World Academy of Science, Engineering and Technology
International Journal of Medical and Health Sciences
Vol:10, No:9, 2016

Development of a Telemedical Network Supporting an Automated Flow Cytometric Analysis for the Clinical Follow-up of Leukaemia

Claude Takenga, Rolf-Dietrich Berndt, Erling Si, Markus Diem, Guohui Qiao, Melanie Gau, Michael Brandstoetter, Martin Kampel, Michael Dworzak

Open Science Index, Medical and Health Sciences Vol:10, No:9, 2016 publications.waset.org/10005624.pdf

***Abstract*—**In patients with acute lymphoblastic leukaemia (ALL), treatment response is increasingly evaluated with minimal residual disease (MRD) analyses. Flow Cytometry (FCM) is a fast and sensitive method to detect MRD. However, the interpretation of these multi-parametric data requires intensive operator training and experience. This paper presents a pipeline-software, as a ready-to-use FCM-based MRD-assessment tool for the daily clinical practice for patients with ALL. The new tool increases accuracy in assessment of FCM-MRD in samples which are difficult to analyse by conventional operator-based gating since computer-aided analysis potentially has a superior resolution due to utilization of the whole multi-parametric FCM-data space at once instead of step-wise, two-dimensional plot-based visualization. The system developed as a telemedical network reduces the work-load and lab-costs, staff-time needed for training, continuous quality control, operator-based data interpretation. It allows dissemination of automated FCM-MRD analysis to medical centres which have no established expertise for the benefit of an even larger community of diseased children worldwide. We established a telemedical network system for analysis and clinical follow-up and treatment monitoring of Leukaemia. The system is scalable and adapted to link several centres and laboratories worldwide.

***Keywords*—**Data security, flow cytometry, leukaemia, telematics platform, telemedicine.

I. INTRODUCTION

ALL is the most frequent leukaemia entity in children and adolescents. Despite continued progress and refinement of therapeutic approaches, disease relapse due to insufficient extinction of leukemic blasts still remains the number one cause of treatment failure. About 15-20% of paediatric patients with the disease still suffer from relapse. FCM is one of the methodologies most useful in this respect, because it is widely available and applicable to most patients. While FCM acquisition can be harmonized straightforwardly, data analysis and interpretation rely largely on operator skills and experience. Hence, these requirements represent the current bottleneck of safely applying the FCM- MRD methodology in a growing community of diagnostic laboratories. Thus, Machine Learning (ML) methods represent an objective methodology for automated analysis of FCM data to cluster/ classify cell populations and determine MRD.

Advancement of telemedicine is strongly affected by the convergence of medicine and continuously progressing information and communication technologies, increasing demands, and finally rising cost pressure from health organizations systems. A number of projects in the last decades have been focusing on platform concepts for hosting telemedical solutions. The frame-work, proposed in [1] can be exploited for research projects in the laboratories of automation process. A robust platform has been proposed in [2] and has extended the state-of-the art by addressing issues of communication security and privacy in the provision of telemedical services.

Several approaches and applications for automated FCM-data analysis were already published. For example, classification of FCM-data based on Support Vector Machine (SVM) learning [3], [4], model-based clustering using multivariate mixture models with t-distributions [5] or skew and heavy-tailed distributions [6], Bayesian statistical approaches using Gaussian distributions [7], [8] and neural networks [9] are supervised learning and statistical methods useful to identify known event populations.

Software packages implementing ML approaches for FCM-data were already published for automated gating purposes. The software FCMClust, proposed in [10], implements a model-based clustering approach based on multivariate t mixture models. It also features outlier removal and data transformation functions. The open-source software FCM, developed in [11] features several data analysis and clustering methods as well as annotation of FCM data with a graphical user interface. However, there is no ready-to-use analysis tool available to implement the complete FCM-data analysis pipeline, as described in [12]. Such a comprehensive tool would contain automated gating features along with data quality control, normalization, outlier removal, and interpretation tasks. Hence, in this paper we develop and validate practically such a pipeline-software package as a ready-to-use FCM-based MRD-assessment prototype-tool for clinical practice in patients with ALL.

C. Takenga is with TU Wien, Faculty of Informatics, Computer Vision Lab, 1040 Vienna, Austria (corresponding author; phone: +49-17663067606; fax: +43-1-58801-18399; e-mail: takenga@caa.tuwien.ac.at).

R-D. Berndt and E. Si are with the Infokom GmbH, 17033 Neubrandenburg, Germany. (e-mail: rberdnt, esi@infokom.de).

M. Diem, M. Gau and M. Kampel are with TU Wien, Faculty of Informatics, Computer Vision Lab, 1040 Vienna, Austria (e-mail: mdiem, mgau, kampel@caa.tuwien.ac.at).

M. Brandstötter and G. Qiao are with the CogVis GmbH, 1040 Vienna, Austra (e-mail: qiao@cogvis.at)

M. Dworzak is with both the Labdia Labordiagnostik, Vienna, Austria and St. Anna Children's Hospital and Children's Cancer Research Institute, Department of Paediatrics, Medical University of Vienna (e-mail: michael.dworzak@ccri.at).

Open Science Index, Medical and Health Sciences Vol:10, No:9, 2016 publications.waset.org/10005624.pdf

II. STATE-OF-THE-ART IN ANALYSIS SOFTWARE

The FCM data used in this project were derived from acquisitions on flow cytometers of the two most prevalent systems used world-wide, i.e. of the two leading companies Becton-Dickinson (BD) and Beckman-Coulter (BC). Hence, data files are based on the respective acquisition software systems of the two companies: FACSDiva™ (BD) and Kaluza™ (BC). Depending on these systems, the files stored in the folders have different data structures. However, they are in both cases based on the FCS format which stores the measurement information for each event as well as the FCM machine settings. FCS is an open file standard defined by the International Society for Advancement of Cytometry (ISAC). Within this AutoFLOW project, each patient is stored in a folder, named according the patient’s pseudonym. The following files are contained with different composition according to BD or BC software peculiarities:

A patient gated with FACSDiva™ (BD):

- Each folder contains multiple FCS files, where each FCS file belongs to one tube and one day. Thus up to 6 files can be present (2 tubes and 3 treatment days).
- Additionally, an XML containing the gating information (GT) is present.
- As naming convention, the XML file is named according the patient pseudonym (e.g. *500.xml*). The names of the FCS files are defined by the acquisition software, and are referenced within the xml file.

A patient gated with Kaluza™ (BC)

- Each folder contains up to 3 Analysis files (3 treatment days; proprietary file format Kaluza™).
- Each Analysis file contains the FCS information as well as the GT.
- As naming convention, the analysis file is named according the patient pseudonym and the treatment day (e.g. 500.d15.analysis).

In the case of FACSDiva™, the xml file is directly parsed, whereas in the case of Kaluza™ the file is first split into the FCS and xml part. Based on the R interface, common data structures for the follow up processing are available.

III. SYSTEM DESCRIPTION

A. Progress beyond the State-of-the-Art

The automated FCM-MRD analysis developed in this work is able to correctly identify and quantify leukemic populations in FCM-data from bone marrow of patients with ALL drawn at clinically relevant time-points during induction therapy without the need of expert knowledge in order to significantly reduce work-load. We will assess the causes and situations of false machine votes in order to be able to further improve the tool in a later phase as needed. In this project, a number of papers describing methods and algorithms developed and used have been published. In [13] a comparison among three different approaches for the automatic detection of leukemic cells in FCM data samples. The first approach is a fully-discriminative baseline based on the SVM. The second approach is divided in two phases, the first is an unsupervised feature learning using a Stacked Auto-Encoder Neural Network, the second phase consists of exploiting the previously trained Neural Network to perform inference on the data classification. The third approach is fully generative and based on the Gaussian Mixture Model estimation and Bayes decision. Results show that a generative model provides better accuracy on MRD estimation with respect to the other two approaches. In [14], [15] this EU-funded project (AutoFLOW) aiming at the standardization and automation of FCM-MRD analysis by ML technology has been introduced. Reference [16] presents a novel semi-automatic approach for classification of biologically meaningful cell populations in imaging cytometry. In [17], a brief overview on the role of Computer Vision and ML in solving medical problems in an automatic (supervised or unsupervised) fashion is given, a case study of FCM data analysis for MRD assessment in ALL is considered. Two different approaches are proposed: the first is generative semi-supervised Gaussian Mixture Model based approach, the latter is a discriminative semi-supervised Deep Learning based approach.

Fig. 1 illustrates the four most promising strategies (based on literature search and preliminary data) for the implementation of the automated gating and cluster labelling that were evaluated in [13], [16], adjusted and validated in this project.

B. Implementation as a Telemedical Network

Modern information technology is increasingly used in healthcare with the goal to improve and enhance medical services and to reduce costs. E-health has been proved to be suitable candidate for the future of health care provision and it has become a very promising market for the industry.

This work introduces a secure and scalable platform which enables the implementation of health applications and their provision based on SaaS (software as a service) technology [2].

The platform is composed of a set of modules. Each module is for a particular service or group of services responsible. These services can be consumed at the application layer by the end-users as SaaS-model or as modular business solution.

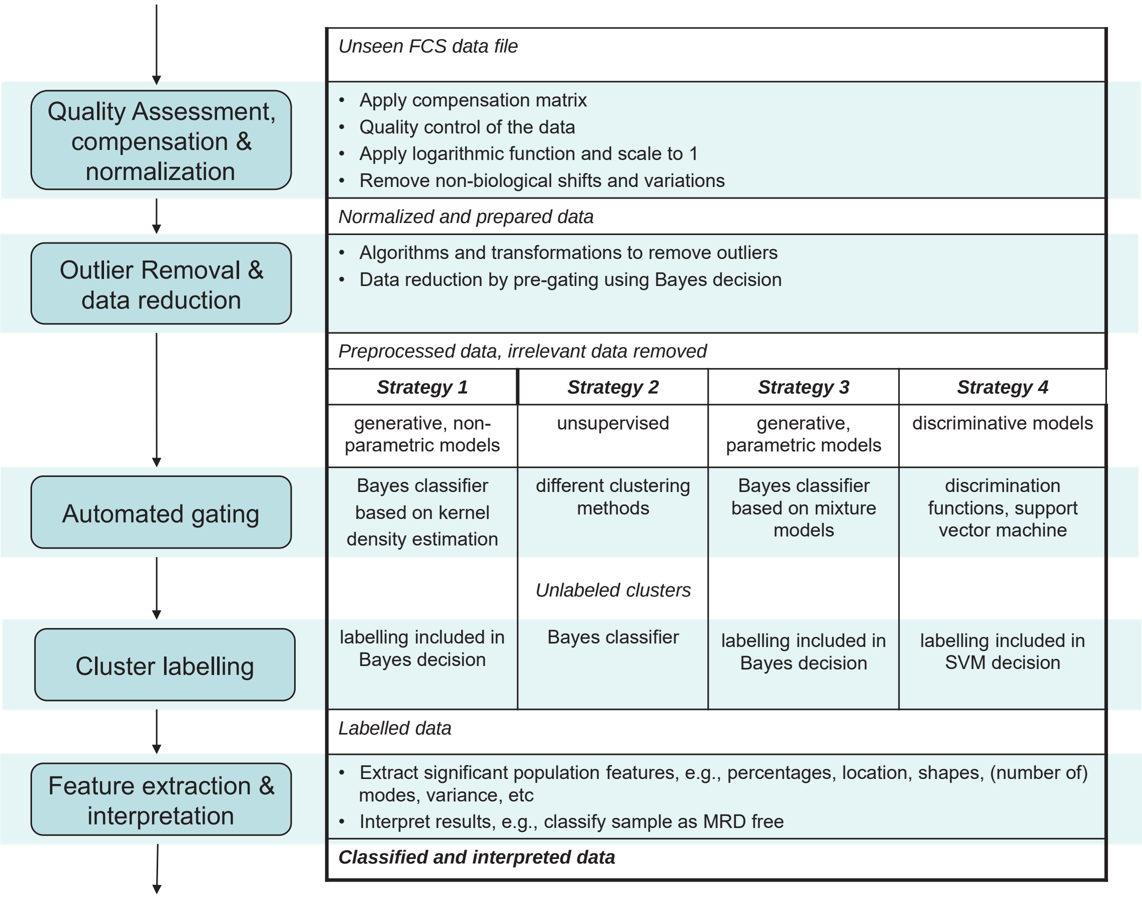
During the conception phase of the platform, special consideration has been paid on the security and the privacy issues. Therefore, session models, high-level data encryption methods, authentication, authorization control and methods ensuring confidentiality have been designed and implemented. Furthermore, medical data and personal data are encrypted and stored separately. The platform includes an authorization mechanism that controls the access. This function sets the authorization to applications, services and some areas of the data according to the user category and privacy constraints.

Figs. 2, 3 illustrate the overall concept of the telemedical network. In the implementation phase, several components have been developed:

- The front-end or client user interface, which enables users interact with the system. A desktop client version (FlowVIEW) and a web-based one have been developed in order to get benefit from both technologies.

Open Science Index, Medical and Health Sciences Vol:10, No:9, 2016 publications.waset.org/10005624.pdf

- Database management system which serves as a secure central storage place.
- Server modules which implements several functionalities of the system that can be called by client software. These backend modules are implemented as web-services.
- Particular attention has been paid to the conception and development of features which enable a high data security and privacy according to the regulations of European countries.
- Implementation of the communication interfaces, which enable a network of labs to work with the system.



Unseen FCS data file

- Apply compensation matrix
- Quality control of the data
- Apply logarithmic function and scale to 1
- Remove non-biological shifts and variations

Normalized and prepared data

- Algorithms and transformations to remove outliers
- Data reduction by pre-gating using Bayes decision

Preprocessed data, irrelevant data removed

Strategy 1	***Strategy 2***	***Strategy 3***	***Strategy 4***
generative, non-parametric models	unsupervised	generative, parametric models	discriminative models
Bayes classifier based on kernel density estimation	different clustering methods	Bayes classifier based on mixture models	discrimination functions, support vector machine
	Unlabeled clusters		
labelling included in Bayes decision	Bayes classifier	labelling included in Bayes decision	labelling included in SVM decision

Labelled data

- Extract significant population features, e.g., percentages, location, shapes, (number of) modes, variance, etc
- Interpret results, e.g., classify sample as MRD free

Classified and interpreted data

Fig. 1 Workflow for different strategies of automated FCM analysis

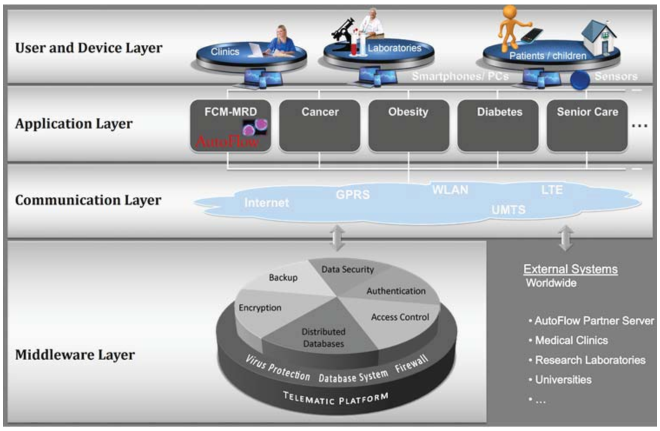


Fig. 2 Layered architecture of the system

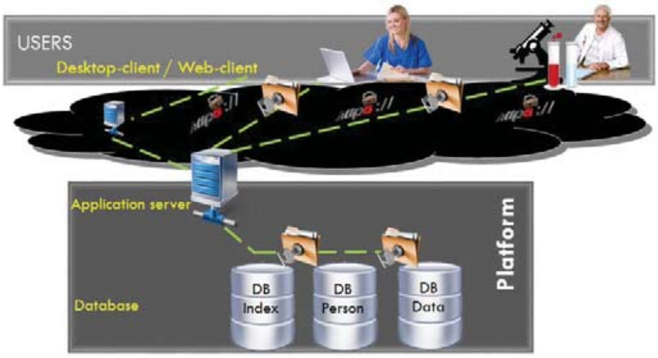


Fig. 3 Concept of the telemedical network

IV. CONCEPTION OF THE HIERARCHICAL USER MANAGEMENT

The system has different levels of users: administrative and user levels. The Administrator of the System (Admin system) has the right to add new labs into the system and create administration account for each lab. This gives the Admin lab the right to add medical experts/clinicians who will have the right to analyse experiments, they also have the right to assign a group of experiments to a clinician and define the period for which the clinician can access these data. The clinician can add new experiments and have the right to analyse a group of experiments assigned to him, concept is illustrated in Fig. 4.

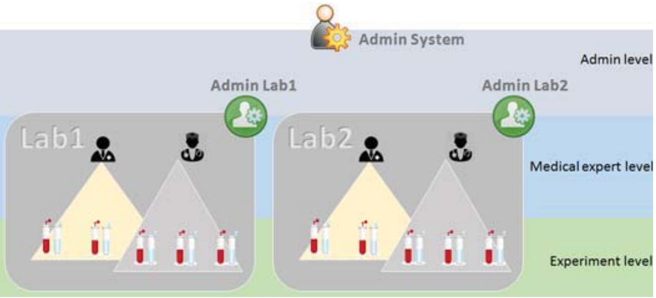
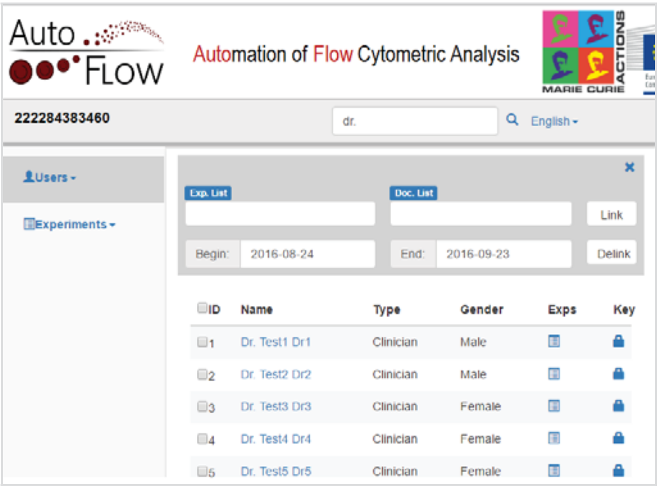


Fig. 4 User management and rights system

The Client which was developed to fulfil these user management tasks is the User Management (FlowADMIN), Figs. 5, 6. Depending on whether the account category is Admin System or Admin Lab, corresponding functionalities described above are made available.



222284383460

dr.

Begin: 2016-08-24

End: 2016-09-23

☐ID	Name	Type	Gender	Exps	Key
☐1	Dr. Test1 Dr1	Clinician	Male	▤	🔒
☐2	Dr. Test2 Dr2	Clinician	Male	▤	🔒
☐3	Dr. Test3 Dr3	Clinician	Female	▤	🔒
☐4	Dr. Test4 Dr4	Clinician	Female	▤	🔒
☐5	Dr. Test5 Dr5	Clinician	Female	▤	🔒

Fig. 5 Screenshot of the User Management Client FlowADMIN

V. CLINICIANS CLIENT APPLICATIONS OF THE SYSTEM

A. Block Structure of the Telemedical Network

The system offers two versions of clients for clinicians in order to remain flexible and take benefits of both desktop and web technologies. Moreover, a user management client is developed using new web technologies HTML5, CSS3 and JavaScript and following the responsive design methods in order to have an application adapted for all screen sizes of desktop PCs and mobile phone devices. Design responds to user's behaviour and environment based on screen size, platform and orientation.

The Telematics platform is conceived in several modules: The FlowONLINE Server Module together with its database DB2 (Fig. 6) which are responsible for data representation on the FlowONLINE is kept separately from the main Application server connected to DB1 which is more focused on administrative tasks. This way, medical data in DB2 and personal data combined with administrative ones in DB1 are stored separately.

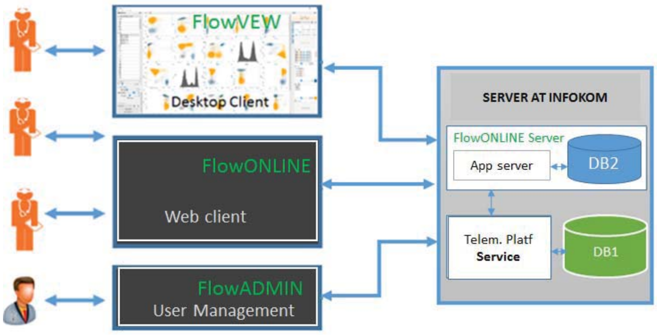


Fig. 6 Block structure of the telemedical Network

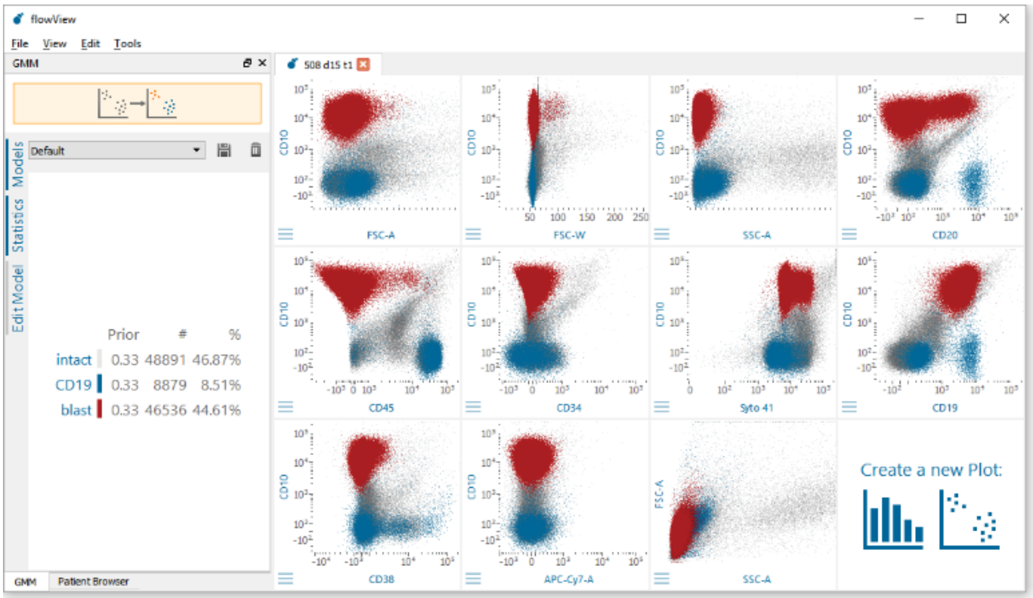
B. Desktop Client for Clinicians-FlowVIEW

FlowVIEW is an interactive visualization software for AutoGating [14]. Users can load FCM data and visualize it similar to conventional software such as Kaluza Analysis. While the number of Blasts and therefore the MRD is nowadays determined through a manual process called Gating, FlowVIEW allows for AutoGating. Hence, users need to push a button which triggers an automated labelling process. The classification of cells into Blasts and Non-Blasts is based on aligning GMMs sampled from the new population with multiple GMMs from the training set [15], [16].

Fig. 7 shows a screenshot of FlowVIEW. Blasts are colored in red while Viable and CD19+ cells are colored gray and blue respectively. The left AutoGating panel shows the statistics of all three populations if the model's priors are set to default (0.33). Functionalities of the developed client FlowVIEW can be watched at the link [18], [19].

Open Science Index, Medical and Health Sciences Vol:10, No:9, 2016 publications.waset.org/10005624.pdf

Open Science Index, Medical and Health Sciences Vol:10, No:9, 2016 publications.waset.org/10005624.pdf



	Prior	#	%
intact	0.33	48891	46.87%
CD19	0.33	8879	8.51%
blast	0.33	46536	44.61%

Fig. 7 Screenshot of the desktop client for clinicians FlowVIEW

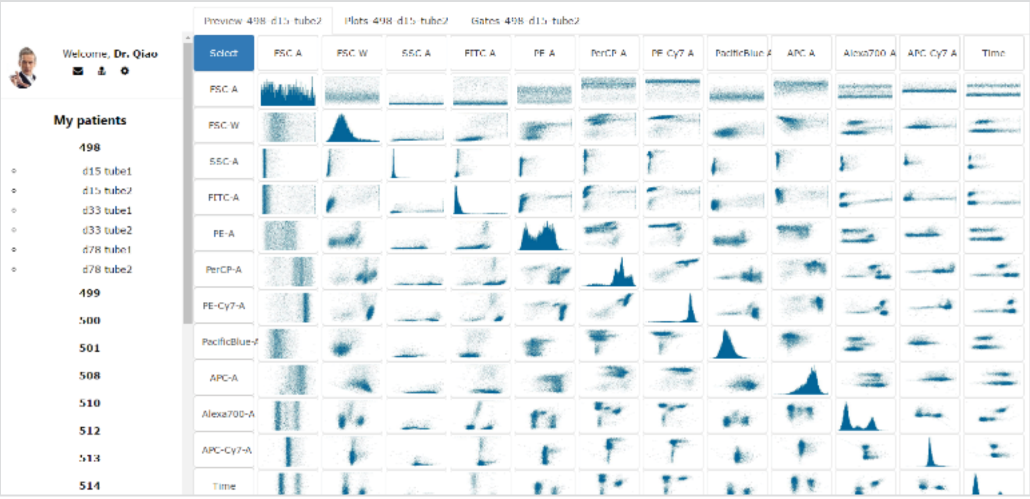


Fig. 8 Screenshot of the web client for clinicians FlowONLINE

C. Web Client for Clinician – FlowONLINE

FlowONLINE is an online platform for AutoGating and visualization of FCS data, Fig. 8. Clinicians could access to this platform when Internet is available without the effort of application installation. They are authorized to view the AutoGating result and plots of their patients, give comments or confirmations about the AutoGating results, and share opinions among different clinicians. The evolution and treatment of individual patient could also be easily traced through the platform.

The backend of FlowONLINE is hosted by Python Django (FlowONLINE Server, Fig. 6) connected with a SQLite database (DB2, Fig. 6) on the same server. The computation and visualization of FCS data is conducted by a module called FlowVIEWEx, which is a Python wrapper for FlowVIEW C++ libraries, developed for this project. Sharing the same library by the desktop client FlowVIEW and the Web Client FlowONLINE, (Fig. 6) reduces needed effort to produce identical results in both clients. The frontend of FlowONLINE is implemented using modern web technologies including HTML5, jQuery, Bootstrap, JavaScript and CSS3.

VI. CONCLUSION

A concept and realization of pipeline-software, as a ready-to-use tool for daily clinical practice for Automated Flow Cytometric Analysis for the Clinical Follow-up of patients with ALL has been presented in this paper.

The fully automated FCM-data analysis was designed and implemented in the project, the prototype software was tested, validated and optimized. In this paper, the processing prototype has been integrated into deployable telemedical software. In order to achieve the proposed objectives, a multidisciplinary network of professionals have been involved into this work.

Open Science Index, Medical and Health Sciences Vol:10, No:9, 2016 publications.waset.org/10005624.pdf

The implementation as a telemedical network saves processing time, effort, reduces costs and errors due to operator-based data interpretation. It enables distribution of automated FCM-MRD analysis to laboratories and other institutions with less expertise in the field.

ACKNOWLEDGMENT

This paper is written as the result of the AutoFLOW project, funded by Marie Curie Industry Academia Partnerships & Pathways (FP7-Marie Curie–PEOPLE-2013-IAPP) under the grant no. 610872.

REFERENCES

[1] S. Holzmüller-Laue, et. al., "A Highly Scalable Information System as Extendable Framework Solution for Medical R&D Projects. *XXIInd International Congress of the European Federation for Medical Informatics*, ISBN 978-1-60750-044-5.

[2] R-D. Berndt, C. Takenga et. al. “SaaS-Platform for Mobile Health Applications”, in *Proceedings of the IEEE-International Multi-Conference on Systems, Signals and Devices*, Chemnitz, Germany March 2012, DOI: 10.1109/SSD.2012.6198120, pp.1-4.

[3] J. Toedling, P. Rhein, R. Ratei, L. Karawajew, R. Spang (2006) Automated in-silico detection of cell populations in FCM cytometry read-outs and its application to leukaemia disease monitoring. *BMC Bioinformatics*, 7:282.

[4] K. Fiser, T. Sieger, A. Schumich, B. Wood, J. Irving, E. Mejstrikova, M. Dworzak. Detection and Monitoring of Normal and Leukemic Cell Populations with Hierarchical Clustering of FCM Cytometry Data (2011). *Cytometry A*. 2011.

[5] K. Lo, RR. Brinkman and R. Gottardo, Automated Gating of FCM Cytometry Data via Robust Model-based Clustering (2008). *Cytometry A*. 2008 Apr;73(4):321-32.

[6] S. Pyne, X. Hu, K. Wang, E. Rossin, TI. Lin, KM. Maier, C. Baecher-Allan, GJ. McLachlan, P. Tamayo, DA. Hafler, PL. De Jager, JP. Mesirov (2009). Automated high-dimensional flow cytometric data analysis. *Proceedings of the National Academy of Sciences of the United States of America,* Volume: 106, Issue: 21, Publisher: National Academy of Sciences; 8519-8524.

[7] CE. Pedreira, ES. Costa, J. Almeida, C. Fernandez, S. Quijano, J. Flores, S. Barrena, Q. Lecrevisse, JJ. Van Dongen, A. Orfao; EuroFCM Consortium (2008). A probabilistic approach for the evaluation of minimal residual disease by multiparameter FCM cytometry in leukemic B-cell chronic lymphoproliferative disorders. *Cytometry Part A* Volume 73A, Issue 12; 1141–1150.

[8] C. Chan, F. Feng, J. Ottinger, D. Foster, M. West, TB. Kepler. Statistical mixture modeling for cell subtype identification in FCM cytometry. *Cytometry A*. 2008 Aug;73(8): 693-701.

[9] M. Wilkins, L. Boddy, C. Morris, R. Jonker. A comparison of some neural and non-neural methods for identification of phytoplankton from FCM *cytometry data. Comput Appl Biosci* (1996) 12(1): doi:10.1093/bioinformatics/12.1.9, pp 9-18.

[10] K. Lo, F. Hahne, R. Brinkman, R. Gottardo. FCMClust: A Bioconductor package for automated gating of FCM cytometry data (2009). *BMC Bioinformatics* 2009, 10:145 doi:10.1186/1471-2105-10-145.

[11] J. Frelinger, TB. Kepler, C. Chan, FCM: Statistics, visualization and informatics for FCM cytometry (2008). *Source Code for Biology and Medicine* 2008, 3:10 doi:10.1186/1751-0473-3-10.

[12] A. Bashashati, R. Brinkman (2009). A Survey of FCM Cytometry Data Analysis Methods. *Advances in Bioinformatics*. Volume 2009, Article ID 584603.

[13] P. Rota, S. Groeneveld-Krentz, M. Reiter. On Automated Flow Cytometric Analysis for MRD Estimation of Acute Lymphoblastic Leukaemia: A Comparison Among Different Approaches, *The IEEE International Conference on Bioinformatics and Biomedicine* (BIBM) 2015.

[14] M. Reiter, M. Gau, P. Rota, F. Kleber, S. Groeneveld-Krentz, A. Schumich, M. Dworzak: An Automated Flowcytometry Data Analysis Support System, *CYTO* 2015, Glasgow.

[15] M. Reiter, J. Hoffmann, F. Kleber, A. Schumich, G. Peter, M. Kampel, M. Dworzak Towards Automation of Flow Cytometric Analysis for Quality-Assured Follow-up Assessment to Guide Curative Therapy for Acute Lymphoblastic Leukaemia in Children, memo – *Magazine of European Medical Oncology* 7(4). 2014.

[16] F. Kromp, M. Reiter, S. Taschner-Mandl, P. Ambros, A. Hanbury. Classification of cellular populations using Image Scatter-Plots. *Proceedings of the 20th Computer Vision Winter Workshop*, p. 113–20, February 9 – 11, 2015. Seggau, Austria; 2015.

[17] P. Rota, F. Kleber, M. Reiter, S. Groeneveld-Krentz and M. Kampel, The Role of Machine Learning in Medical Data Analysis. A Case Study: Flow Cytometry, in the proceedings *of International Conference on Computer Vision Theory and Application VISIGRAPP*, 2016, Rome.

[18] FlowVIEW Functionalities: https://www.youtube.com/watch?v=fu0V76Cppa4 Accessed (March 29th, 2016).

[19] M. Reiter, F. Kleber, J. Hoffmann, M. Dworzak: “Automation of MRD Measurements in Flow Cytometry to Guide Curative Therapy for ALL in Children”. *4th Munich Biomarker Conference. Munich*. 2014.