HHS Public Access

Author manuscript

Proc SPIE Int Soc Opt Eng. Author manuscript; available in PMC 2019 May 01.

Published in final edited form as:

Proc SPIE Int Soc Opt Eng. 2019 February ; 10953: . doi:10.1117/12.2506032.

A web-based system for statistical shape analysis in temporomandibular joint osteoarthritis

Loic Michoud[a], **Chao Huang**[b], **Marilia Yatabe**[a], **Antonio Ruellas**[a], **Marcos Ioshida**[a], **Beatriz Paniagua**[c], **Martin Styner**[b], **João Roberto Gonçalves**[d], **Jonas Bianchi**[a,d], **Lucia Cevidanes**[a], and **Juan-Carlos Prieto**[b]

[a]Dept. of Orthodontics and Pediatric Dentistry, University of Michigan, 1011 N University Ave, Ann Arbor, MI, USA 48109

[b]Dept. of Statistics and Operations Research, University of North Carolina at Chapel Hill, Chapel Hill, Hanes Hall, Campus Box 3260, NC, USA 27599

[c]Kitware, Inc., 101 East Weaver Street, Carrboro, NC, USA 25710

[d]Dept. of Pediatric Dentistry, Sao Paulo State University (Unesp), School of Dentistry 1680 Humaita St, Araraquara, SP, Brazil 14801-385.

Abstract

This study presents a web-system repository: Data Storage for Computation and Integration (DSCI) for Osteoarthritis of the temporomandibular joint (TMJ OA). This environment aims to maintain and allow contributions to the database from multi-clinical centers and compute novel statistics for disease classification. For this purpose, imaging datasets stored in the DSCI consisted of three-dimensional (3D) surface meshes of condyles from CBCT, clinical markers and biological markers in healthy and TMJ OA subjects. A clusterpost package was included in the web platform to be able to execute the jobs in remote computing grids. The DSCI application allowed runs of statistical packages, such as the Multivariate Functional Shape Data Analysis to compute global correlations between covariates and the morphological variability, as well as local p-values in the 3D condylar morphology. In conclusion, the DSCI allows interactive advanced statistical tools for non-statistical experts.

Keywords

Web-platform; Biomarkers; Meshes; Statistics; Temporomandibular Joint Disorders; Osteoarthritis

1. INTRODUCTION

In the field of medical research, one of the main objectives is to develop tools that can be widely used by the greatest number of physician, dentists, researcher, and the general public[1,2]. It is therefore essential to make these new tool features accessible to all interest users. The most powerful algorithms used in medical/dental data processing, such as neural networks[3] and shape statistic (Multivariate Functional Shape Data Analysis – MFSDA[4]), require very large sample sizes to be effective. Standardized protocols for collection of medical/dental data in different clinics or hospital are required to solve problems related to

control of sample dimensionality and heterogenic. The first step in any study in a multitude of fields is to collect and store the large number of data of different types such as clinical, biological ad even 3D images. The Data Storage for Computation and Integration (DSCI) that we have created utilizes a database collected by clinicians researching temporomandibular joint (TMJ) osteoarthritis (OA). The security of the data is also a concern and while other technologies, such as Blockchain[5], have been proposed as platforms for management of data collections in a secure environment, the current prototypes seem to be lees secure than our current database and we used a JSON web token[6] that is a highly encrypted token. In addition, the cluster post technology is used to send tasks to remote servers such as the Umich Flux technology. It allows the users to send a big amount of computing tasks without having trouble with the run rime because that technology is much faster than standard computers. Our long-term goal is to create repositories for clinical studies maintaining the data in a distributed computational environment to allow contributions to the database for multi-clinical centers and to shares trained models for TMJ classification. The novel shape statistics requires advanced computational power and is accessible via the website, where it is possible to execute computing intensive task outside the clinical centers.

Initially is it necessary to create a tool on the website DSCI to test correlations between the clinical and biological data and the 3D condylar meshes, in order to know which area of the condyles could become related with the covariates. Then, the MFSDA tests could be predictive of TMJ OA prognosis, i.e. whether a patient is at risk for worsening clinical symptoms and condylar degeneration. The MFSDA is a statistical package that are no correlated with each other, and which are strongly correlated with the principal variables in the study. The pre-processing plugin is a package available on the website to run those correlations in order to test correlations among covariates as well as their correlation with the principal components to determine which covariates to include in the MFSDA model. The purpose of this study is to create this web-system repository, DSCI for Osteoarthritis of the TMJ (TMJ OA) to maintain the data in a distributed computational environment and allow contributions from multi-clinical center, compute novel shape statistics, train a neural and allow interactive visualization of data stored and computational results.

2. METHODS

2.1 Multivariate functional shape data analysis (MFSDA)

Suppose that we observe an image dataset for n unrelated subjects. Without loss of generality, we focus on a compact set, denoted as $D \subset R^t$, which is general enough to cover curves (t = 1), contours (t = 2), and surfaces (t = 3). It is assumed that $\{d_1, \ldots, d_{NV}\}$ are NV grid points (or vertices) on D from the template file. Specifically, for the i-th subject, we observe a $J \times 1$ vector of shape measurements corresponding to each grid point d, denoted as $y_i(d) = (y_{i1}(d), \ldots, y_{iJ}(d))^T$, and a $p \times 1$ vector of covariates (e.g., age, gender, group information, and biological markers), denoted as x_i. The MFSDA equation(1) is defined as

$$y_{ij}(d) = X_i^T\left(B_j(d) + n_{ij}(d) + e_{ij}(d) \quad (1)$$

where B_j is a $p \times 1$ vector of fixed effects, $n_i(d) = (n_{i1}(d), \ldots, n_{iJ}(d))^T$ characterizes both subject-specific and location-specific variability, and $e_i(d) = (e_{i1}(d), \ldots, e_{iJ}(d))^T$ are measurement errors. It is also assumed that $n_i(d)$ and $e_i(d)$ are mutually independent. Compared with the standard linear regression model, MFSDA explicitly accounts for spatial smoothness, spatial correlation, and the low-dimensional representation of functional shape responses [7,8] Under model Equation 1, we investigate whether there is a statistically significant morphological difference caused by some covariate of interest or linear combination of covariates of interest, a local Wald-type test statistic is used.[9] To estimate all unknown parameters in model (1), we employ a weighted least squares (WLS) method based on the multivariate local polynomial kernel smoothing technique[7]. In general, MFSDA consists of four steps: 1) Fit a model under the null hypothesis which yields $B_j(d)^*$, $n_{ij}(d)^*$, $e_{ij}(d)^*$ for all i and d. 2) Generate random bootstrap samples from all i and d. 3) For the bootstrap samples, calculate the Wald-type statistic T_n. 4) The *p-value* of T_n is calculated using an approximation method by a X^2 – type random variable.[8–10]

2.2 Pearson correlations and principal component score (PCA)

The Pearson and PCA package were implemented on the DSCI web platform to tests both the Pearson correlations and the principal component scores among clinical or biological variables. The main aim of this module is to reduce the dataset to allow more efficient processing time when running the MFSDA model. In order to avoid the use of repetitive features of the dataset, the package allow the users to test the Pearson correlations among biological or clinical variables to select covariates not correlated with each other. The Pearson's correlation coefficient is the covariance of the two variables divided by the product of their standard deviation. That's why it has values between −1 and +1 where 1 is a total linear correlation and −1 a total negative correlation. This first step will be very useful to select which covariate to include in the statistical model when looking at the results of the principal component analysis (P). This second step is used to reduce the number of covariates to include in the model. The principal component uses an orthogonal transformation to convert correlated variables into linearly uncorrelated variables called principal components that explain a given percentage of the information contained in all the dataset. We then test the Pearson Correlation between the principal components and the covariates to detect which variables could be used as a principal component. To avoid repetitive features the user has to check if the covariates highly correlated with the principal components are correlated with each other. Indeed, all the covariates to be included in the statistical model must be uncorrelated with each other and highly correlated with each principal component. A covariate is considered as correlated if its p-value is below 0.05. In order to facilitate the use of this package, a pdf file is created when processing the statistical tool that summaries all the results in tables using different color bar.

2.2 Implementation of the DSCI web platform

The DSCI system is a web-based application currently hosted on the Elastic Computing Cloud (EC2) of Amazon web services. Access to DSCI is handled using JSON Web Tokens (JWT), with JWT encryption of each user login. The DSCI facilitates and centralizes multiple data sources. It keeps track of data from previous experiments and assures the reproducibility of future experiments. In the TMJ study, the different types of data used are: responses to clinical questionnaires, protein levels in plasma and saliva (measured with quantitative microarrays via spreadsheet files), and imaging data via the DataBaselnteractor plugin[11], available on open-source software called 3D-Slicer[12]. It allows the user to upload a VTK file[13] on the database used by the platform. Node is a Javascript engine that facilitates building scalable network applications. Using Hapi as the server framework, we are able to build services and focus on writing reusable application logic instead of pure infrastructure. Hapi is fully REST and orchestrates communication between all components in the system. The tool used for storage is Couchdb, a NoSQL type database[14], i.e., it does not store data and relationships in tables. Instead, each database is a collection of independent JSON documents. JSON is a flexible format and facilitates encoding data without enforcing a predefined rigid structure. In order to merge all the data stored in the system, the map reduce algorithm is used. It indexes the data in the system, using a unique patient anonymized id and users are allowed access after approval by the administrator. The Clusterpost package is used in our web platform to send tasks to remote computing grid in order to run plug-ins such as MFSDA that require high processing power. This package allows the users to execute the jobs through the plug-ins in remote grids using a REST api and nodes with Hapijs in the server side application. Concerning TMJ OA study, the use of Umich Flux technology (remote computing grid) is essential to be able to send a big amount of computing tasks at the same time. The Clusterpost execution is then useful to get the output data of all those runs which can be downloaded from the website. The interactive display available on the web site facilitates the understanding of the runs. The 3D visualization of the morphological data is handled by the visualization toolkit (vtk.js) and the visualization of the results has been developed using D3.js[15]. It is a JavaScript library for visualizing data using web standards.

3. RESULTS

The MFSDA package built associations between covariates (biological markers or clinical markers) and the morphological variability. This package computed the global correlations with morphological variability as well as local p-values in the 3D condylar morphology (Figure 1) The MSFDA tool worked with the coordinates of 3D meshes stored in VTK file format. A multivariate varying coefficient model then tested the association between the multivariate shape measurements, the demographic information and other clinical and biological variables.

The output values are stored in comma-separated value (csv) files but also in JSON files to fit with the documents used in the web platform. (Fig. 2)

The preprocessing model is currently available on the web platform as a plug-in. The user needs to create a project, choose the number of components to process (number of principal

components always lower than the total number of variables) and select the type of patient to study (OA or Control in TMJ study). The visualization of the state of the remote runs is displayed on the website (Fig 3) and is used to download the output files, display specific information from the script and show the state of the run (queue, run, done or failed).

The d3 display is fully interactive, indeed the user has the possibility to set his own color bar and scale to plot the results according to his expectations (Fig 4).

4. CONCLUSION

This is the first work to implement the computation of MFSDA in a web-based system that facilitates computation and access and to multiple data sources. This website can be used for distributed learning storage and management of data collected at different clinics or hospital, and training of algorithms bases on a neural network, in order to increase its accuracy. We developed efficient web-based data management, mining, and analytics that integrates and analyze clinical, biological, and high dimensional imaging data from TMJ OA patients. The Data Storage for Computation and Integration (DSCI) remotely computes machine learning, image analysis, and advanced statistics from patients with and without TMJ-OA. Our long-term goal is to create and maintain the data in a distributed computational environment to allow contributions to the database from multi-clinical centers and to share trained models for TMJ classification.

ACKNOWLEDGMENTS

This study was partially supported by NIF1 grants DE R01DE024450 and R21DE025306.

REFERENCES

[1]. Gomes LR, Gomes M, Jung B, Paniagua B, Ruellas AC, Gonsalves JR, Styner MA, Wolford L and Cevidanes L, “Diagnostic index of 3D osteoarthritic changes in TMJ condylar morphology,” Proc. SPIE-the Int. Soc. Opt. Eng 9414, Hadjiiski LM and Tourassi GD, Eds., 941405 (2015).

[2]. Cevidanes LHS, Hajati A-K, Paniagua B, Lim PF, Walker DG, Palconet G, Nackley AG, Styner M, Ludlow JB, Zhu H and Phillips C, “Quantification of condylar resorption in temporomandibular joint osteoarthritis,” Oral Surgery, Oral Med. Oral Pathol. Oral Radiol. Endodontology 110(1), 110–117 (2010).

[3]. de Dumast P, Mirabel C, Cevidanes L, Ruellas A, Yatabe M, Ioshida M, Ribera NT, Michoud L, Gomes L, Huang C, Zhu H, Muniz L, Shoukri B, Paniagua B, Styner M, Pieper S, Budin F, Vimort J-B, Pascal L, et al., “A web-based system for neural network based classification in temporomandibular joint osteoarthritis.,” Comput. Med. Imaging Graph. 67, 45–54 (2018). [PubMed: 29753964]

[4]. Huang C, Thompson P, Wang Y, Yu Y, Zhang J, Kong D, Colen RR, Knickmeyer RC, Zhu H and Alzheimer’s Disease Neuroimaging Initiative., “FGWAS: Functional genome wide association analysis,” Neuroimage 159, 107–121 (2017). [PubMed: 28735012]

[5]. Zheng Z, Xie S, Dai H, Chen X and Wang H, “An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends,” Proc. - 2017 IEEE 6th Int. Congr. Big Data, BigData Congr 2017 (2017).

[6]. Jones M, Bradley J and Sakimura N, “JSON Web Token (JWT)” (2015).

[7]. Yandell BS, Wand MP and Jones MC, “Kernel Smoothing,” Technometrics (1996).

[8]. Huang M, Nichols T, Huang C, Yu Y, Lu Z, Knickmeyer RC, Feng Q, Zhu H and Alzheimer’s Disease Neuroimaging Initiative., “FVGWAS: Fast voxelwise genome wide association analysis of large-scale imaging genetic data,” Neuroimage 118, 613–627 (2015). [PubMed: 26025292]

[9]. Chen C-H, Gutierrez ED, Thompson W, Panizzon MS, Jernigan TL, Eyler LT, Fennema- Notestine C, Jak AJ, Neale MC, Franz CE, Lyons MJ, Grant MD, Fischl B, Seidman LJ, Tsuang MT, Kremen WS and Dale AM, "Hierarchical Genetic Organization of Human Cortical Surface Area," Science (80-.). 335(6076), 1634–1636 (2012).

[10]. Zhang J-T and Chen J, "Statistical inferences for functional data," Ann. Stat. 35(3), 1052–1079 (2007).

[11]. "DatabaseInteractor.", https://www.slicer.org/wiki/Documentation/4.8/Modules/DatabaseInteractor, 2017.

[12]. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin JC, Pujol S, Bauer C, Jennings D, Fennessy F, Sonka M, Buatti J, Aylward S, Miller JV, Pieper S and Kikinis R, "3D Slicer as an image computing platform for the Quantitative Imaging Network," Magn. Reson. Imaging (2012).

[13]. "VTK.js.", https://kitware.github.io/vtk-js/index.html., 2018.

[14]. "Seamless multi-master sync, that scales from Big Data to Mobile, with an Intuitive HTTP/JSON API and designed for Reliability.", https://couchdb.apache.org, 2017.

[15]. Dr.js., "Data-Driven Documents," https://d3js.org/., 2017.

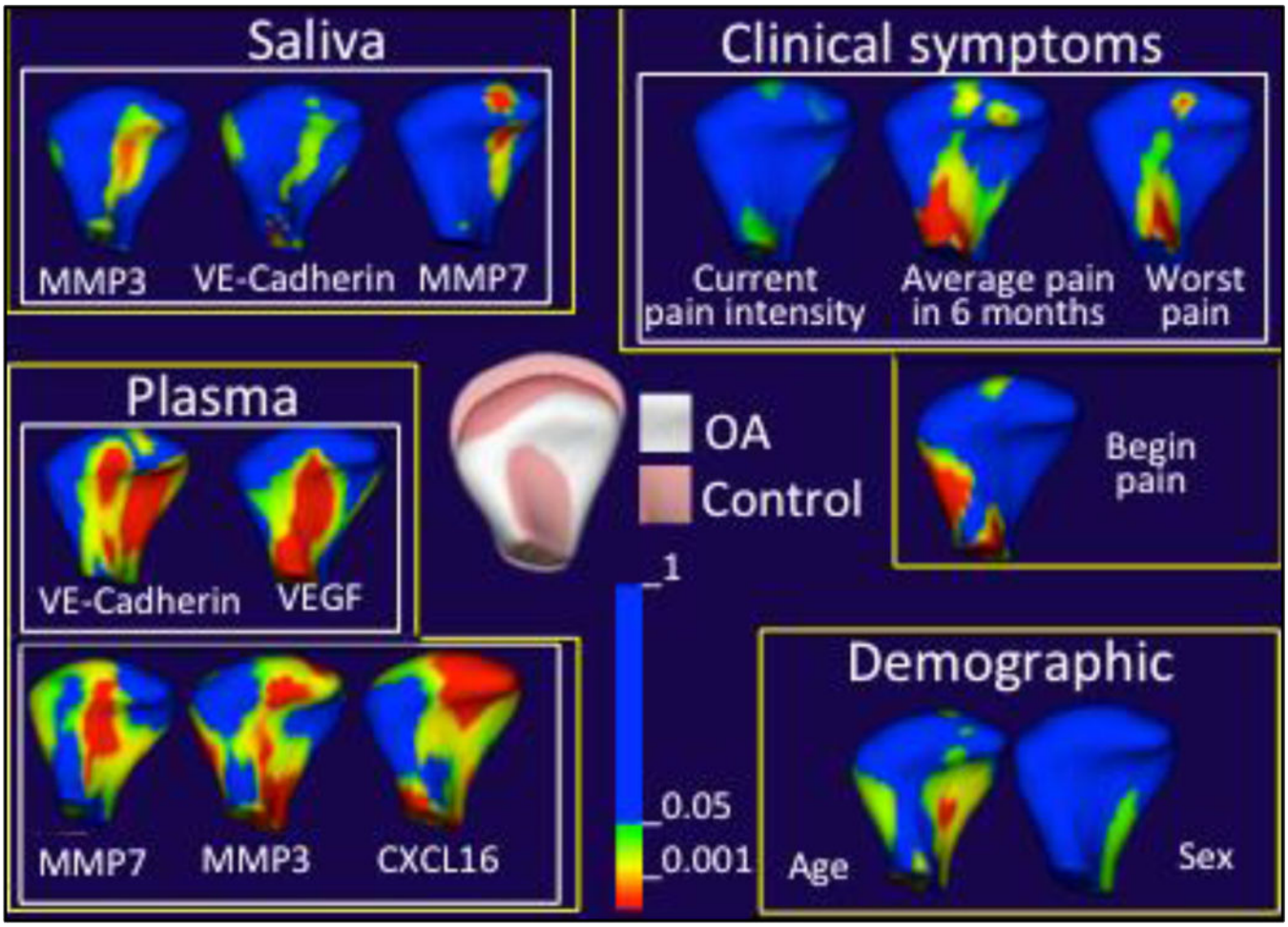

Figure 1.
p-values visualization of clinical and biological markers on a 3D mesh computed by MFSDA.

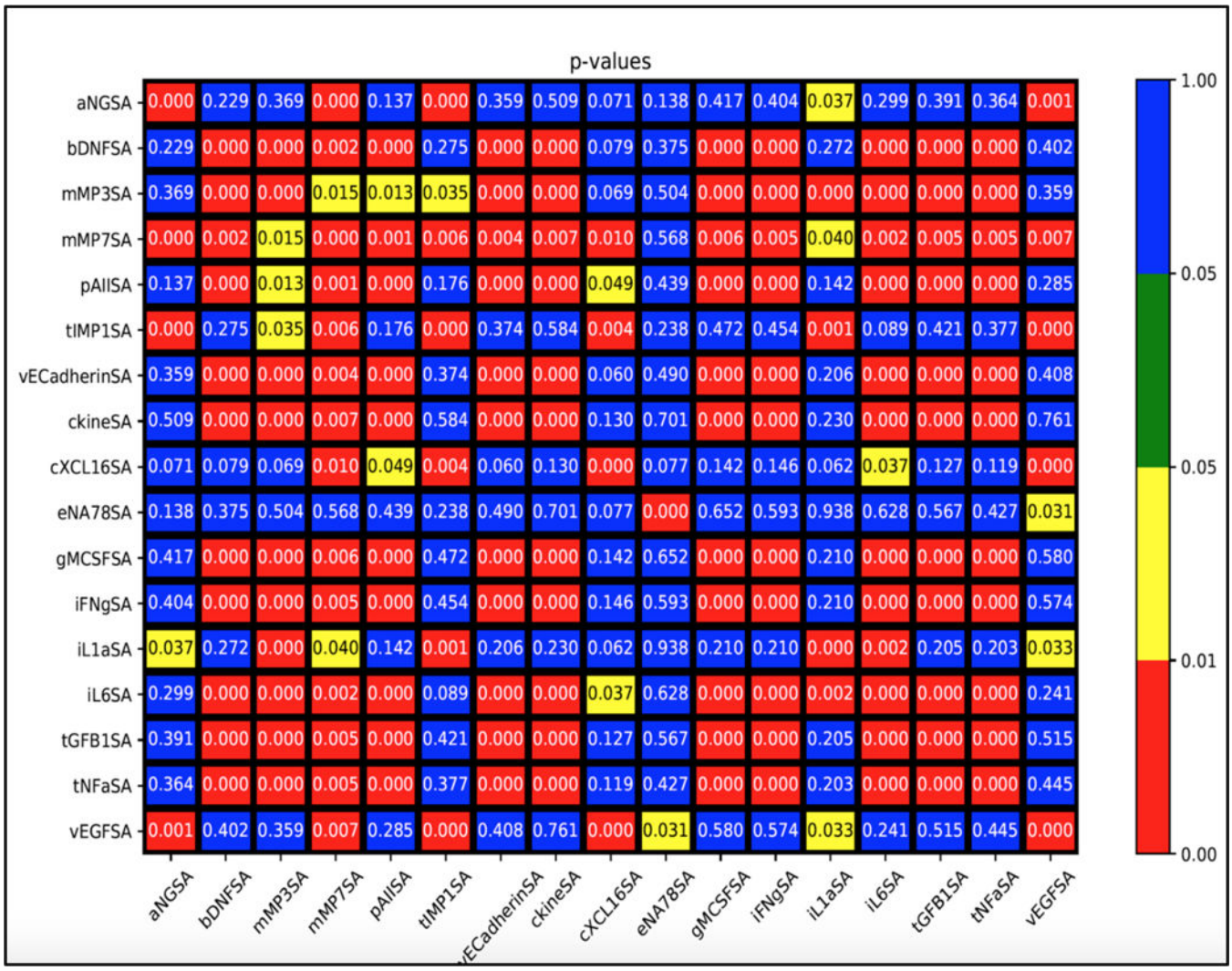

Figure 2.
p-values among all the covariates computed by the preprocessing package.

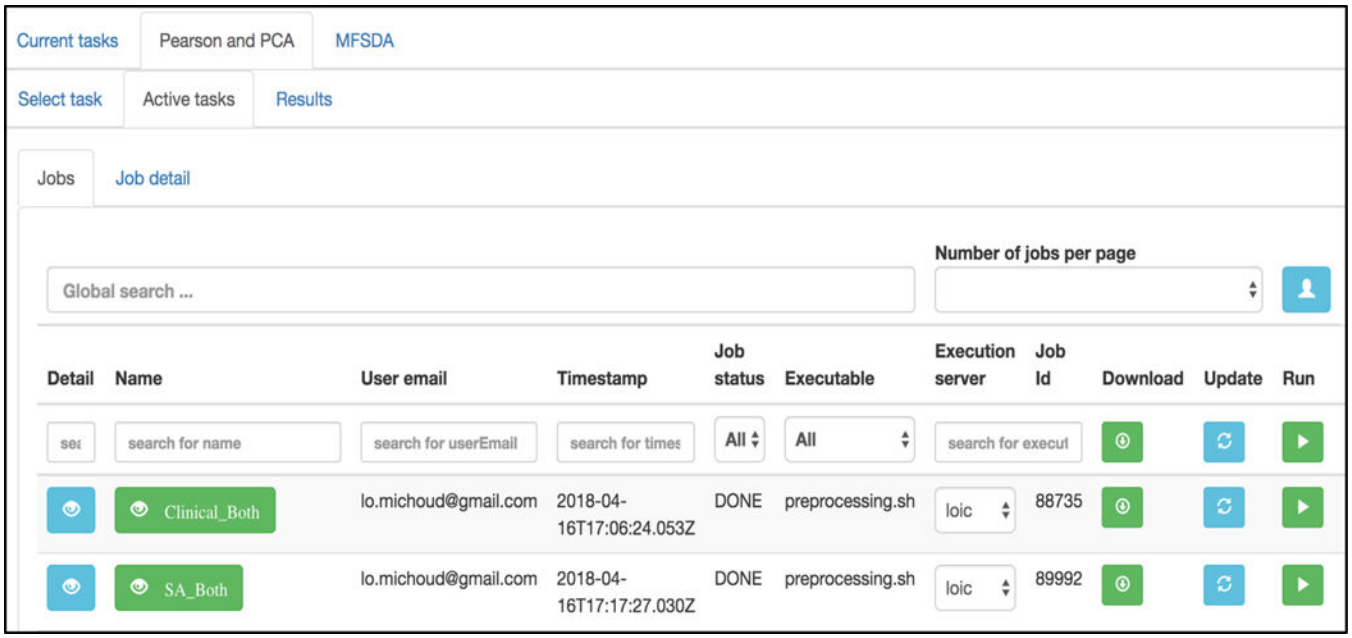

Figure 3.
Remote runs visualization.

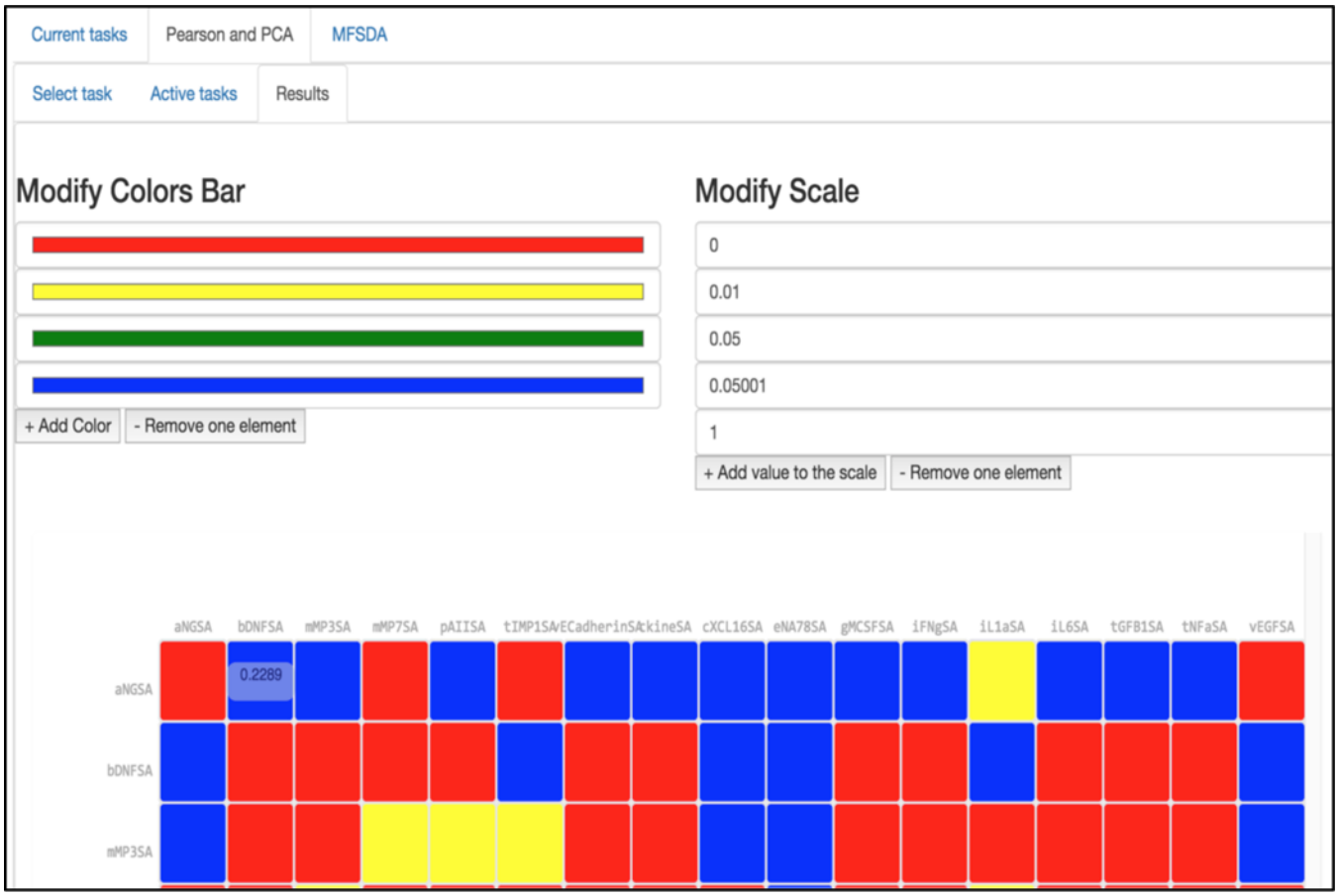

Figure 4.
Display of the preprocessing results using d3.js.