1.1	Federated Multiple Factor Analysis	10
5.2	Input Files	13
5.2.1	Configuration	13
5.2.2	Results folder.....	13
6	Results.....	14
6.1	Standalone app.....	14
6.2	Workflow application	14
7	Open issues.....	18
8	Conclusion	18
9	References	19
10	Other supporting documents / figures / tables (if applicable).....	20

1 Table of acronyms and definitions

A	astrocytomas
concentris	concentris research management gmbh
D	Deliverable
DoA	Description of the Action
GBM	glioblastomas
gfPCA	global federated Principal Component Analysis
HaDEA	Health and Digital Executive Agency
MFA	Multiple Factor Analysis
MS	Milestone
OMICS	“OMICS” is a rapidly evolving, multi-disciplinary biological research field that encompasses genomics, epigenomics, transcriptomics, proteomics, and metabolomics.
Patients	In this deliverable, we use the term “patients” for all research subjects. In FeatureCloud, we will focus on patients, as this is already the most vulnerable case scenario and this is where most primary data is available to us. Admittedly, some research subjects participate in clinical trials but not as patients but as healthy individuals, usually on a voluntary basis and are therefore not dependent on the physicians who care for them. Thus, to increase readability, we simply refer to them as “patients”.
PCA	Principal Component Analysis
PM	Person month
mRNA	messenger ribonucleic acid
O	oligodendrogliomas
OA	mixed oligoastrocytomas
SDU	Syddansk Universitet
sfPCA	separated federated Principal Component Analysis
UHAM	Universität Hamburg
WHO	World Health Organization
WP	Work Package

2 Objectives of the deliverable based on the Description of Action (DoA)

The FeatureCloud Work Package 5 (WP5) develops federated algorithms for unsupervised learning tasks, primarily for, but not limited to, the medical domain.

In the previous deliverable (D5.1), we have presented unsupervised preprocessing methods, including normalization, standardizations, and lower-dimensionality projection methods (principal component analysis) as well as in D5.2 general-purpose clustering methods, based on gaussian mixture models and k-means based clustering algorithms, including a sophisticated visualization application. The objective of this deliverable (D5.3) is to extend the previous methods to fulfill the task “**Federated multi-OMICS system-medicine based clustering**” (Task 2).

Here, we strive to build specialized clustering methods which allow the separation of patients based on multiple OMICS types simultaneously, and as such, deliver a more holistic view of the patients and enable a system medical interpretation of the groups, based on the driving factors specifying the *de novo* discovered patients’ groups. It is important to note that in this deliverable, we are evaluating the mere performance of our implementation and demonstrate that we can recover the same information as with a non-federated version. The deduction of novel biological insights, as well as the medical interpretation of results, is beyond the scope of the deliverable. In fulfillment of the deliverable, the following milestones (MS) have been also been reached within the last reporting period of the FeatureCloud project:

- **MS36:** Implementation of federated multi-OMICS system-medicine based clustering software
- **MS37:** Testing federated multi-OMICS system-medicine based clustering completed

3 Executive Summary

With current developments of wet lab technologies, researchers are increasingly able to collect data rapidly from multiple so-called OMICS data from the same sample and thus receiving a more complete view of the state of the sample and ongoing processes. “OMICS” is a rapidly evolving, multi-disciplinary biological research field that encompasses genomics, epigenomics, transcriptomics, proteomics, and metabolomics. The goal is to understand aberrations and mechanisms leading to disease, for instance, by accounting for all interplaying factors, from genetics to proteomics. The data integration of these different modes has been a long-standing problem in research, and a multitude of different approaches has been developed.

After a careful review of the existing literature, we are presenting a federated version of the Multi-Factor Analysis (MFA) [1], which is fully integrated into the FeatureCloud ecosystem [2] and available as an app for download¹. MFA is based on the projection of multiple OMICS modalities into a shared latent variable space with unified variance, which is subsequently decomposed into singular values to identify driving factors of the dataset.

These methods were tested on real-world datasets such as brain tumors [3], and we were able to demonstrate that the results correspond to those obtained by the non-federated version. Furthermore, the full integration into the FeatureCloud platform allows the factor analysis to form the basis for subsequent clustering, directly feeding into our clustering applications² and visualization apps³. With this integration, FeatureCloud supports a full-fledged multi-OMICS analysis for the de novo phenotyping of patients based on an arbitrary number of OMICS modes.

¹ <https://featurecloud.ai/app/multiple-factor-analysis>

² <https://featurecloud.ai/app/fc-federated-kmeans>

³ <https://featurecloud.ai/app/fc-cluster-visualization-app>

4 Introduction (Challenge)

4.1 Multi-OMICS analysis

As previously mentioned, multi-OMICS analysis (i.e., the analysis of numerous modes of data per sample simultaneously) has been a long standing problem in data integration for biomedical data. As described in [4]: “Analysis of multi-OMICS data along with clinical information has taken the front seat in deriving useful insights into the cellular functions. Integration of multi-OMICS data providing information on biomolecules from different layers seems to be promising to understand complex biology systematically and holistically.”

With multi-OMICS analysis numerous questions can be answered: From Biomarker prediction and disease insights to disease subtyping. The most relevant methods considered for this deliverable are those allowing for disease subtyping. This poses an unsupervised learning problem, with the goal of grouping similar patients based on multiple OMICS modes together.

4.2 The FeatureCloud Setting

The federated setting considered in this deliverable is as follows: The patient data is distributed over several sites:

- Each site holds data on different patients
- Each site has performed the same data collection, i.e., each site has the same features for every patient

We do not consider partially overlapping feature space between the sites, nor different preprocessing approaches. That means all sites have agreed on an exhaustive and detailed data processing protocol before the federated study and provisioned the data accordingly to the FeatureCloud platform. Furthermore, there are no missing values and quality control ensures the absence of outliers or technical measurement errors. These limitations can be relaxed with additional steps of federated data homogenization and appropriate adjustments to the study protocol, but these are beyond the scope of the deliverable.

4.3 Approaches to Multi-OMICS

In general, unsupervised approaches to a multi-OMICS analysis for disease subtyping can be differentiated into following classes of methods [4]:

- **Similarity-based approaches:** These methods establish pairwise similarity between the patients using all available OMICS data. Once a pairwise similarity is established, the data can be clustered using a common clustering tool that works on a similarity matrix for the identification of the final disease subtypes.
- **Bayesian approaches:** Those approaches seek to utilize a joint latent variable model for modeling the associations between different factors while reducing the dimensionality. The parameterized models are trained in a Bayesian approach, maximizing the likelihood of the observed data given the model. Later, the latent spaces can be clustered and interpreted.
- **Multivariate approaches:** These approaches seek to project data from multiple-OMICS data into a lower-dimensional subspace, which allows for the subsequent lossy reconstruction of the original data. The lower-dimensionality projection serves later as the basis for the final clustering.

- Fusion-based approaches: Here, the data is also projected into lower-dimensional spaces, but not in a combined space; instead, OMICS-individual. Subsequently, those lower dimensional spaces are aligned and fused into a combined space by their specific fusion procedure.

Overall, all methods have their merits and drawbacks. For the deliverable, we considered the following factors for the final methods selection:

- “Federatability”: Here, we evaluated the ability of the method to be federated. For instance, pairwise approaches are inherently hard to federate, since a pairwise similarity calculation would require the exchange of raw patients’ data in different hospitals, or approximative measures.
- Performance/Flexibility: The considered methods should have proven suitable to numerous use-cases and should be flexible enough to be employed for a wide range of OMICS data.
- Popularity: we also considered the number of citations for the tools as a measure of popularity of the method. For FeatureCloud, it is of great importance to replicate popular methods in a federated setting to reduce the burden of starting a federated analysis.

Based on the considerations above, we have produced a prototypical implementation of iCluster [5], which is an expectation-maximization-based approach to the multi-OMICS analysis. Further, due to the popularity, its flexibility and closeness to the principal component analysis, we have decided to implement the MFA [3], which uses joint lower-dimensionality projections for the subtyping. Ultimately, internal evaluations of the performance and result quality have led us to focus our development efforts solely on MFA for a fully functional and tested FeatureCloud App. The MFA implementation is described in the remainder of the deliverable.

5 Methodology

This section provides an overview of the methodology employed in this study. Our focus is on the development of a clustering-based tool accessible through the online FeatureCloud platform. Our goal is to leverage the power of the Multiple Factor Analysis (MFA) algorithm and integrate it into the FeatureCloud library, enabling users to obtain Factor Scores that reveal the underlying patterns within a multi-OMICS dataset. Additionally, by leveraging other available applications within the FeatureCloud platform, users can visualize and extract meaningful insights from these patterns. The first step in our methodology is to incorporate the MFA algorithm into the FeatureCloud library. MFA relies on the already federated Principal Component Analysis (PCA) application. By integrating the MFA results with other available applications, users can generate visualizations that facilitate the exploration and interpretation of the identified patterns. For example, users can plot a factor map based on the Factor Scores obtained from MFA using the published federated K-means app allowing the users to visualize the relationships between different samples or variables.

5.1 Multiple Factor Analysis

Multiple Factor Analysis [1] is a statistical method allowing us to discover a common structure of a multivariate data set. MFA builds on the principles of Principal Component Analysis (PCA) to capture the relationship of several sets of variables pertaining to the same observations. While it is a generalized mathematical approach and was not originally developed in the context of biomedical applications, it is useful in the context of multi-OMICS. Jointly analyzing multiple OMICS provides a comprehensive and holistic understanding of the underlying biological interactions. This approach was first implemented and applied by de Tayrac et al. [3].

MFA is performed over 5 steps:

1. The K tables representing K_j OMICS with the same number of samples and number of features are collected.
2. PCA is performed on each K_j OMICS to retrieve the singular values.
3. The K_j OMICS are normalized by dividing the first singular value denoted σ_1^j for every K_j OMICS denoted as Z_j .
4. The normalized Z_j OMICS are concatenated yielding a K table.
5. PCA is performed on the concatenated K table.

The resulting output is used to derive the global factor scores allowing us to jointly analyze the similarities and dissimilarities between the OMICS. Recall in deliverable 5.1, that PCA decomposes the concatenated K table into a set of mutually singular vectors $K = U\Delta V^T$, where U and V are the left and right singular vectors of K and Δ represents the diagonal matrix of singular values. The global factor scores can be computed as $F = M^{-\frac{1}{2}}U\Delta$, where $M \in R^{n \times n}$ is a diagonal matrix of n samples. The matrix represents the *mass* $m_i = \frac{1}{n}$, which determines the importance of all samples. When extracting the first two principal components by columns of the global factor scores, we can compute the total inertia (i.e., variance) to indicate how much of the data can be explained by the principal components. Given a $\Delta \in R^{n \times n}$ diagonal matrix of singular values denoted δ , the inertia of a principal component is derived by $\tau_i = \frac{\lambda_i}{\sum_j \lambda_j}$ where i is the i th principal component and the eigenvalue $\lambda = \delta^2$. It is possible to project each omic to the global space by first computing the projection matrix $P = M^{-\frac{1}{2}}U\Delta^{-1}$ and then compute every omic by $F_j = T \times (Z_j Z_j^T) P$, where T is the number of OMICS and F_j is the partial factor score for the omic K_j .

In this section, we provide an implementation of MFA extending an existing algorithm PCA [6] running on the FeatureCloud ecosystem describing the variables that are local and shared among the hospitals.

5.1.1 Federated Multiple Factor Analysis

One of the challenges developing an algorithm on the FeatureCloud ecosystem is to preserve the privacy for all patients. Furthermore, we assume the hospitals provide a data set with samples as columns and OMICS as rows. **Figure 1** shows a state diagram of a federated version of MFA. The blue states represent the global computation, where hospitals are sharing data together while red states represent the local computation. All the hospitals are calculating their own data locally. The purple states represent the aggregator-only computations, in which the aggregator collects the required data from the hospitals and performs necessary operations, such as the projection matrix and the global factor score. The state diagram follows the main ideas of the centralized MFA, as explained previously. The separate PCA are based on Hartebrodt's work [6] to initially generate the first singular value σ_1^j used for scaling K_j OMICS. Once the OMICS are scaled and concatenated as a single data set, the aggregator performs the same PCA to retrieve the projection matrix. After computing the MFA application, the federated K-means application can be used to generate a visualization map of the clustered data.

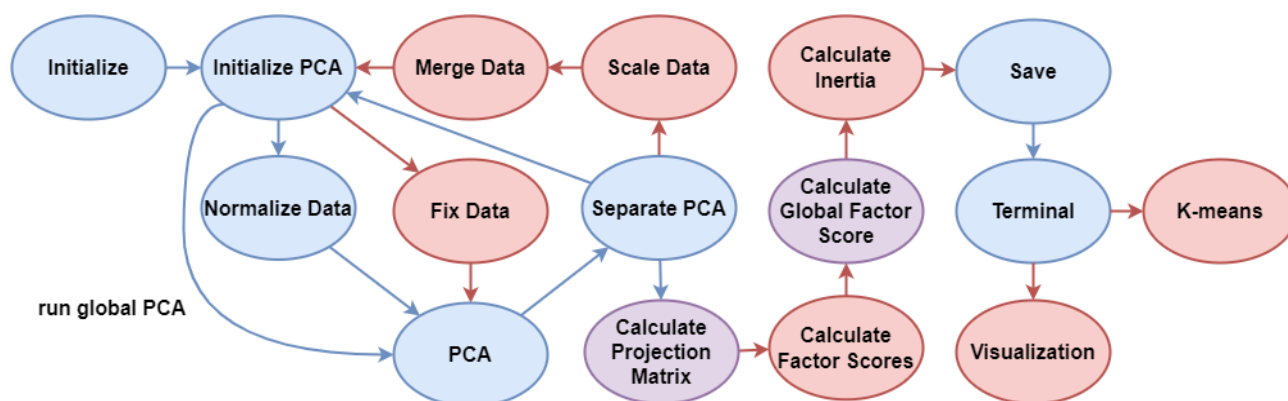


Figure 1: A state diagram of the federated version of MFA. Blue, red and purple are respectively the global, local, and aggregator-only computations.

In **Figure 2**, we see a diagram that shows how MFA computes the factor scores in a federated fashion for two clients denoted $H1$ and $H2$. The steps are similar to Section 5.1, where we use the multi-OMICS data set provided by [3] as an example to show how the genes and samples are computed. In the first step, the clients provide their sample data. $H1$ provides a genomic of 17 samples and 68 genes, and transcriptomic of 17 samples and 356 genes. $H2$ does the same with genomics of 26 samples and 68 genes, and transcriptomic of 26 samples and 68 genes.

The samples are normalized across the rows to zero mean and unit variance. In separated federated PCA (sfPCA), the sample data are scaled by dividing their first singular value for each omic. The scaled sample data are concatenated, where the global federated PCA (gfPCA) is performed to return the decomposition of the sample data $U\Delta V^T$. The left singular vector U remains local to preserve privacy. The singular values and the right vector Δ and V^T are shared globally.

Finally, the federated MFA computes the local variables such as the partial factor scores F_i , projection matrices P , and the global factor score F similarly to Section 5.1. The diagonal mass matrix M is applied globally across all clients i.e., we use the global number of samples to compute the global factor score and the projection matrix.

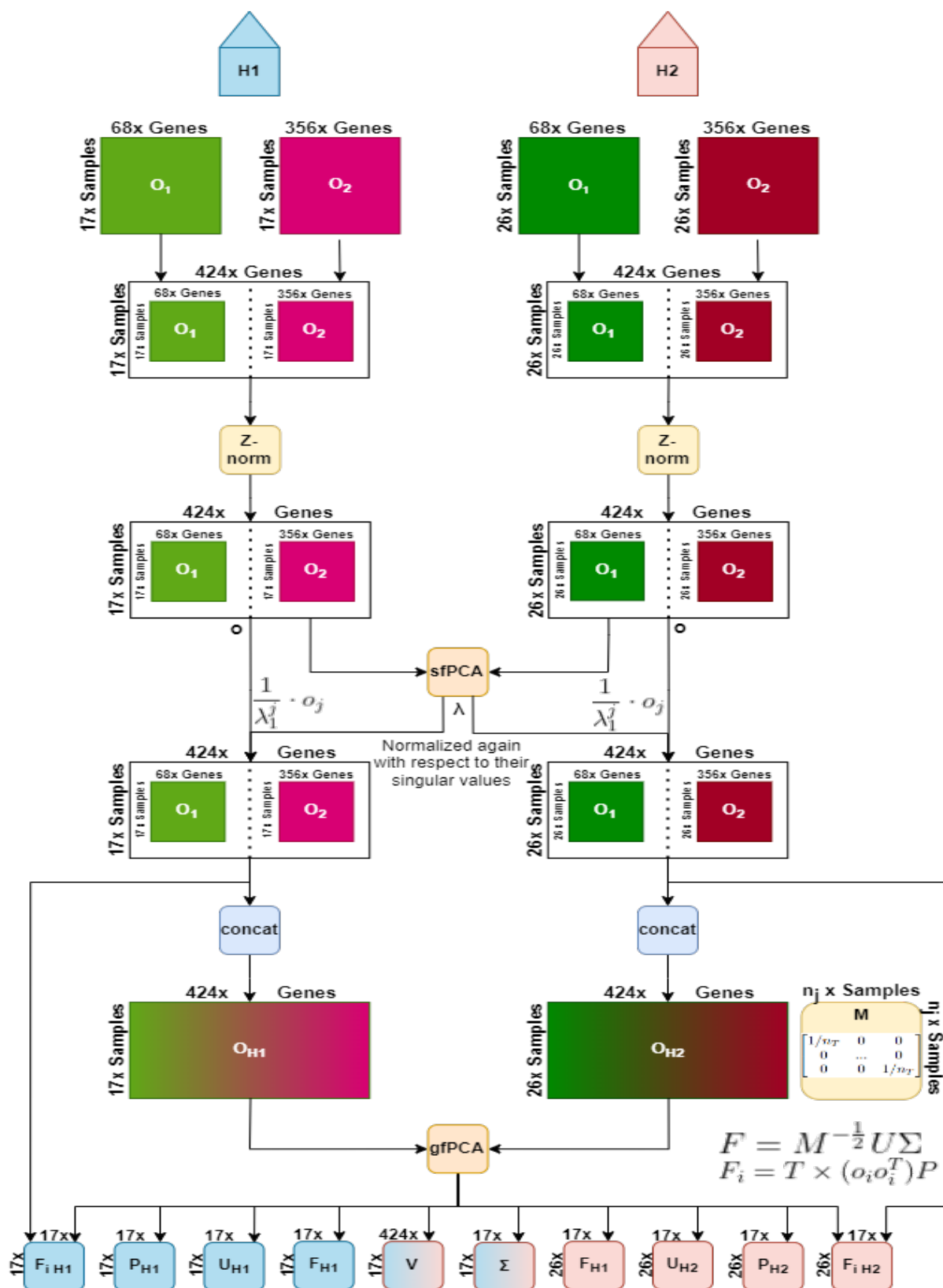


Figure 2: A diagram of federated MFA transforming the OMICS data for two different clients. The output colored blue and red are local to their clients. The color gradient indicates that the variables are shared by both clients.

5.2 Input Files

This section briefly describes how to set up the app by providing the configuration and the input files as well as what kind of output the user would expect once the app finishes.

5.2.1 Configuration

We have defined a configuration file allowing the user to define the names of the input and output files. An example can be seen on **Figure 3**. It is important for the different hospitals to agree on the same naming convention of the input files. For example, the hospital 1 and 2 have respectively the files “bredel_1_1.csv”, “bredel_1_2.csv”, “bredel_2_1.csv”, and “bredel_2_2.csv”. The first number indicates the hospital, the second number is the OMIC type. This allows our federated MFA to distinguish the data files across all hospitals. It is also expected that the rows must be the samples and the columns the features.

```
fc_mfa:
  input:
    datas: bredel_
    extension: .csv
    delimiter: "," # field separator in the input file
  output:
    dir: mfa # output files are saved in 'mfa' folder
    global_factor_score: global_factor_score.tsv
    inertia: inertia.tsv
    omic_factor_score: omic_factor_score
    factor_score_2PC: two_principal_components_factor_score.tsv
```

Figure 3: Example configuration file. The user can modify the naming of the files.

5.2.2 Results folder

The resulting folder as shown in **Figure 4** contains various files such as the global factor score, the partial factor score representing the OMICS and the inertia which is used to derive the total variability. The user can integrate the app into a workflow within the FeatureCloud environment to plot a factor map from the first two principal components in the “global_factor_score.tsv” file, allowing them to explore relationships, patterns and differences among the different types of OMICS data. The “inertia.tsv” file is the same across all hospitals. The “global_factor_score.tsv”, “omic_factor_score.tsv” and “two_principal_components_factor_score.tsv” are local to the hospital. The number of “omic_factor_score.tsv” files depend on the number of OMICS.

```
/mfa
├── global_factor_score.tsv
├── inertia.tsv
├── omic_factor_score_1.tsv
├── omic_factor_score_2.tsv
└── two_principal_components_factor_score.tsv
```

Figure 4: Example results file. When applying the federated MFA onto a dataset of two OMICS, the folder would contain 2 omic factor scores.

6 Results

This section is divided into two subsections: the Multi-OMICS Dataset and the Workflow Application. In the former subsection, we present the outcomes of applying MFA to a well-established dataset. Furthermore, we compare the performance of the federated version of MFA with a centralized solution, highlighting the similarities in the performance of the federated approach.

In the latter subsection, we explore the compatibility of our approach with the FeatureCloud application. We demonstrate how different modules within the workflow can seamlessly interact with each other. Specifically, we investigate the integration of the k-means clustering algorithm after the MFA process to visually display the resulting mapping.

6.1 Standalone app

To illustrate the correctness and consistency of our developed federated MFA, we have compared a multi-OMICS data set containing glioma tumor samples on the prior work by de Tayrac et al. [3]. Their centralized MFA is applied on a joined genomic and transcriptomic sample data. The sample data contains 43 samples, whereas the genomic has 68 features and transcriptomic has 356 features. **Figure 5** shows the centralized solution by de Tayrac et al. [3] and our federated solution. The colored points are following the World Health Organization (WHO) classification: A, astrocytomas; GBM, glioblastomas; O, oligodendrogliomas; OA, mixed oligoastrocytomas.

We illustrate that there are no significant differences between the centralized and the federated solution. The structure is still preserved despite running in the federated fashion. However, the total inertia of centralized (33.7 %) and federated (41.3 %) solutions differs. The federated PCA developed by Hartebrodt et al. [6] has omitted some of the least significant eigenvalues, hence the variation in total inertia.

6.2 Workflow application

The MFA app is ready to be integrated in the FeatureCloud workflow with many other apps. Using the previous work from Deliverable 5.1 and 5.2, we have built a workflow using our federated MFA and federated K-Means clustering in the particular order to illustrate its full potential. To simulate the workflow application for datasets with no labels, we have removed the labels of the glioma data set from Section 5.1 and applied the data to our workflow. The workflow has been executed for two clients, where each client has their own sample data. Recall in Section 5.2.2, MFA outputs among others a file named `two_principal_components_factor_score.tsv`. This file is used for the K-Means clustering. We have merged the clustering results from the clients to demonstrate the usability of our app. In a real life setting, the clients would only see the clustering results with respect to their own sample data.

In **Figure 6**, we see the visualization of the clustering for the entire sample data using our federated MFA and K-Means clustering with K equals 4. We have included the gold standard for comparison from de Tayrac et al. [3]. As expected, the samples are mostly difficult to separate into groups especially for subtypes A and OA, while it captures most of the GBM and O subtypes.

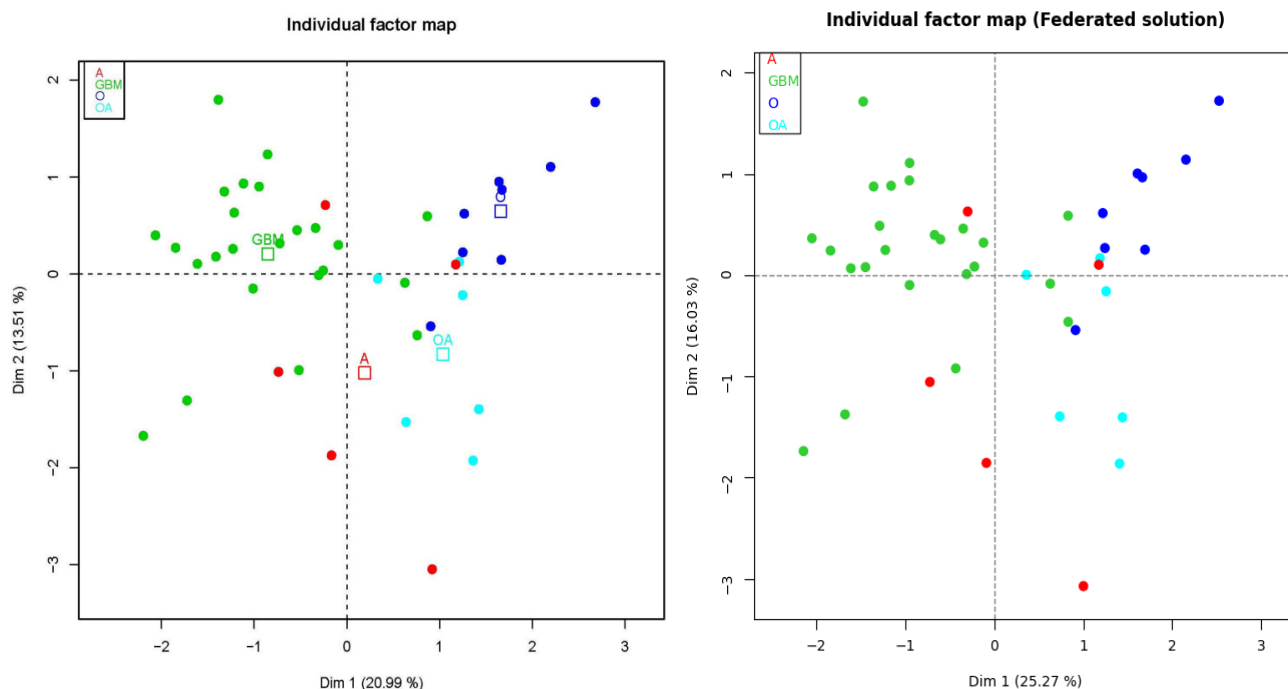


Figure 5. Left side: The individual factor map from the centralized solution provided by de Tayrac et al. [3]. Right side: The federated solution from our algorithm. The algorithm still captures the structure accurately despite running in the federated fashion.

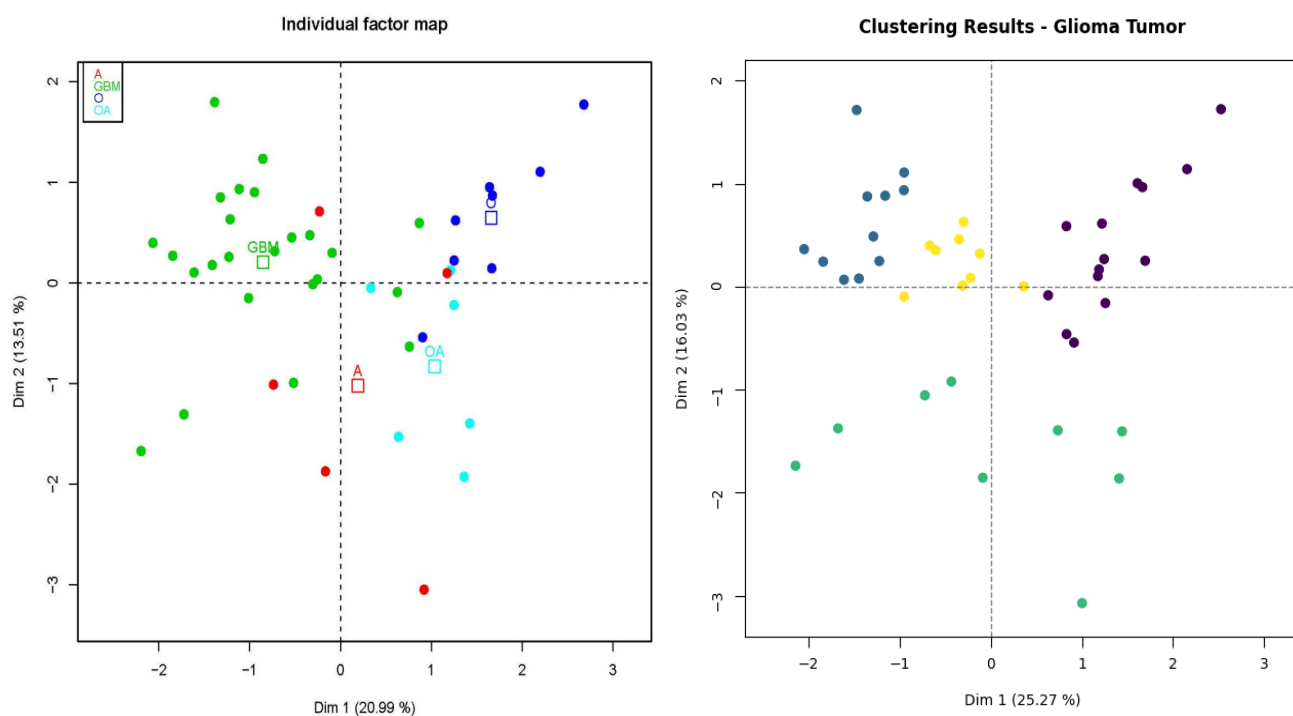


Figure 6: Clustering results after running MFA and K-Means clustering in a workflow. Left side: Gold standard dataset provided by de Tayrac et al. [3]. Right side: Clustering results.

Additionally, we have run the workflow on a genomic and transcriptomic breast cancer dataset by Shen et al. [5] for K equals 4 in the same workflow setting and merged the clustering results as a single dataset. There are 354 samples and 41 features for each omic type. With the total inertia of 19.64 %, we show that breast cancer can be divided into multiple cancer subtypes. **Figure 7** shows the visualization of the clustering results for breast cancer.

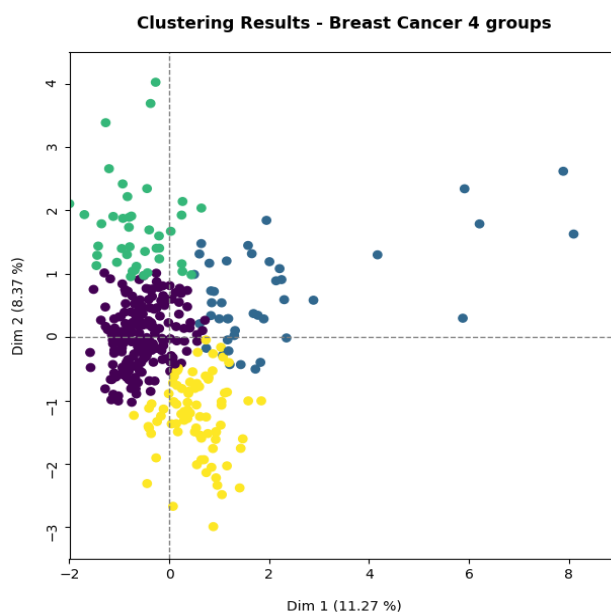


Figure 7: Clustering results after running MFA and K-Means clustering in a workflow.

In a real life setting, the clients can only see the clustering results on their own multi-OMICS sample data. We use **Figure 7** to demonstrate that MFA runs in a federated setting with a number of samples larger than features. **Figure 8** shows the actual results for every client.

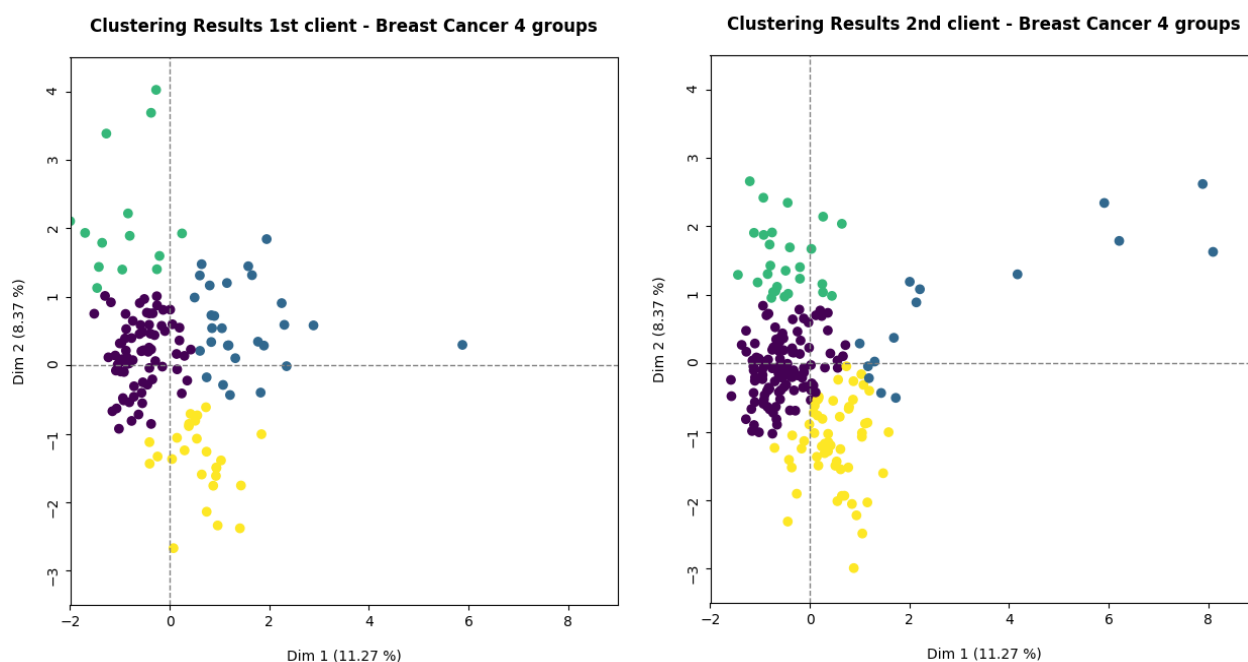


Figure 8: Clustering results for client 1 and 2. Both only have access to their multi-OMICS sample.

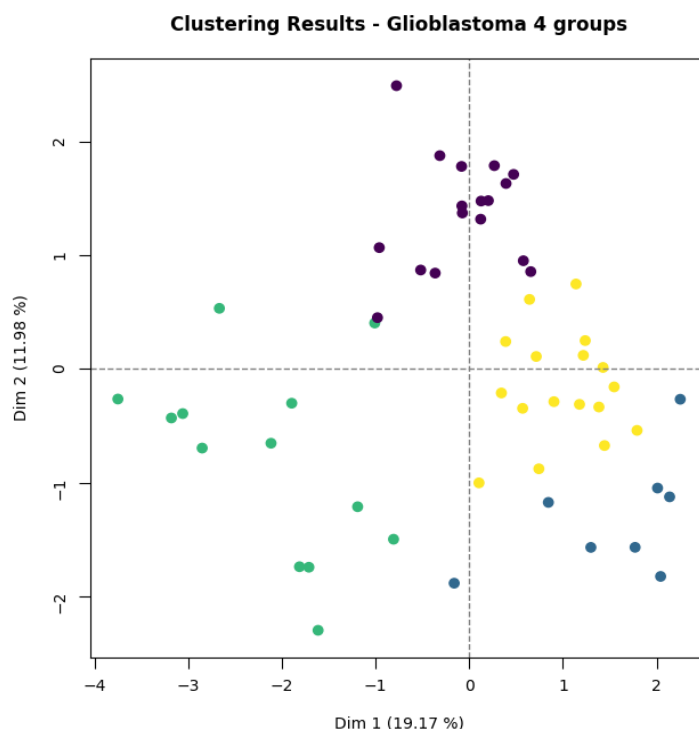


Figure 9: Clustering results for Glioblastoma after running MFA on three different OMICS namely genomics, epigenomics, and transcriptomics.

Finally, we have performed additional tests for one of the most malignant glioma tumors, glioblastoma, by using three different types of OMICS namely genomics, epigenomics, and transcriptomics provided by Subramanian et al. [4]. More specifically, we have performed the experiment with 55 samples with 1599 features belonging to copy numbers, 1514 features to methylation, and 1740 features associated with mRNA. The workflow setting is the same as previous tests, that is, we split the distribute the datasets by samples to two different clients. The MFA algorithm derives the factor scores and the k-means clusters the results. We have deemed that K equals 4 is the best representative of the subtypes as seen in Figure 9.

7 Open issues

- Privacy considerations arise when employing federated PCA, particularly regarding the amount of information contained within the covariance matrix and eigenvectors, and the acceptability of sharing them. In PCA, participants must have access to the complete eigenvectors, resulting in a minimal information leakage. Consequently, further investigation is required to determine if it is actually a problem.
- The decision to prioritize code reusability and simplify development by employing a sequential approach for computing PCAs using individual OMICS datasets, as shown in the diagram, introduces a trade-off. While it eases development complexity, it comes at the cost of increased computational requirements.

8 Conclusion

In conclusion, this deliverable showcases the successful usage of our federated PCA to develop a tool for a more specialized dataset. The Multiple Factor Analysis significantly improves the analysis of multi-OMICS data by effectively capturing the relationships among multiple sets of variables observed on the same samples. By jointly analyzing the various OMICS datasets, a comprehensive and holistic understanding of the underlying biological interactions is achieved.

While the PCA module in FeatureCloud is not directly integrated into the workflow of the MFA application, it serves as the structural backbone for this specialized tool. This architectural decision was made to streamline development and enhance usability. By utilizing the atomic nature of the PCA module, the application simplifies the configuration process, allowing users to fill a straightforward configuration file instead of dealing with complex intermediate steps. This design choice ensures ease of use, as it is only necessary to deal with one application.

Within the context of the FeatureCloud project, our primary objective is to develop applications that respect all sensible medical data. In alignment with this approach, the federated MFA algorithm presented in this deliverable follows this core principle to ensure the exchange of only aggregated model parameters or intermediate results that will not compromise crucial information.

In this deliverable we present a version of the federated MFA tested in different datasets and compared to results given by the centralized version of the algorithm. The development of this algorithm also proves the proper functionality of the FeatureCloud store PCA app. The combination of the MFA with another clustering app can yield useful visualization maps to facilitate the understanding of the results. The design of the project could have handled multiple directions but we chose the one which enhanced the final user experience and the easiest to develop.

9 References

1. Escofier B, Pagès J. Analyses factorielles simples et multiples: objectifs, méthodes et interprétation. 2008.
2. Matschinske J, Späth J, Nasirigerdeh R, Torkzadehmahani R, Hartebrodt A, Orbán B, et al. The FeatureCloud AI Store for Federated Learning in Biomedicine and Beyond. arXiv [cs.LG]. 2021. Available: <http://arxiv.org/abs/2105.05734>
3. de Tayrac M, Lê S, Aubry M, Mosser J, Husson F. Simultaneous analysis of distinct Omics data sets with integration of biological knowledge: Multiple Factor Analysis approach. BMC Genomics. 2009;10: 32.
4. Subramanian I, Verma S, Kumar S, Jere A, Anamika K. Multi-OMICS Data Integration, Interpretation, and Its Application. Bioinform Biol Insights. 2020;14: 1177932219899051.
5. Shen R, Olshen AB, Ladanyi M. Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis. Bioinformatics. 2009;25: 2906–2912.
6. Hartebrodt A, Röttger R, Blumenthal DB. Federated singular value decomposition for high dimensional data. arXiv preprint arXiv:220512109. 2022. Available: <http://arxiv.org/abs/2205.12109>

10 Other supporting documents / figures / tables (if applicable)

Package	Version
bios	0.1.2
bottle	0.12.21
joblib	1.1.0
jsonpickle	2.2.0
numpy	1.22.4
pandas	1.4.2
pydot	1.4.2
pyyaml	6.0

Source code

- Multiple Factor Analysis:
 - <https://github.com/NicolaiT/multiplefactoranalysis/tree/linear-design>