Tackling Data Related Challenges in Healthcare Process Mining using Visual Analytics

[1] Kennedy O. Ondimu; [2] Kelvin K. Omieno; [3] Geoffrey M. Muchiri; [4] Ismael A. Lukandu

[1] Institute of Computing and Informatics, Technical University of Mombasa
Box 90420 -8011 Mombasa, Kenya

[2] School of Computing and Informatics, Masinde Muliro University of Science and Technology
Box 190-50100, Kakamega, Kenya

[3] School of Computing and Information Technology, Murang'a University of Technology
Box 75-10200 Muranga, Kenya

[4] Faculty of information Technology, Strathmore University
Box 59857-00200 Nairobi, Kenya

Abstract - Data-science approaches such as Visual analytics tend to be process blind whereas process-science approaches such as process mining tend to be model-driven without considering the "evidence" hidden in the data. Use of either approach separately faces limitations in analysis of healthcare data. Visual analytics allows humans to exploit their perceptual and cognitive capabilities in processing data, while process mining represents the data in terms of activities and resources thereby giving a complete process picture. We use a literature survey on both Visual analytics and process mining in the healthcare environments, to discover strengths that can help solve open problems in healthcare data when using process mining. We present a visual analytics approach in solving data challenges in healthcare process mining. Historical data (event logs) obtained from organizational archives are used to generate accurate and evidence based activity sequences that are manipulated and analyzed to answer questions that could not be tackled by process mining. The approach can help hospital management and clinicians among others, audit their business processes in addition to providing important operational information. Other beneficiaries include those organizations interested in forensic information regarding individuals and groups of patients.

Keywords: *Healthcare, visual analytics, process mining, challenges*

1. Introduction

In some countries such as the Netherlands, it is a requirement that healthcare procedures be standardized and auditable [1]. Unlike in industry such as manufacturing, healthcare procedures are unique. In healthcare, individual actors, say doctors, though initially trained to treat patients using specific procedures and drugs, the procedures and drugs are continuously changing; likewise individual patients may present non-standard symptoms or suffer complications. Interest in healthcare is beyond individual hospitals. Organizations such as insurance companies, governments and even international organizations are interested in an efficient healthcare system. In the United States of America, healthcare accounts for 17% of the GDP and employs 11% of the countries workers, yet its costs have been growing at 5% in the last decade making it unsustainable and a major contributor to the high national debt levels projected in the next two decades [2]. One of the leading contributors to these situation is that little has been invested in healthcare efficiency optimization [2].

Many authors including [17] categorize healthcare processes into two: (1) Organizational/Administrative processes - include resources (staff, machines and equipment) and their scheduling. (2) Clinical/Medical Processes - these constitute the decisions made by qualified clinicians on the treatment of a patient such as diagnosis, laboratory tests, imaging, prescriptions and drug administration, discharge and follow-up.

Clinical and Administrative processes in healthcare have high complexity due to the many decisions and procedures captured, are highly dynamic, increasingly multidisciplinary and ad hoc; hence cannot fit in a-priory model. The only accurate way to know what is happening is to mine the process model from historical data. On the other hand, optimization or improvement of a process model alone cannot improve the process due to dynamism of procedures and other factors.

Process mining is able to build a process from historical data that is accurate and evidence based. It is *good at portraying processes that have many vertices (activities) and flow lines (edges)*. However, Process mining is unable to deal with complexity, resulting in unreadable process models and does not portly what is happening to individual records or even groups of record [3].

Visual analytics (VA) is defined as "combining automated analysis techniques with interactive visualizations for an effective understanding, reasoning and decision making on the basis of very large and complex data sets" [4]. The technique is human-centered, whereby a human solves a complex problem with the help of a computer. It is an important candidate for handling information overloads and complexity as well as being interactive [5]. The disadvantage of Visual analytics is that it is process blind (e.g. event sequences do not show the relationship between different variants say sharing of resources, convergence of work, parallelism, et cetera). However, it is good at portraying the "evidence" hidden in the data e.g longest paths, variants, et cetera.

This paper's aim is to address three data challenges when answering frequently posed questions in healthcare process mining [6]. They include:

RQ1: What are the most followed paths and what exceptional paths are followed?

RQ2: Are there differences in care paths followed by different patient groups?

RQ3: Do we comply with internal and external guidelines?

2. Related Literature

A search of literature using the keywords visual analytics, process mining, healthcare and challenges resulted in four categories of reports including: (1) those that involve both visual analytics and process mining, (2) visual analytics; (3) process mining and (4) other interesting findings.

Among those that employed both visual analytics and process mining is [3] and [7], the later being validation of the formers work. Riemers and Ramos did not address any of the three research questions but their work generally describes and recommends a combined approach of the two technologies that is also recommended widely in the process mining community. A study by [8] lists a set of six open challenges in process mining four of which can be mapped on to the three research questions in this study. The challenge of combining the purely automatic analysis associated with process mining and visualization methods has been highlighted by a number of researchers [9] [10] [11][12].

Some research reports that deployed visual analytics and used electronic health records or log data don't describe the data extraction method [10] [11][12]. However they mention the need for a combined approach to problem solving as Monroe went ahead to develop Eventflow toolkit. Other research that focused on healthcare used visual analytics and data mining instead of process mining [13][14]. This approach also used by the IBM Watson research is based on data mining and is used to answer predetermined questions.

In a study to apply process mining in healthcare, [15] used process mining without visual analytics. They however ended up recommending development of new techniques and or using existing techniques in an innovative way to obtain understandable high level information instead of the spaghetti-like models showing all details. A similar suggestion to handle abstraction and complexity using integration, visualization and interaction though not based on healthcare or the two technologies was made by [16]. Visual analytics has capacity to handle [16] suggestions on abstraction and complexity. Research involving healthcare workflows that employed process mining without visual analytics identified one of the challenges to be the inability of process mining to portray one-to-many and many-to-many relationships between Proclets e.g. mini-processes [17], a challenge that can to some extent be handled by visual analytics.

The combination of visual data exploration with process mining algorithms makes complex information structures more comprehensible and facilitates new insights [8]. Data-science approaches tend to be process blind (e.g. event sequences do not show the relationship between different variants say sharing of resources, convergence of work etc) whereas process-science approaches tend to be model-driven without considering the "evidence" hidden in the data (e.g longest paths, variants, etc). *VA is good at handling abstraction and complexity in data, it is not ideal for portraying processes that have so many vertices (activities) and flow lines (edges).* The implication is that data science may not be of much use in process issues such as control flow or dealing with concurrency but can support in issues to do with aggregation and abstraction and sequence. Likewise process science may not be that efficient in abstraction aggregation or sequence issues but can help in showing the relationship between different data elements and activities. For example, process science can help show the control flow issues such as bottlenecks while data science can help analyze the data around it. Indeed VA can be used to answer some of the data challenges encountered in application of process mining in healthcare as posed in [6].

Obviously, both worlds need to be connected and integrated [18]. ProM an open source framework available from www.processmining.org and Eventflow toolkit developed by University of Maryland are among tools that are applicable to process mining and visual analytics respectively.

On the other hand research on visual analytics also reports a number of challenges. Appropriate combination of the strengths of intelligent automatic data analysis with the visual perception and analysis capabilities of the human user has been recommended [19]. Kohlhammer further underscores the need for faster new solutions to highly complex problems, as targeted by visual analytics, through the promotion of interdisciplinary research. He also calls for use of incremental improvements to generate broader support in various application fields using visual analytics. Typical data analysis using visual analysis involves search for known patterns from data and prediction of unexpected information remains a challenging problem [20]. Finally there is need for integrative solutions for interactive visualization of the data. Although there are many sophisticated results and paradigms from the visualization community, integrated solutions, e.g. within business hospital information systems, are rare today [21]. Most of the challenges highlighted with regard to visual analytics can be tackled by applying visual analytics to solving the three process mining data challenges by [6].

3. Method

To answer the research questions, we use event-log data from the Radiology department of the AMC University teaching hospital, a Dutch hospital of the Netherlands that is freely available at www.processmining.org.

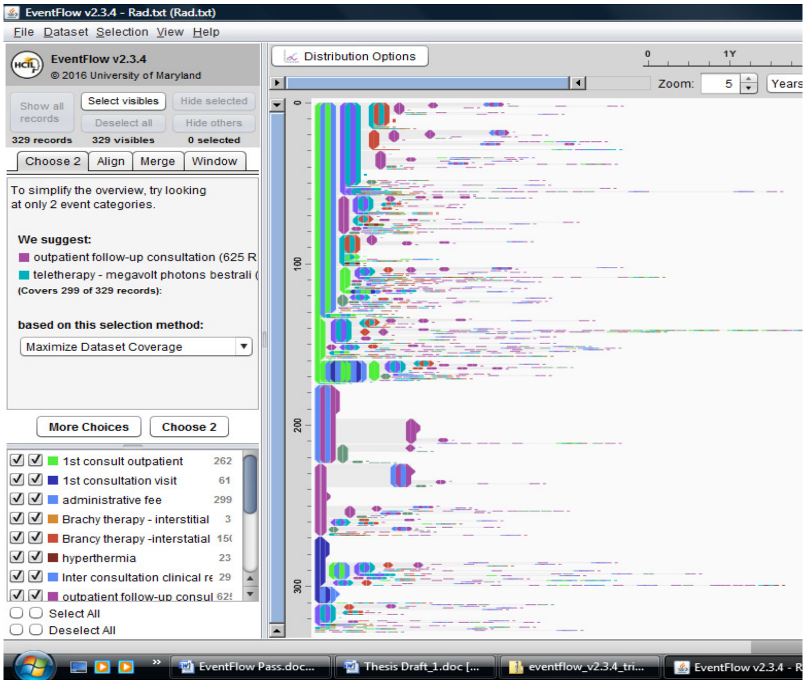
The event-log was converted from XSS and saved as MXML for ProM and txt for Eventflow. Other departments were filtered out to obtain event-logs relating to radiotherapy department only. As with most hospital information systems a number of common issues were noted [6]. Some of the issues could be addressed yet others had no immediate solution as follows: (1) event-log has average abstraction e.g. one event referring to an individual task which is a positive factor. (2) accuracy of the data viewed in three sub dimensions was as follows: the data had average granularity (hours) with the resources in block form e.g. RATH and SRTH; low directness of registration (timestamps are registered manually e.g. almost all end at 2:00 pm - the next day); and low correctness (timestamps not logged correctly in the chosen granularity e.g. mixed date formats, some start times coming after the end time), duplication of records and single entries that do not constitute a procedure or trace in a process.

Apparently, such data challenges are a problem common in both VA and process mining [8]. This could be attributed to the many actors, some who do not follow standard procedures, responsible for entering data into the systems that store such data. Each event in the data had a CaseID, TaskID, Start timestamp, End timestamp, diagnosis Code and resource. The event-log had 329 cases consisting of 4322 events. Using Eventflow and the radiology event-log, the three research questions are answered as follows.

RQ1: What are the most followed paths and what exceptional paths are followed

3.1 Most followed paths

The data that was earlier on saved as text is used as input to Eventflow. A raw event-log view of radiology department is generated representing all the information, some of which is unnecessary and too complex for the eye to pick out required information. The activities of each record/case are represented taking into account the timestamps on the overview panel as shown in Figure 1. The timeline panel shows the same information in a format similar to a Gantt chart



To simplify the overview, try looking at only 2 event categories.

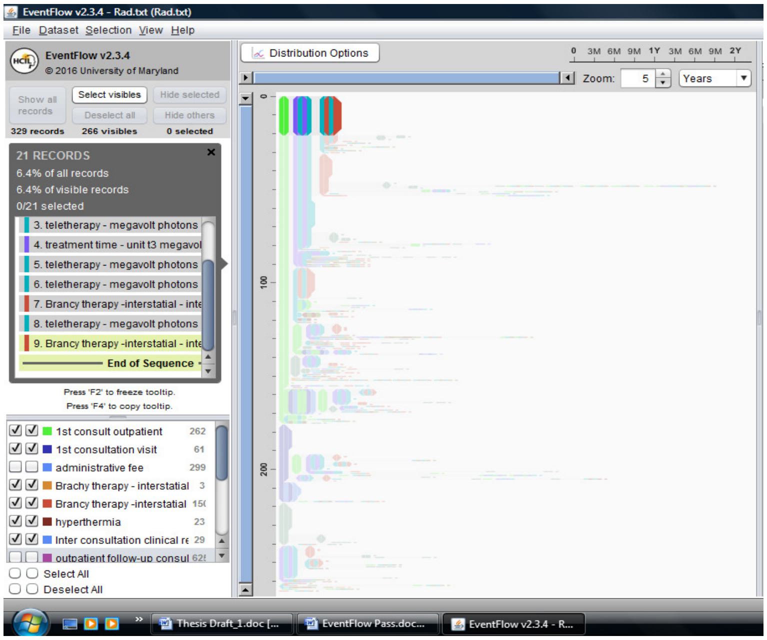
We suggest:
🟪 outpatient follow-up consultation (625 R
🟦 teletherapy - megavolt photons bestrali (
(Covers 299 of 329 records):

based on this selection method:
Maximize Dataset Coverage

☑	☑	🟩	1st consult outpatient	262
☑	☑	🟦	1st consultation visit	61
☑	☑	🟦	administrative fee	299
☑	☑	🟧	Brachy therapy - interstitial	3
☑	☑	🟥	Brancy therapy -interstatial	15(
☑	☑	🟫	hyperthermia	23
☑	☑	🟦	Inter consultation clinical re	29
☑	☑	🟪	outpatient follow-up consul	62!

Fig. 1: Un-optimized Radiology department showing all case activities

To reduce the complexity, administrative fee and outpatient follow-up consultation were checked off since they either don't constitute treatment or it is common to almost all patients at the end hence cannot help in discriminating between traces. This reduces the overview complexity of the view, placing the most common sequence at the top as depicted in Figure 2. The overview panel is ranked based on number of records in ascending order.



21 RECORDS
6.4% of all records
6.4% of visible records
0/21 selected
3. teletherapy - megavolt photons
4. treatment time - unit t3 megavol
5. teletherapy - megavolt photons
6. teletherapy - megavolt photons
7. Brancy therapy -interstatial - inte
8. teletherapy - megavolt photons
9. Brancy therapy -interstatial - inte
End of Sequence

Press 'F2' to freeze tooltip.
Press 'F4' to copy tooltip.

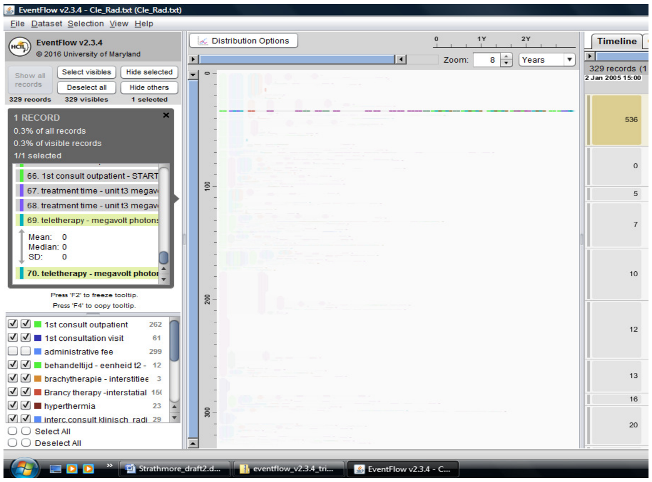
☑	☑	🟩	1st consult outpatient	262
☑	☑	🟦	1st consultation visit	61
☐	☐	🟦	administrative fee	299
☑	☑	🟧	Brachy therapy - interstatial	3
☑	☑	🟥	Brancy therapy -interstatial	15(
☑	☑	🟫	hyperthermia	23
☑	☑	🟦	Inter consultation clinical re	29
☐	☐	🟪	outpatient follow-up consul	62!

Fig.2: Most followed sequence/trace in the radiology process

The events in the trace consist of: (1) 1st consultation outpatient, (2) treatment time unit t3 megavolt, (3) Teletherapy – megavolt photons bestrali and Branchy therapy – interstitial – intensi. On completion, the sequence/trace constitutes 6.4% of all records.

3.2 Exceptional paths

By checking off administrative fee which does not constitute treatment complexity is reduced. The view presents a number of mostly single records that are exceptionally long such as case number 536 that is highlighted with 70 events in figure 3. Others can consist of single events without any other service following. Exceptional paths are usually of special interest to the process owners.


EventFlow v2.3.4 - Cle_Rad.txt (Cle_Rad.txt)

1 RECORD
0.3% of all records
0.3% of visible records
1/1 selected

66. 1st consult outpatient - START
67. treatment time - unit t3 megav
68. treatment time - unit t3 megav
69. teletherapy - megavolt photons
Mean: 0
Median: 0
SD: 0
70. teletherapy - megavolt photo

Press 'F2' to freeze tooltip.
Press 'F4' to copy tooltip.

☑ ☑ 1st consult outpatient 262
☑ ☑ 1st consultation visit 61
☐ ☐ administrative fee 299
☑ ☑ behandeltijd - eenheid t2 - 12
☑ ☑ brachytherapie - interstitiee 3
☑ ☑ Brancy therapy -interstatial 15(
☑ ☑ hyperthermia 23
☑ ☑ interc.consult klinisch radi 29
○ ○ Select All
○ ○ Deselect All

Fig 3: Exceptional Record number 536

RQ2: Different care paths followed by different patient groups

Different patient and groups often follow different care paths despite having the same diagnosis. To generate the view, consultation and administrative fee categories are checked off to reduce complexity since they are only used to determine diagnosis and payment rather than actual treatment. Creation of such a view requires an attribute file that matches the cases with the specific diagnosis. All case are then grouped according to diagnosis code as shown in Figure4.

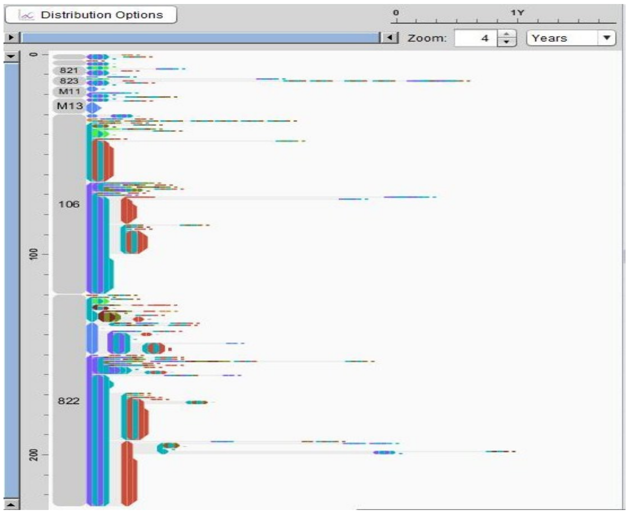


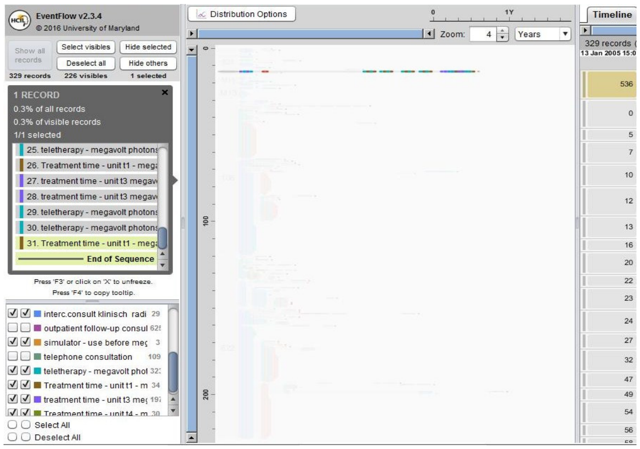
Figure 4: Different care paths for different patient groups having same diagnosis

In Figure 4, the largest number of patients is in diagnosis code 822 with 106 cases, followed by diagnosis code 106 with 90 cases, the rest being less than 8 cases each. Both diagnosis codes 822 and 106 split into more than 10 care paths each.

RQ3: Compliance with internal and external guidelines

Eventflow has sufficient capacity to present either group or individual cases with regard to time and activity sequence. It is for example possible to determine the time that a particular activity took place and the time between various activities in an individual case or same variant sequence by checking off the distribution options.

A number of decisions can be made just by checking if the patient actually entered process from the first activity or exited after the last activity as expected. Such information is revealed by pointing the curser over the particular record and either taking it to the first of last activity which will be displayed on the control panel as shown in Figure 5.



329 records 226 visibles 1 selected

4 Years

329 records (
13 Jan 2005 15:0

536
0
5
7
10
12
13
16
20
22
23
24
27
32
47
49
54
56

1 RECORD
0.3% of all records
0.3% of visible records
1/1 selected

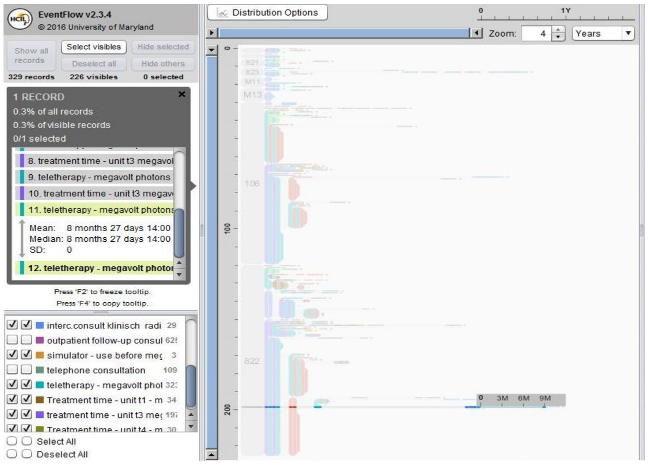
25. teletherapy - megavolt photons
26. Treatment time - unit t1 - mega
27. treatment time - unit t3 megav
28. treatment time - unit t3 megav
29. teletherapy - megavolt photons
30. teletherapy - megavolt photons
31. Treatment time - unit t1 - mega
End of Sequence

Press 'F3' or click on 'X' to unfreeze.
Press 'F4' to copy tooltip.

☑ ☑ 🟦 interc.consult klinisch radi 29
□ □ 🟪 outpatient follow-up consul 625
☑ ☑ 🟧 simulator - use before meg 3
□ □ 🟩 telephone consultation 109
☑ ☑ 🟦 teletherapy - megavolt phot 323
☑ ☑ 🟫 Treatment time - unit t1 - m 34
☑ ☑ 🟪 treatment time - unit t3 meg 197
☑ ☑ 🟩 Treatment time - unit t4 - m 30
○ ○ Select All
○ ○ Deselect All

Fig 5: Last activity in a record sequence

The period between two adjacent or non-adjacent activities in a sequence can also be revealed as shown in Figure 6 on the overview and control panels.



329 records 226 visibles 0 selected

4 Years

1 RECORD
0.3% of all records
0.3% of visible records
0/1 selected

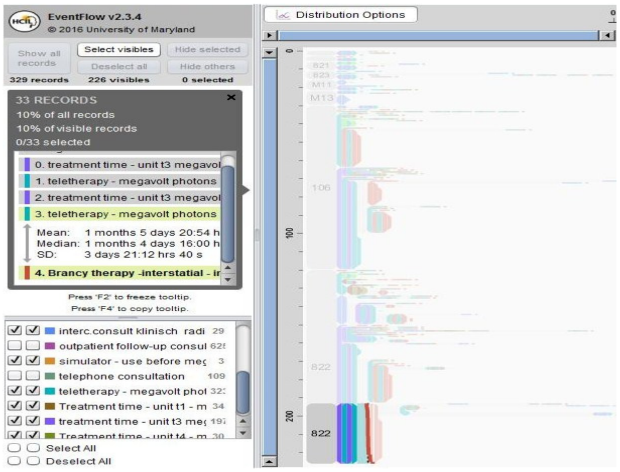
8. treatment time - unit t3 megavol
9. teletherapy - megavolt photons
10. treatment time - unit t3 megav
11. teletherapy - megavolt photons
Mean: 8 months 27 days 14:00
Median: 8 months 27 days 14:00
SD: 0
12. teletherapy - megavolt photor

Press 'F2' to freeze tooltip.
Press 'F4' to copy tooltip.

☑ ☑ 🟦 interc.consult klinisch radi 29
□ □ 🟪 outpatient follow-up consul 625
☑ ☑ 🟧 simulator - use before meg 3
□ □ 🟩 telephone consultation 109
☑ ☑ 🟦 teletherapy - megavolt phot 323
☑ ☑ 🟫 Treatment time - unit t1 - m 34
☑ ☑ 🟪 treatment time - unit t3 meg 197
☑ ☑ 🟩 Treatment time - unit t4 - m 30
○ ○ Select All
○ ○ Deselect All

Fig 6: Period between two activities in a sequence.

The two treatment activities in question are highlighted in the overview panel and from the tooltip as in figure 7. The treatment activities teletherapy – megavolt photons bestrali and Branchy therapy – interstitial-intensi have a mean time of 1 month 5 days 20.54 hrs, a median of 1 month 4 days 16:00 hrs and SD of 3 days 21:12 hrs 40 s between them. The lengths of the individual activities are got from the control panel.



329 records 226 visibles 0 selected

33 RECORDS
10% of all records
10% of visible records
0/33 selected

0. treatment time - unit t3 megavol
1. teletherapy - megavolt photons
2. treatment time - unit t3 megavol
3. teletherapy - megavolt photons
Mean: 1 months 5 days 20:54 h
Median: 1 months 4 days 16:00 h
SD: 3 days 21:12 hrs 40 s
4. Brancy therapy -interstatial - ir

Press 'F2' to freeze tooltip.
Press 'F4' to copy tooltip.

☑ ☑ 🟦 interc.consult klinisch radi 29
□ □ 🟪 outpatient follow-up consul 625
☑ ☑ 🟧 simulator - use before meg 3
□ □ 🟩 telephone consultation 109
☑ ☑ 🟦 teletherapy - megavolt phot 323
☑ ☑ 🟫 Treatment time - unit t1 - m 34
☑ ☑ 🟪 treatment time - unit t3 meg 197
☑ ☑ 🟩 Treatment time - unit t4 - m 30
○ ○ Select All
○ ○ Deselect All

Fig 7: Distribution information between two activities in treatment

4. Discussions

Visual analytics overcomes the challenge of identification of the most followed path/trace that is encountered in process mining. This is achieved by grouping all records with similar sequences together and by sorting; the most followed sequence will be placed either at the top or bottom. In addition to identification of the most followed path/trace, more information on the event-log is available by pointing at the particular sequence and reading the information displayed on the control panel. It should be noted that cases that enter an activity (START) are often split, such that the ones that complete the activity (END) are different in number as regards a particular trace. For example in figure 2, Branchy therapy – interstitial – intense starts with 7.9% of all records but only 6.4% complete the process in the most followed trace.

The earlier activities in a trace are part of a higher percentage of records since they are also part of other sequences such as: (1) 1st consultation outpatient (53.5% of all records), (2) treatment time unit t3 megavolt (28% of all records), (3) Teletherapy – megavolt photons bestrali: (25.2% of all records) and (4) Branchy therapy – interstitial – intense (7.9% of all records). The implication is that changes in the earlier activities of a most followed trace have a higher impact or weight compared to the later activities.

Exceptional paths that are depicted by very long sequence represent either very complex cases, or an indication of ineffective treatment that should be of concern to the heath care facility. Such may indicate misdiagnosis, wrong prescription, resistance to drugs or a new disease.

A similar diagnosis, different care path, is not surprising since despite a common diagnosis, individual peculiarities of the patient determine the care path. However, the variation in care paths is expected to reflect some categorization such as age sets, and progression level of ailment (mild, intermediate, and severe). Such revelations can help management to restructure the section into subsections handing the different care paths if the numbers are significant such as in diagnosis code 822 and 106.

Questions pertaining to compliance to internal and external guidelines can be answered using visual analytics. Individual records can be inspected on aspects such as order of activities in a procedure, timing between activities, first and last activities. In comparison to process mining which treats all cases as part of a process model by eliminating those that do not fit, visual analytics accounts for each case and presents an easy way to audit all cases. The possibility of inspecting each patient record can be extended to resource performance

5. Conclusions

Complexity and abstraction in healthcare data is a challenge when using process mining. Using an integrative approach, a number of previously open problems when using process mining can be solved using visual analytics. Three challenges in healthcare process mining including identification of most followed paths and exceptional paths; differences in care paths followed by different patient groups with same diagnosis; and compliance with internal and external guidelines are solvable using visual analytics. The ability of visual analytics to reveal evidence hidden in the data can also help process owners in operational running.

Acknowledgments

My acknowledgements go to the following organizations: (1) National Commission for Science, Technology and Innovation for funding this research; (2) The Human-Computer Interaction Lab, University of Maryland, for the Eventflow research licence; and (3) Fluxicon for the Disco academic licence

References

[1] J. K. Helderman, F. T. Schut, T. E. van der Grinten, W. P. van de Ven, "Market-oriented health care reforms and policy learning in the Netherlands", Journal of Health Politics, Policy and Law, Vol. 30, No.1-2, 2005, pp. 189-210.

[2] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, & A.H. Byers, "Big data: The next frontier for innovation, competition, and productivity", 2011.

[3] P. Riemers, "Process improvement in Healthcare: a data-based method using a combination of process mining and visual analytics", MSc thesis, Technology Management, Eindhoven University of Technology, Eindhoven, The Netherlands, 2009.

[4] D.A. Keim, F. Mansmann, & J. Thomas, "Visual analytics: how much visualization and how much analytics?", *ACM* SIGKDD Explorations Newsletter, Vol. 11, No. 2, 2010, pp. 5-8.

[5] D. Keim, F. Mansmann, J. Schneidewind, J. Thomas, H. Ziegler, "Visual analytics: Scope and challenges. Visual Data Mining", 2008, pp. 76-90.

[6] R.S. Mans, W.M. van der Aalst, R.J. Vanwersch, A.J. Moleman, "Process mining in healthcare: Data challenges when answering frequently posed questions", In Process Support and Knowledge Representation in Health Care, Springer Berlin Heidelberg, 2013, pp. 140-153.

[7] L. T. Ramos, "Healthcare Process Analysis: validation and improvements of a data-based method using process mining and visual analytics". MSC thesis, Operations Management and Logistics, Eindhoven University of Technology, Eindhoven, The Netherlands, 2009.

[8] T. Gschwandtner, "Visual Analytics Meets Process Mining: Challenges and Opportunities", In International Symposium on Data-Driven Process Discovery and Analysis, Springer, Cham, 2015 December, pp. 142-154.

[9] W.M. van der Aalst, M. de Leoni, A. H. ter Hofstede, "Process mining and visual analytics: Breathing life into business process models", BPM Center Report BPM-11-15, BPMcenter. org, Vol. 17, 2011, pp. 699-730.

[10] M. De Leoni, M. Adams, W.M. Van Der Aalst, A.H. Ter Hofstede, "Visual support for work assignment in process-aware information systems: Framework formalization and implementation". Decision Support Systems, Vol. 54, No. 1, 2012, pp. 345-361.

[11] M. Monroe, "Interactive Event Sequence Query and Transformation", Doctorial dissertation, Computer Science, University of Maryland, Maryland, USA, 2014.

[12] M. Ozkaynak, O. Dziadkowiec, R. Mistry, T. Callahan, Z. He, S. Deakyne, E. Tham, "Characterizing workflow for pediatric asthma patients in emergency departments using electronic health records", Journal of biomedical informatics, Vol. 57, 2015, pp. 386-398.

[13] A. Perer, D. Gotz, "Data-driven exploration of care plans for patients. In CHI'13 Extended Abstracts on Human Factors in Computing Systems", ACM, 2013 April, pp. 439-444.

[14] J.A. Fitzgerald, A. Dadich, "Using Visual Analytics to improve hospital scheduling and patient flow", Journal of Theoretical and Applied Electronic Commerce Research, Vol. 4, No. 2, 2009, pp. 20-30.

[15] R.S. Mans, M.H. Schonenberg, M. Song, W.M.P van der Aalst, P.J.M. Bakker, "Application of process mining in healthcare–a case study in a Dutch hospital", In Fred, A., Filipe, J. and Gamboa, H. (Eds.): BIOSTEC 2008, CCIS 25 Springer-Verlag Berlin Heidelberg, 2008, pp. 425- 438.

[16] G. M. Nugteren, "Process Model Simplification", MSC thesis, Business Information systems, Eindhoven University of Technology, Eindhoven, The Netherlands, 2010.

[17] R.S. Mans, "Workflow support for healthcare domain", PhD thesis, Operations Management and Logistics, Eindhoven University of Technology, Eindhoven, The Netherlands, 2011.

[18] P. Harmon, R. Tregear, W.M.P. van der Aalst, et al, "Questioning BPM: 109 Answers by 33 Authors to 15 Questions About Business Process Management" Meghan-Kiffer Press, Tampa, USA, 2016.

[19] J. Kohlhammer, D. Keim, M. Pohl, G. Santucci, G. Andrienko, "Solving problems with visual analytics", Procedia Computer Science, Vol. 7, 2011, pp. 117-120.

[20] S.H. Hong, K.L. Ma, K. Koyamada, "Big Data Visual Analytics", In NII Shonan Meeting Report, 2015 November, pp. 2186-7437.

[21] C. Turkay, F. Jeanquartier, A. Holzinger, H. Hauser, "On computationally-enhanced visual analysis of heterogeneous data and its application in biomedical informatics". In Interactive knowledge discovery and data mining in biomedical informatics, Springer, Berlin, Heidelberg. 2014, pp. 117-140.

Authors –

Kennedy O. Ondimu is a PhD candidate in information Technology at Masinde Muliro University of Science and Technology as well as a lecturer at Technical University of Mombasa. He has published a number of papers and book chapters in the area of IT. He has wide experience as a lecture, academic leadership and as director of ICT at University.

Dr. Kelvin K. Omieno is Lecturer and also Founding and Current Dean, School of Computing and Informatics (SCI), Masinde Muliro University of Science and Technology, Kenya (www.sci.mmust.ac.ke). He holds a PhD in Business Information Systems of Jaramogi Oginga Odinga University of Science & Technology (Kenya). He has MSc in Information Technology and Bachelor of Science in Computer Science (First Class Honors) from Masinde Muliro University of Science and Technology (Kenya). Dr. Omieno has been involved in a number of research projects of ICTs for Development, Data Analytics, Computational Grid Project, Health Informatics, E-learning systems and E-waste management in Kenya. Besides, he has published widely in journals and conference proceedings in Information technology and ICTs for development. He is a professional member of the Association for Computing Machinery (ACM), the largest association of computing professionals globally and is a reviewer with three International Journals.

Prof. Geoffrey M. Muketha is a professor of software engineering at Murang'a University of Technology and current Dean, school of Computing and Information Technology. He has published widely as well as supervised several graduate students at masters and PhD level. His interests are in Software metrics, automated static code analysis and structural quality of software.

Prof Ismail A. Lukandu is an Associate professor at Strathmore University, faculty of Information Technology and current Dean of research in the University. He has published widely as well as supervised several graduate students at masters and PhD level. His interests are in Information Systems, Modeling and simulation, Database marketing and Information Technology among others.