A Literature Study of Visual Analysis in an Educational Context

1st Luiza Hansen
dept. Computer Science
Universidade de Brasília
Brasília, Brazil
ORCID 0000-0002-2848-3441

2nd Vinicius R. P. Borges
dept. Computer Science
Universidade de Brasília
Brasília, Brazil
ORCID 0000-0003-1254-8420

3rd Maristela Holanda
dept. Computer Science
Universidade de Brasília
Brasília, Brazil
ORCID 0000-0002-0883-2579

***Abstract*—Research Full Paper - Visual Analytics is an emerging field that enables detection of the expected and discovery of the unexpected. This technique has been used in numerous areas such as education, where the motivation is to understand and improve the teaching and learning processes. This paper presents a systematic literature study of the field of visual analysis in an educational context in order to provide more insights into this research area. Therefore, 128 papers were related to this topic. The majority of them were developed in the United States, although Spain, China, the United Kingdom and Brazil are in the ranking of countries that published the most. According to this literature study, 72% of the found papers were published after 2015, showing that Visual Learning Analytics is an emerging field, also, there are several conferences that have included this topic in their publications. Additional analyses were made, focusing on the discovery of the most used algorithms, which include bar charts, line charts, pie charts, Heatmap, and others; whether data mining is commonly combined with Visual Learning Analytics; which educational level is most analyzed in the literature: high school, undergraduate or graduate; and whether any subject has more approaches as well as whether the initial computing classes have been analyzed.**

***Index Terms*—Visual analysis, education, literature**

I. Introduction

Nowadays a great deal of information is constantly being generated, making the acquisition of knowledge and the visualization of data a major challenge. In this way, it is important to understand the purposes of visualization techniques that seek to graphically represent datasets, so the visual representation generated may explore the capacity of human perception [1]. Hence, domain specialists and users of several knowledge domains can interpret and understand the spatial relations of the underlying data displayed in the visual outputs, leading to the obtaining of implicit and potentially useful knowledge.

According to the definition, visualization techniques provide several advantages to data analysis such as [2] [3] [4]: understanding information faster, due to the fact that sight is the human sense with the greatest capacity for capturing information in the shortest periods of time; viewing huge amounts of data in an intuitive and cohesive way; discovering atypical values and recognising patterns and relations; and involving the user by conveying a message, which might generate impact, work as an extension of human memory and also assist in the cognitive process.

Due to these several advantages, data visualization is already employed in various areas, such as stock market tracking [5], consulting movie databases [6] and in applications in educational areas. One of the main fields of visualization research in an educational context is the constant dropout of students in university majors, an issue that has several consequences, including social, economic and human.

The education field presents several challenges that concerns the research community and society. Culligan, et al. in [7] created a tool to keep track of students, allowing teachers to identify students at risk of failure or disengagement and enabling early interventions. Hegde in [8] used Principal Component Analysis (PCA), a technique of dimensionality reduction, in order to predict student dropout. Zhang et al. in [9] used Massive Open Online Course (MOOC) data to anticipate if a student would complete the course or not, also suggesting new courses according to each profile.

In this context, this work aims to implement a systematic literature study, in order to explore the impacts of visual analysis processes in Education as well as identify gaps in the literature.

The remainder of this paper is structured as follows: Section II shows the methodology steps followed in this work, presenting in the Sections III, IV, V more details of each stage. The analysis of the papers is in Section VI. Finally, Section VII concludes the article.

II. Research method: a process of systematic mapping

As the main purpose of this work concerns exploring research on visual analysis in an educational context, the methodology of this study was divided into three steps: Search for Papers, First Cycle and lastly the Second Cycle. These steps are shown in Figure 1. In each step, the papers were filtered aiming to select the most relevant ones, thus allowing to establish the basis of the study..

In order to achieve a literature study without returning biased papers or those that are not related to the theme [10], it is important to define the indexed academic databases where papers will be searched and to build a query string. To this end, during the phase Search for Papers, Scopus and Web of Science (WoS) were selected as databases, as well as

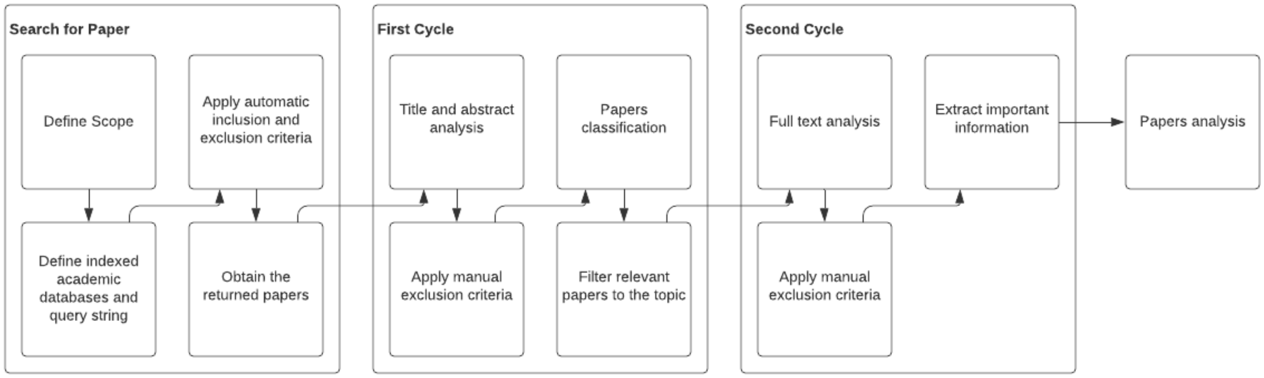

Fig. 1. Methodology defined for this research.

using VOSviewer [11] and the opinion of experts to build the query string. VOSviewer is a software tool to build a visual representation of bibliometric networks. VOSviewer has data mining functionality that can be used to build and visualize networks of co-occurrence of important terms extracted from a scientific literature body.

In order to get only relevant papers from the application of the query string in the databases, it was necessary to implement the exclusion criteria, which aim to remove the false-positive papers from the research. Hence, the automatic exclusion criteria to be applied in the databases is defined as follows: 1) analyse papers only from the last decade (2010 until 2019); 2) analyse papers only from computer science and social science areas; and 3) analyse papers from journals and conferences. The results of the Search for Papers are presented in Section III.

First Cycle is the next methodology step, which aims to detect and filter the papers related to visualization analysis of educational data. In this phase, the manual exclusion criteria were applied. Therefore, 1) all duplicated papers were removed; 2) documents classified as full books were discarded; 3) non-English-language papers were not analysed; 4) papers not available for download were removed; and 5) those that were not full papers were discarded.

After the manual exclusion criteria application, the papers' themes were analysed by reading the titles and abstracts of each one. It led to a pre-classification with the following categories, explained in Table I: 3D, Online, Games, Mobile, VR, Review, FE, Learning and VDA. The First Cycle results are presented on Section IV.

The next phase, Second Cycle, aims to analyse papers classified as VDA and extract relevant information in order to carry out the literature study. For that, it was necessary to read the entire paper. Also, the following information were extracted from the papers: Type - what is the purpose of the paper, LMS - if any LMS (*Learning Management System*) was analysed, API - if any API (*Application Programming Interface*) was used for data analysis, Open - if it has open algorithms, Algorithm - which visualization algorithms were used, Data Mining - if data mining algorithms were used, Source - which data sources are used, Major - if any courses were analyzed, Educational Level - which educational levels were analysed (K-12, undergraduate, graduate) and Class - if any classes were analysed.

TABLE I
CATEGORIES CLASSIFICATION FOR THE FIRST CYCLE.

3D	papers using 3D technology
Online	paper related to distance education
Games	papers related to games
Mobile	papers developed to cellphones
VR	papers using virtual reality technology
Review	papers of literature review
FE	papers out-of-scope
Learning	papers related with teaching of a tool/subject
VDA	papers using visual analysis

Reading the entire paper enables the identification of some details that are not traceable when only reading the abstract, for this reason, the information extraction was carried out at this stage. Similarly, the manual exclusion criteria were reapplied and some papers reclassified, since reading the entire paper revealed that, contrary to the impression gained from reading only the abstract, it did not belong to the VDA class. The Second Cycle results are presented in Section V.

At the end of the Second Cycle, only papers relevant to the area of visual analysis of educational data, with relevant information already extracted, were obtained. In the Analysis phase, a literature study was carried out, according to the following questions:

1) What is the number of documents by year?
2) Which sources publish the most on this research topic?
3) Which countries publish the most?
4) Which universities publish the most?
5) Are introductory programming courses analysed?
6) Which educational levels are analysed?

7) Which visualization algorithms are most common?
8) Is visualization used with Data Mining algorithms? Which algorithms are most common?

These questions were drafted in order to discover gaps in the research of the last decade in the education area. The analysis and the answers to the literature study questions are presented in Section VI.

III. Search for Papers

The aim of the first phase, Search for Papers, is to obtain papers related to the visual analysis of educational data from the databases previously set (Scopus and WoS). Thus, the query string had to be defined.

The process started by defining the key words to be used to retrieve papers in the considered databases, such as "visualisation" and "educational data". After that, the software VOSviewer was used in order to find new related words, increasing the query string. The opinion of experts was important for selecting relevant words and also adding any significant word not already identified. In the end, the query string was:

(("visualization techniques" OR "visualisation techniques" OR "visualization approach" OR "visualisation approach" OR "algorithm visualization" OR "algorithm visualisation" OR "data visualization" OR "data visualisation" OR "information visualization" OR "information visualisation" OR "visualization tool" OR "visualisation tool" OR "visual analytics")
AND
("educative" OR "educational dataset" OR "educational data" OR education)
AND
(student* OR undergraduate*))

After the implementation of the query string in the databases, a filter using the automatic exclusion criteria was applied, obtaining 823 papers from Scopus and 205 papers from WoS, totalling 1028 papers.

IV. First Cycle

In the First Cycle, the manual exclusion criteria were applied to the papers obtained from the Search for Papers. As a result 175 papers were selected as duplicated, 21 papers were classified as literature review and 17 documents were classified as full book. At the end of this process, 815 papers remained to be classified according to the aforementioned categories.

Although the defined criteria for manual exclusion states that it is necessary to analyse if the papers are classified as full-papers and are available for download, in the First Cycle only the abstract and title of the papers were analysed. Therefore, these exclusion criteria were not identified at this phase.

The 815 selected papers were classified as defined in Table I. The result is shown in Figure 2, where 248 papers were classified as FE, 247 as Learning, 22 as Games, 17 as 3D, 16 as VR, 12 as Online and 7 as mobile. However, only the 246 papers of VDA were analysed in the Second Cycle.

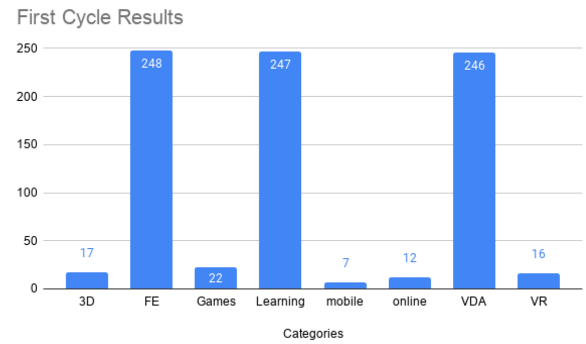

Fig. 2. First Cycle results.

V. Second Cycle

The main activity of the Second Cycle is the complete reading of the 246 papers obtained from the First Cycle and the extraction of relevant details from the text. The full reading highlighted some aspects that were not apparent during the abstract reading. In this sense, some papers classified as VDA had to be re-classified according to the other predefined categories.

Besides the re-classification, it was also necessary to apply the manual exclusion criteria one more time after the complete reading of the papers. During this process, 6 papers were found presenting titles and abstracts in English language and body texts written in Spanish language. Moreover, 16 papers were not classified as full papers and 44 were not available for download. These papers were excluded from the analysis.

At the end of the Second Cycle, 26 papers were reclassified as FE, 24 as Learning and 2 as Games, obtaining 128 papers on VDA. The final classification is illustrated in Figure 3. Furthermore, relevant information such as paper type, LMS, API, open source, visualization algorithms, data mining, data sources, majors, educational levels and classes were extracted from VDA papers to be analysed in the next phase.

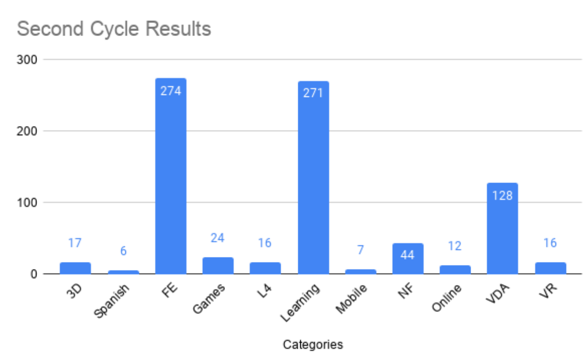

Fig. 3. Second Cycle results.

VI. Analyses

In this section, during the Analysis phase, the questions defined in Section II will be answered. This was possible due to an investigation of the 128 papers obtained after the Second Cycle.

A. *What is the number of documents by year?*

To analyse the interest in the research topic, it was necessary to note the number of publications by year. Figure 4 illustrates these results, which shows the increasing research on visual analysis of educational data. The largest number of publications was in 2017, though it decreased in 2018, which still presented a higher number when compared to previous years. Given that the research was carried out in November 2019, some publications from this year might not have been indexed when the research was conducted.

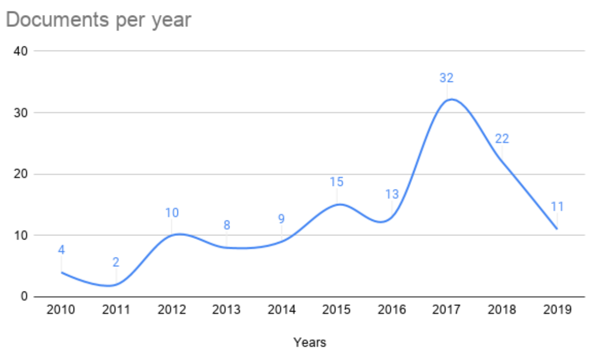

Fig. 4. Documents per year.

B. *Which sources publish the most on this research topic?*

In order to analyse the most relevant conferences and journals in this research topic, it is important to verify which sources publish the most. Two distinct sources were identified from the returned papers, journals and conferences. Figure 5 illustrates the number of papers published by source.

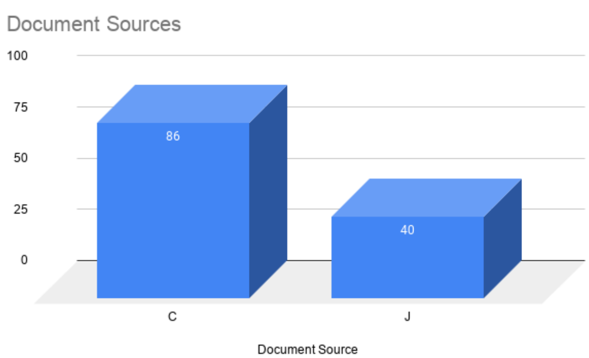

Fig. 5. Documents per source type.

Most of the papers, 86, were from conferences, a total of 67.5%. These papers have 60 distinct sources, and the "International Conference on Advanced Learning Technologies" from IEEE has the largest number of publications on the topic according to the result of the study. This conference has six papers published [12] [13] [14] [15] [16] [17].

Table II contains a list of the conferences with the largest number of papers published found in this study. The Frontiers in Education (FIE) from IEEE had just one paper [18] returned in this study.

TABLE II
CONFERENCES THAT PUBLISHED MOST ON THIS TOPIC.

Conferences	**Publisher**	**H5 index**	**Number of publications**
International Conference on Advanced Learning Technologies	IEEE	18	6
CEUR Workshop Proceedings		9	5
Annual Conference and Exposition	ASEE	9	4
International Conference on Computer Supported Education	Springer	14	3
IEEE Global Engineering Education Conference	IEEE	7	3
International Conference on Computers in Education	APSCE	11	3

Besides that, 40 papers from journals were found, a total of 31.7%. These papers have 29 different sources, where the journal "IEEE Transactions on Learning Technologies" from IEEE has the largest number of publications found on this topic. Five papers were published [19] [20] [21] [22] [23].

Table III shows the journals with the largest number of papers found in this study.

TABLE III
JOURNALS THAT PUBLISHED MOST ON THIS TOPIC.

Journals	**Publisher**	**H5 index**	**Number of publications**
IEEE Transactions on Learning Technologies	IEEE	25	5
Computers in Human Behavior	Elsevier	125	3
International Journal of Emerging Technologies in Learning	Kassel University Press GmbH	19	3
International Journal of Engineering Education	Dublin Institute of Technology Tempus Publications	23	2
Research in Learning Technology	Association for Learning Technology	23	2

C. *Which countries publish the most?*

The number of countries participating in the found publications were evaluated in order to find the countries that are concerned with visual analysis of educational data. Figure 6 illustrates the results.

The United States (USA) is the country with the highest participation in publications, 27 papers were found. Followed by Spain (15 papers), China (14 papers), United Kingdom

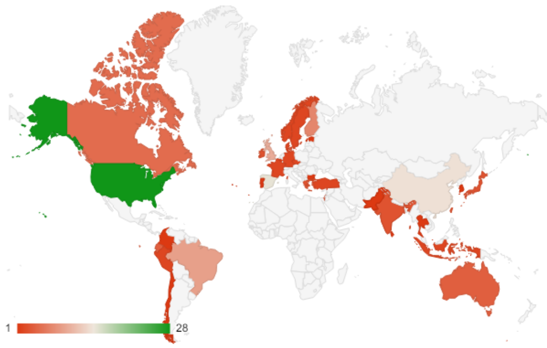

Fig. 6. Papers per country.

(10 papers) and Brazil in fifth place with 9 contributions in publications related to the topic. Table IV reports the countries that published the most on visual analysis of educational data.

TABLE IV
NUMBER OF PAPERS PER COUNTRY.

Countries	Number of publications
United States	27
Spain	15
China	14
United Kingdom	10
Brazil	9
Finland	7
Canada	5
Australia	4
Ecuador	4

D. Which universities publish the most?

As well as analysing the country, the universities with most participation in publications were also examined. One hundred and fifty different universities were found, where Tampere University of Technology in Finland had the largest number, a total of 4 participations [24] [25] [26] [27].

Table V shows the universities that had the largest number of participations in the topic and its respective number.

TABLE V
NUMBER OF PAPERS PER UNIVERSITY.

Universities	Country	Number of publications
Tampere University of Technology	Finland	4
Escuela Superior Politécnica del Litoral	Ecuador	3
University of Salamanca	Spain	3
Tsinghua University	China	3
Federal University of Uberlandia	Brazil	3
Universitat Oberta de Catalunya	Spain	3
Universidad Politécnica de Madrid	Spain	3

The United States was identified as the country that published the most concerning the research topic, though its publications are distributed among 32 universities. The American universities that published the most are: City University of New York [17] [28], Purdue University [29] [30] and University of California [30] [31]. Each university had published two papers according to the proposed methodology of this research.

E. Are introductory programming courses analysed?

After reading the papers, it was noticed that a small number of them analysed courses, only 30. Furthermore, 7 papers examine introductory courses, such as Basic Computer Programming [16], Basic Linux Commands [14], Introduction to Financial Information [32], Introduction to Programming Course [33], Introduction to Social Data Analysis [34], Introductory Computer Science Course [35] and Introductory Object-Oriented Programming Course [36].

In addition, classes from different majors were analysed in the papers, including math, languages, finance, and others. Those related to computer classes were the following: Algorithms and Programming [17], Artificial Intelligence [37], Basic Computer Programming [16], Basic Linux Commands [14], Data Structures and Algorithms [19], Database Systems [38], Design and Analysis of Algorithms [22], Functional Programming [39], Fundamentals of IT [40], Graph Theory [41], Introduction to Programming Course [33], Introduction to Social Data Analysis [34], Introductory Computer Science Course [35], Introductory Object Oriented Programming Course [36], Logic for Computer Science [17], Mathematics for Computing [42], Operating Systems [40], Programming Foundation (C language) [41], R Programming Design Course [43], Software Engineering [44] [18] [45] and Systems Architecture [46].

F. Which educational levels were analysed?

The majority of the papers that considered specific educational levels in their visual analysis processes employed data regarding undergraduate programs, 60 papers. Nevertheless, 8 papers took into account data related to graduate programs. Moreover, the analysis of data from K-12 school was found in 13 papers. These results are shown in Figure 7.

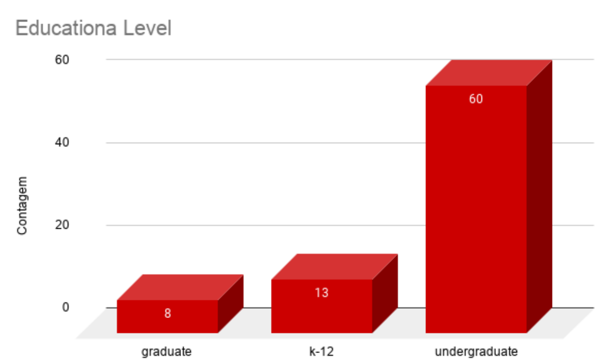

Fig. 7. Papers according to the educational level.

G. Which visualization algorithms are most common?

As all papers explored visual analysis in education, it is interesting to map the most employed visualization algorithms.

TABLE VI
NUMBER OF PAPERS PER VISUALIZATION ALGORITHMS.

Visualization algorithms	Number of publications
bar chart	46
line chart	24
heatmap	14
pie chart	14
network	11
scatter plot	11
timeline	10
box plot	7

Furthermore, those papers employed, at least, one visualization algorithm as described in Table VI.

The most of visualization algorithms are popular and well-known approaches, such as bar charts, line charts, pie charts, scatter plots. However, more complex algorithms were found, like t-SNE [31], Parallel Coordinates [18] [45] [47] [48], Ontologies [49], among others.

H. Is visualization used with Data Mining algorithms? Which algorithms are most common?

Considering the 128 papers related to visual analysis of educational data obtained in the Second Cycle, 42 papers proposed a data mining process with the support of visualization techniques. Clustering algorithms are commonly employed and were identified in 18 distinct papers. Tree-based algorithms were used in 11 papers and Latent Dirichlet Allocation (LDA) is described in 3 papers.

The clustering techniques covered some algorithms, like K-Means [48] [50] [28] [41] [51] [52], Multilevel Clustering [53], Hierarchical Clustering [50], Microsoft Sequence Clustering [54], Fuzzy C-Means [48], XMeans [54]; well as the tree algorithms that include Conditional Inference Tree, Decision Tree [40] [52] [55], J48 [52] [55], DecisiontreeC5.0 [56], Enhanced Random Tree [55], Process Trees [57] and Random Forest [37] [58].

VII. CONCLUSION

This paper proposed a literature study about visual analysis of educational data, accomplished in three steps, Search for Papers, First Cycle and Second Cycle. In the Search for Papers, a query string was built and was implemented in Scopus and WoS databases, where the automatic exclusion criteria were applied. In the First Cycle, a classification of the papers was made after manual exclusion criteria were employed. The Second Cycle extracted the relevant information from the papers classified as VDA. These steps enabled them to be analysed.

The analysis included verifying the number of papers per year, their sources, and also the leading countries that carry out research in the topic. Besides that, the analysis provided information on whether a study of introductory classes was made and which educational levels were analysed. Furthermore, it included the main visualization algorithms and whether data mining was used.

The findings of this work include that visualization in educational context is a topic in which academic interest is increasing. However, there are some gaps in the literature, for example, according to the analysis, just a few papers analyse introductory computer subjects, on the other hand, the data of undergraduate students are frequently examined in the research. The first contact with programming is during under-graduate study, where 30% of the students end up failing these introductory subjects [59] [60]. This study highlights that analysing introductory computer subjects is both relevant to educational improvement, and also fills a gap in the current literature.

Data mining and visualizations techniques are used together to improve results. Although, in the educational area, only 32% of the returned papers use both techniques. In conclusion, that there are still areas to be explored when using visualization with data mining in an educational context. It is expected that this study will provide useful guidance and also encourage new research developments for the emerging field of visual analysis in an educational context.

ACKNOWLEDGEMENTS

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) Finance Code 001 and Fundação de Apoio a Pesquisa do Distrito Federal (FAP-DF). Maristela Holanda thanks the FAP-DF Brazil, for the postdoctoral grant, Edital UnB-DPG 05- 2018 Bolsa de Pos-doutorado no Exterior FAP-DF.

REFERENCES

[1] M. Khan and S. S. Khan, "Data and information visualization methods, and interactive mechanisms: A survey," *International Journal of Computer Applications*, vol. 34, no. 1, pp. 1–14, 2011.

[2] H. A. Do Nascimento and C. B. Ferreira, "Uma introdução à visualização de informações," *Visualidades*, vol. 9, no. 2, 2011.

[3] L. F. Estivalet, "O uso de ícones na visualização de informações," Ph.D. dissertation, 2000.

[4] H. A. Do Nascimento and C. B. Ferreira, "Visualização de informações–uma abordagem prática," in *XXV Congresso da Sociedade Brasileira de Computação, XXIV JAI. UNISINOS, S. Leopoldo–RS*, 2005.

[5] T. Dwyer and P. Eades, "Visualising a fund manager flow graph with columns and worms," in *Proceedings Sixth International Conference on Information Visualisation*, 2002, pp. 147–152.

[6] C. Ahlberg and B. Shneiderman, "Visual information seeking using the filmfinder," in *Conference Companion on Human Factors in Computing Systems*, 1994, pp. 433–434.

[7] N. Culligan, K. Quille, and S. Bergin, "Veap: A visualisation engine and analyzer for press#," in *Proceedings of the 16th Koli Calling International Conference on Computing Education Research*, 2016, pp. 130–134.

[8] V. Hegde, "Dimensionality reduction technique for developing undergraduate student dropout model using principal component analysis through r package," in *2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, 2016, pp. 1–6.

[9] T. Zhang and B. Yuan, "Visualizing mooc user behaviors: A case study on xuetangx," in *Springer International Conference on Intelligent Data Engineering and Automated Learning*, 2016, pp. 89–98.

[10] J. Akoka, I. Comyn-Wattiau, and N. Laoufi, "Research on big data–a systematic mapping study," *Computer Standards & Interfaces*, vol. 54, pp. 105–115, 2017.

[11] VOSviewer, vOSviewer - visualizing scientifics landscape. Disponível em: https://www.vosviewer.com/.

[12] K. Mouri, F. Okubo, A. Shimada, and H. Ogata, "Bayesian network for predicting students' final grade using e-book logs in university education," in *2016 IEEE 16th International Conference on Advanced Learning Technologies (ICALT)*, 2016, pp. 85–89.

[13] I. G. Rojas and R. M. C. Garcia, "Towards efficient provision of feedback supported by learning analytics," in *2012 IEEE 12th International Conference on Advanced Learning Technologies*, 2012, pp. 599–603.

[14] R. Venant, P. Vidal, and J. Broisin, "Evaluation of learner performance during practical activities: an experimentation in computer education," in *2016 IEEE 16th International Conference on Advanced Learning Technologies (ICALT)*, 2016, pp. 237–241.

[15] A. Amigud, J. Arnedo-Moreno, T. Daradoumis, and A.-E. Guerrero-Roldan, "A method for thematic and structural visualization of academic content," in *2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT)*, 2017, pp. 230–234.

[16] D. A. T. de Menezes, D. L. Florêncio, R. E. D. Silva, I. D. Nunes, U. Schiel, and M. S. de Aquino, "David—a model of data visualization for the instructional design," in *2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT)*, 2017, pp. 281–285.

[17] R. Etemadpour, Y. Zhu, Q. Zhao, Y. Hu, B. Chen, M. A. Sharier, S. Zheng, and J. G. S. Paiva, "Academic performance analysis supported by a web-based visual analytics tool," in *2019 IEEE 19th International Conference on Advanced Learning Technologies (ICALT)*, vol. 2161, 2019, pp. 158–162.

[18] M. Á. Conde, F. J. García-Peñalvo, D. A. Gómez-Aguilar, and R. Theron, "Visual learning analytics techniques applied in software engineering subjects," in *2014 IEEE Frontiers in Education Conference (FIE) Proceedings*, 2014, pp. 1–9.

[19] T. Auvinen, L. Hakulinen, and L. Malmi, "Increasing students' awareness of their behavior in online learning environments with visualizations and achievement badges," *IEEE Transactions on Learning Technologies*, vol. 8, no. 3, pp. 261–273, 2015.

[20] C. Mejia, B. Florian, R. Vatrapu, S. Bull, S. Gomez, and R. Fabregat, "A novel web-based approach for visualization and inspection of reading difficulties on university students," *IEEE Transactions on Learning Technologies*, vol. 10, no. 1, pp. 53–67, 2016.

[21] S. Gutierrez-Santos, E. Geraniou, D. Pearce-Lazard, and A. Poulovassilis, "Design of teacher assistance tools in an exploratory learning environment for algebraic generalization," *IEEE Transactions on Learning Technologies*, vol. 5, no. 4, pp. 366–376, 2012.

[22] J. A. Velazquez-Iturbide, O. Debdi, N. Esteban-Sanchez, and C. Pizarro, "Greedex: A visualization tool for experimentation and discovery learning of greedy algorithms," *IEEE Transactions on Learning Technologies*, vol. 6, no. 2, pp. 130–143, 2013.

[23] B. Yousuf and O. Conlan, "Supporting student engagement through explorable visual narratives," *IEEE Transactions on Learning Technologies*, vol. 11, no. 3, pp. 307–320, 2017.

[24] K. Silius, A.-M. Tervakari, and M. Kailanto, "Visualizations of user data in a social media enhanced web-based environment in higher education," in *2013 IEEE Global Engineering Education Conference (EDUCON)*, 2013, pp. 893–899.

[25] A.-M. Tervakari, K. Silius, J. Koro, J. Paukkeri, and O. Pirttilä, "Usefulness of information visualizations based on educational data," in *2014 IEEE Global Engineering Education Conference (EDUCON)*, 2014, pp. 142–151.

[26] K. Silius, A.-M. Tervakari, and M. Kailanto, "Visualizations of user data in a social media enhanced web-based environment in higher education," in *2013 IEEE global engineering education conference (EDUCON)*, 2013, pp. 893–899.

[27] K. Kuosa, J. Koro, A. Tervakari, J. Paukkeri, and M. Kailanto, "Content analysis and visualizations: Tools for a social media-enhanced learning environment," in *2014 International Conference on Interactive Collaborative Learning (ICL)*, 2014, pp. 361–367.

[28] E. Sklar, C. Jansen, J. Chan, and M. Byrd, "Toward a methodology for agent-based data mining and visualization," in *Springer International Workshop on Agents and Data Mining Interaction*, 2011, pp. 4–15.

[29] M. Li, W. Wu, J. Zhao, K. Zhou, D. Perkis, T. N. Bond, K. Mumford, D. Hummels, Y. V. Chen, and M. Potel, "Careervis: hierarchical visualization of career pathway data," *IEEE Computer Graphics and Applications*, vol. 38, no. 6, pp. 96–105, 2018.

[30] C. G. P. Berdanier, N. T. Buswell, and Z. Zhao, "Data visualization for time-resolved real-time engineering writing processes," in *ASEE Annual Conference and Exposition, Conference Proceedings*, 2018.

[31] Z. A. Pardos and L. Horodyskyj, "Analysis of student behaviour in habitable worlds using continuous representation visualization," *arXiv preprint arXiv:1710.06654*, 2017.

[32] Á. Hernández-García, I. González-González, A. I. Jiménez-Zarco, and J. Chaparro-Peláez, "Visualizations of online course interactions for social network learning analytics," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 11, no. 07, pp. 6–15, 2016.

[33] K. Jankowski, A. Knutas, J. Ikonen, and J. Porras, "Automated social network analysis of online student collaboration activity," in *Proceedings of the 16th International Conference on Computer Systems and Technologies*, 2015, pp. 326–333.

[34] K. Dobashi, "Automatic data integration from moodle course logs to pivot tables for time series cross section analysis," *Procedia computer science*, vol. 112, pp. 1835–1844, 2017.

[35] T. Dragon and C. Lindeman, "Automated assessment of students' conceptual understanding: Supporting students and teachers using data from an interactive textbook," in *2017 IEEE International Symposium on Multimedia (ISM)*, 2017, pp. 567–572.

[36] J. Barria-Pineda, J. Guerra-Hollstein, and P. Brusilovsky, "A fine-grained open learner model for an introductory programming course," in *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization*, 2018, pp. 53–61.

[37] A. Segal, S. Hindi, N. Prusak, O. Swidan, A. Livni, A. Palatnic, B. Schwarz *et al.*, "Keeping the teacher in the loop: Technologies for monitoring group learning in real-time," in *Springer International Conference on Artificial Intelligence in Education*, 2017, pp. 64–76.

[38] V. Echeverria, R. Martinez-Maldonado, R. Granda, K. Chiluiza, C. Conati, and S. B. Shum, "Driving data storytelling from learning design," in *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*, 2018, pp. 131–140.

[39] M. Palomo-Duarte, J. M. Dodero, I. Medina-Bulo, E. J. Rodríguez-Posada, and I. Ruiz-Rube, "Assessment of collaborative learning experiences by graphical analysis of wiki contributions," *Interactive Learning Environments*, vol. 22, no. 4, pp. 444–466, 2014.

[40] P. Guleria and M. Sood, "Predictive data modeling: Educational data classification and comparative analysis of classifiers using python," in *2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC)*, 2018, pp. 740–746.

[41] T. Zhang, W. Dong, H. Shi, R. Liu, and H. Sun, "Design and implementation of teaching analysis system based on data mining," in *2019 Chinese Control And Decision Conference (CCDC)*, 2019, pp. 4400–4405.

[42] W. L. Cambruzzi, S. J. Rigo, and J. L. Barbosa, "Dropout prediction and reduction in distance education courses with the learning analytics multitrail approach." *Journal of Universal Computer Science*, vol. 21, no. 1, pp. 23–47, 2015.

[43] X. Li, X. Zhang, W. Fu, and X. Liu, "E-learning with visual analytics," in *2015 IEEE Conference on e-Learning, e-Management and e-Services (IC3e)*, 2015, pp. 125–130.

[44] T. T. Bilgin, "Parameter free visual exploration tool for mining school records," *The International journal of engineering education*, vol. 28, no. 3, pp. 744–754, 2012.

[45] M. Á. Conde, F. J. García-Peñalvo, D.-A. Gómez-Aguilar, and R. Therón, "Exploring software engineering subjects by using visual learning analytics techniques," *IEEE Revista Iberoamericana de Tecnologias del Aprendizaje*, vol. 10, no. 4, pp. 242–252, 2015.

[46] R. M. Crespo García, A. Pardo, C. Delgado Kloos, K. Niemann, M. Scheffel, and M. Wolpers, "Peeking into the black box: visualising learning activities," *International Journal of Technology Enhanced Learning*, vol. 4, no. 1-2, pp. 99–120, 2012.

[47] S. Nakamura, K. Kaneko, Y. Okada, C. Yin, and H. Ogata, "Cubic gantt chart as visualization tool for learning activity data," in *23rd International Conference on Computers in Education, ICCE 2015*. Asia-Pacific Society for Computers in Education, 2015, pp. 649–658.

[48] B. Dogan and A. Y. Camurcu, "Visual clustering of multidimensional educational data from an intelligent tutoring system," *Computer Applications in Engineering Education*, vol. 18, no. 2, pp. 375–382, 2010.

[49] M. R. Cantal and J. L. C. Peña, "Visualization models for virtual learning environments," in *2015 International Conference on Information Society (i-Society)*, 2015, pp. 16–21.

[50] T. Zhang and B. Yuan, "Visualizing mooc user behaviors: A case study on xuetangx," in *Springer International Conference on Intelligent Data Engineering and Automated Learning*, 2016, pp. 89–98.

[51] J. D. Garofalakis, A. Maritsas, and F. Oikonomou, “Assisting school units management with data mining techniques and gis visualization,” in *International Conference on Computer Supported Education*, 2017, pp. 331–338.

[52] I. M. Félix, A. P. Ambrósio, P. S. Neves, J. Siqueira, and J. D. Brancher, “Moodle predicta: A data mining tool for student follow up.” in *CSEDU (1)*, 2017, pp. 339–346.

[53] F. Gutiérrez, K. Seipp, X. Ochoa, K. Chiluiza, T. De Laet, and K. Verbert, “Lada: A learning analytics dashboard for academic advising,” *Computers in Human Behavior*, p. 105826, 2018.

[54] D. Riofrío-Luzcando, J. Ramírez, C. Moral, A. de Antonio, and M. Berrocal-Lobo, “Visualizing a collective student model for procedural training environments,” *Springer Multimedia Tools and Applications*, vol. 78, no. 8, pp. 10 983–11 010, 2019.

[55] A. Arunachalam, A. T. Raja, and S. Perumal, “Enhanced constructive decision tree classification model for engineering students data,” *International Journal of Recent Technology and Engineering*, 2019.

[56] V. Shankarataman, S. Gottipati, and J. R. Lin, “Extracting implicit suggestions from students’ comments–a text analytics approach,” 2017.

[57] N. Rossol, I. Cheng, and M. Mandal, “A workflow based process visual analyzer (proviszer) for teaching and learning,” in *Springer International Symposium on Visual Computing*, 2010, pp. 406–415.

[58] E. Gomede, F. H. Gaffo, G. U. Briganó, R. M. De Barros, and L. d. S. Mendes, “Application of computational intelligence to improve education in smart cities,” *Sensors*, vol. 18, no. 1, p. 267, 2018.

[59] M. Guzdial and E. Soloway, “Teaching the nintendo generation to program,” *Communications of the ACM*, vol. 45, no. 4, pp. 17–21, 2002.

[60] J. Bennedsen and M. E. Caspersen, “Failure rates in introductory programming,” *AcM SIGcSE Bulletin*, vol. 39, no. 2, pp. 32–36, 2007.