Article Title

Systematic Literature Review of Data Science, Data Analytics and Machine Learning Applied into Healthcare Engineering Systems

Authors details

- Name: Roberto J. Salazar Reyna
- Email address: roberto.salazar@udem.edu
- Affiliation: School of Engineering and Technology, Universidad de Monterrey, San Pedro Garza Garcia, Mexico

- Name: Fernando Gonzalez Aleu
- Email address: fernando.gonzalezaleu@udem.edu
- Affiliation: School of Engineering and Technology, Universidad de Monterrey, San Pedro Garza Garcia, Mexico

Biographies and acknowledgements

Roberto J. Salazar Reyna is an Industrial and Systems Engineering undergraduate student at Universidad de Monterrey (UDEM) in Mexico and an Industrial and Systems Engineering MS candidate at Binghamton University, State University of New York (SUNY) in the United States. His research focuses on healthcare engineering systems, including operations management, quality, safety and data analytics applied into healthcare engineering systems. Roberto Salazar is the corresponding author of this paper.

Fernando Gonzalez Aleu is an Associate Professor at Universidad de Monterrey (UDEM) in Mexico. He received his BS in Mechanical and Management Engineering at UDEM, a MSc with specialty in Manufacturing Systems at ITESM (Mexico) in 1999, a MSc in Industrial and Systems Engineering at Virginia Tech in 2015 and a PhD in Industrial and Systems Engineering at Virginia Tech in 2018. His research focuses on the application of continuous improvement programs and projects, including Kaizen Events and Lean Six Sigma projects. He is a member of the Institute of Industrial and Systems Engineers and the American Society for Engineering Management.

Structured Abstract

- **Purpose –** The objective of this papers is to assess and synthesize the published literature related with the application of data analytics, big data, data mining and machine learning into healthcare engineering systems.
- **Design/methodology/approach –** A systematic literature review (SLR) was conducted to obtain the most relevant papers related with the research study from three different platforms: EBSCOhost, ProQuest and Scopus. The literature was assessed and synthesized based on an analysis of characteristics related with the publications, their authors and their content.
- **Findings –** From the SLR, 576 publications were identified and analyzed. The research area seems to show characteristics of a growing field with new research areas evolving and applications being explored. In addition, the main authors and collaboration groups publishing in this research area were identified throughout a social network analysis; this could lead new and current authors identify researchers with common interests on the field.
- **Research limitations/implications –** The use of the SLR methodology does not guarantee that all relevant publications related with the research are will be covered and analyzed; however, the platforms were selected based on the authors' previous knowledge about the nature of the publications that could be found on each of them.
- **Originality/value –** To the best of the authors' knowledge, this paper represents the most comprehensive study regarding the authors who have publications of the field of data analytics, big data, data mining and machine learning applied into healthcare engineering systems.
- **Keywords:** data analytics, big data, machine learning, healthcare systems, systematic literature review
- **Article classification:** Literature review

Introduction

Data science is a "set of fundamental principles that support and guide the principled extraction of information and knowledge from data" (Foster and Tom, 2013). It involves the use and development of algorithms, processes, methodologies and techniques for understanding past, present and future phenomena through the analysis of data with the goal of improving decision making. Data scientists and data analytics must be able to view business problems from a data perspective to be able to leverage the benefits of its application on the organization. The healthcare industry is one of the world's largest, most critical and fastest-growing industries that is evolving through significant challenges in recent times (Raghunath *et al.*, 2013). It is considered as a data-driven industry and has historically generated large amount of data, driven by record keeping, compliance and regulatory requirements, and patient care (Wullianallur and Viju, 2014). However, the healthcare industry is also considered as a highly inefficient industry where one-third of its expenditures are wasted and do not contribute to better quality outcomes, according to a report from the Institute of Medicine. While the healthcare system continues to apply industrial and systems engineering tools with the aim of achieving an effective coordinated system, data analytics have the potential to improve care, save lives and lower costs by identifying associations and understanding trending and patterns within the data, which can lead to safer, faster, higher quality outcomes and more efficient medical practices.

Data science has several areas and disciplines within itself, thus there is no universal agreement in literature regarding its components and interactions. Dahl Winters (2015) developed a Venn diagram to visualize the three main fields of data science (i.e. data analytics, big data, methods and algorithms) and their intersections (i.e. data mining, machine learning, software tools, big data analytics) based on a two-axis diagram (i.e. on the *x* axis: experimental versus theoretical; on the *y* axis: descriptive versus prescriptive). On the other hand, Frank *et al.*, (2016), developed a schematic visualization (i.e. Efron-triangle) of the main fields constituting data science (i.e. domain knowledge, statistics/mathematics, computer science) and their intersections (i.e. machine learning, biostatistics, data engineering), based on the original data science Veen diagram created by Drew Conway (Conway, 2013). Taking into consideration the significant role data science can take to achieve better outcomes in healthcare systems, it would be relevant to understand to what extent each field/area has been applied, and its maturity state, in healthcare systems, along with the authors researching that field/area. Therefore, the aim of this paper is to assess and synthesize the published literature related with the impact, benefits, implications, challenges, opportunities or trends of data science exclusively in healthcare systems.

In order to achieve this aim, the authors used a SLR as the research methodology. SLRs aim to address the published literature of a specific research field by identifying, evaluating and integrating the findings of all relevant studies that address a set of research questions while being objective, systematic, transparent and replicable. However, in order for highly relevant publications to be identifiable, they must be indexed in targeted platforms/databases (Lefebvre *et al.*, 2011). To ensure this, the authors have strategically selected platforms that contained medical databases to ensure adequate coverage of the research area and designed a search strategy that allowed the capture of as many significant publications as possible.

After the final set of publication was obtained for this study, three different dimensions were assessed and evaluated to synthesize information: publication characteristics, authors characteristics and content characteristics. These were identified based on preliminary work defining relevant criteria to evaluate the

maturity of a research area (Heather *et al.*, 2013). The publication characteristics analyses included an examination on the publication trends over time as well as the characteristics of the publications' sources associated with the final paper set, which in this case were primarily academic journals, given the nature of the publication set. The authors characteristics examination included an investigation of past and new authors quantity and patterns of collaborations between them through social network analyses. Investigation of content characteristics, for this work's purpose, refers to analyze the scope in which the areas/fields within data science (e.g. data analytics, machine learning, data mining) have been addressed in healthcare systems, in which medical departments and to treat which diseases/disorders. Thus, the research questions addressed for this study are:

RQ1. Publications characteristics:

a. Which trend exist in publication pattern over time for this research area?
b. What type of sources are publishing the works?
c. Which are the sources with the highest frequency of published works on the field?
d. Which are the main study fields from the sources publishing the works?

RQ2. Authors characteristics:

a. How many authors are contributing to this area? To what extent are new authors contributing?
b. To what extent authors are collaborating between them in this research area?
c. How is the distribution of number of authors per publication?

RQ3. Content characteristics:

a. Which are the most frequently mentioned data science fields applied into healthcare systems?
b. Which are the top medical departments where data science is being studied and applied?
c. Which are the top diseases/disorders being addressed through data science approaches?
d. Which are the main study approaches on the theoretical publications set?
e. Which are the main application objectives on the case study publications set?
f. Which are the new emerging research lines related with this research area?

This paper is divided into three main section: research methodology (i.e. SLR conduction), results (publication characteristics, authors characteristics and content characteristics) and conclusions and future research.

Literature Review

In order to determine to what extent the literature related with data science applied into healthcare systems has been analyzed, a comparison study of previous literature systematic reviews was conducted. Table I shows the results of such study, sorted by publication year. The literature review conducted in 2015 aimed to discuss the perspectives of the evolving use of big data in science and healthcare, and to examine some of the opportunities and challenges. The literature review conducted in 2015 discussed big data applications in four majors biomedical subdisciplines: bioinformatics, clinical informatics, imaging informatics and public health informatics. The literature review conducted in 2017 reviewed big data sources and techniques in the health sector and identified which of these techniques are the most used in the prediction of chronic diseases. Once again, the first literature review conducted in 2018 reviewed big data analytics applications and challenges in its adoptions in healthcare, and identified strategies to overcome them. The second literature review conducted in 2018-and the most extensive one-provided a systematic review of the development of the fields of multiple healthcare sub-areas, data mining techniques, types of analytics, data and data sources, as well as possible directions. Finally, the last literature review conducted in 2018 assessed and synthesized how the big data phenomenon has contributed to better outcomes for the delivery of healthcare services.

One interesting finding from these systematic literature reviews is the fact that none of them conducted social network analyses related with authors publishing in this research field, which represented a gap within this field to be covered. This study, in addition of being the most updated one, analyzed a significantly higher number of publications in comparison with the others, included a theoretical approach study as well as a social network of the authors publishing in the research field aiming to help new and current researches identify researchers who have similar interests and research lines within this field and that are collaborating in study groups for the diffusion of knowledge.

Category	Big Data in Science and Healthcare: A Review of Recent Literature and Perspectives	Big Data Application in Biomedical Research and Health Care: A Literature Review	A Systematic Review of Techniques and Sources of Big Data in the Healthcare Sector	Concurrence of big data analytics and healthcare: A systematic review	A Systematic Review on Healthcare Analytics: Application and Theoretical Perspective of Data Mining	Data mining and predictive analytics applications for the delivery of healthcare services: a systematic literature review	This paper
Year	2014	2015	2017	2018	2018	2018	2019
Papers analyzed	0	68	32	58	117	22	576
Sources of big data	**No**	**No**	**Yes**	**Yes**	**No**	**No**	**No**
Sources of healthcare data	**No**	**No**	**No**	**Yes**	**No**	**No**	**No**
Big data analytical techniques	**No**	**No**	**Yes**	**Yes**	**Yes**	**No**	**Yes**
Application areas of big data/data mining	**Yes**	**Yes**	**No**	**Yes**	**Yes**	**Yes**	**Yes**
Platforms of big data	**No**	**No**	**Yes**	**No**	**No**	**No**	**No**
Big data definitions	**No**	**No**	**No**	**Yes**	**No**	**No**	**No**
Keywords network	**No**	**No**	**No**	**No**	**Yes**	**No**	**Yes**
Distribution of publications	**No**	**No**	**No**	**No**	**Yes**	**Yes**	**Yes**
Distribution of journals	**No**	**No**	**No**	**No**	**Yes**	**Yes**	**Yes**
Types of analytics	**No**	**No**	**No**	**No**	**Yes**	**No**	**Yes**
Classification by disease	**No**	**No**	**No**	**No**	**Yes**	**No**	**Yes**
Data mining algorithm tool/software	**No**	**No**	**No**	**No**	**No**	**Yes**	**No**
Authors' network	**No**	**No**	**No**	**No**	**No**	**No**	**Yes**
Theoretical approach	**No**	**No**	**No**	**No**	**No**	**No**	**Yes**

Table I. Systematic literature reviews (SLR) comparison table

Research methodology

The systematic literature review approach used throughout this bibliometric analysis was the one proposed by Heather *et al.*, (2016), adapted from David *et al.*, (2003) and the approach presented in the Cochrane Handbook (Julian and Sally 2011; David *et al.*, 2003), composed of seven steps:

1. *Problem definition*: the research team identifies the research area and defines the research objectives.
2. *Scoping study*: the research team identifies the desired study scope and conducts a 'traditional' literature review to find relevant publications related with the research area.
3. *Search strategy*: the research team evaluates the scoping set of papers by identifying potential search terms. Then they define the strategy by defining the databases/platforms to be searched, Boolean phrases, search tools, limiters and filters and exclusion criteria.
4. *Exclusion criteria*: the research team excludes those publications not directly related with analytics, data mining, big data and machine learning applied in healthcare engineering systems.
5. *Data collection*: the research team collects bibliometric data and criteria identified based on the aim of the research study.
6. *Data analysis*: the research team conducts the bibliometric analysis based on the aim of the research study.
7. *Reporting*: presentation of findings and results.

Problem definition

Throughout the literature, there are multiple publications regarding the use of data science, data analytics and machine learning algorithms applied into healthcare systems; however, it is not clear to what extent authors contributing to this research area are collaborating with each other to diffuse new knowledge and significant findings. For this reason, an SLR aiming to synthesize the current published literature can benefit to guide future development and evolution of this research area.

Scoping study

The scoping study was conducted through two activities. First, an initial list of search terms was set based on the authors' previous knowledge about the research area. Second, the research team identified six main publications related with the research area using three platforms (EBSCOhost, ProQuest and Scopus): Malik *et al*. (2018), Saiful *et al*. (2018), Hansen *et al*. (2014), Jake *et al*. (2016), Susel *et al*. (2017) and Nishita and Anil (2018), from which new search terms were added to the initial list. These six publications represented the scoping set of papers and were then used to create the search strategy.

Search strategy

The initial search strategy protocol consisted of 5 single search terms (data analytics, big data, data mining, machine learning and healthcare), three platforms (EBSCOhost, ProQuest and Scopus), utilization of Boolean operators (AND/OR), all fields search and two main exclusion criteria–published in academic journals and written in English language. This search strategy was tested and modified multiple times to

identify a final set of relevant publications for this research area. First, to increase the sensitivity of the search, synonyms (e.g. data analysis, analysis of data, mass data, massive data), techniques (e.g. data processing, text mining, deep learning), more specific concepts (e.g. artificial intelligence, business intelligence, internet of things) and the term "health care" (due to the lack of standardization between *healthcare* and *health care* in publications and academic texts) were added into the original search terms using the OR Boolean operator. Second, also to increase sensitivity, the Boolean phrase would be applied for abstracts instead of all field or all text, which helped control the scope. Lastly, conference materials were considered in the publications' search. Table II shows the final search strategy protocol used in this work. The search strategy was executed to identify all relevant papers up through July 2019.

Apply exclusion criteria

A total of 8,529 publications were identified and screened based on the exclusion criteria listed in Table I. First, duplicated publications (1,395 or 13.36% of the raw results) were removed. Second, publications that were not related with data science fields (2512 or 29.45%) were excluded. Third, publications that did not address data science fields applied exclusively to healthcare systems 4,006 or 46.96%) were excluded (e.g. paper that mentioned healthcare systems or used as an application example, but whose research was not focused on them, were excluded). It is relevant to mention that the search terms within the data science filed appeared on multiple papers related with "smart cities", another popular application field for data analytics. Finally, publications without an electronic file (38 or 0.44%) were excluded.

A total of 576 publications, representing 6.76 per cent of the raw results, was accepted as the final publication set for this research. For purposes of this research, these 576 publications were then classified into two separate sets based on their research approach: a theoretical publications set (consisting of 105 publications) and an applications publication set (consisting of 471 publications). The theoretical publications set included publications who mainly focused on studying and analyzing the strengths, weaknesses, opportunities, threats, challenges, capabilities, trends, benefits and promises of data science, data analytics and machine learning algorithms applied into healthcare systems as a whole. On the other hand, the applications publication set included publications related with case studies of data science, data analytics and machine learning algorithms applied into healthcare systems that addressed a specific problem, disease, medical condition or medical disorder.

To investigate the extent to which this research area is expanding, synthesizing and assessing the literature in the three dimensions outlined earlier (publications characteristics, authors characteristics and content characteristics) becomes a significant task. Each of these includes the analysis of one or more criteria, as reported in the following section.

Components of search	Explanation
Data science concept:	Search terms: Data analytics (8 search terms): analytics, data analytics, data analysis, analysis of data, informatic, informatics, health information technology, health information technologies Big data (6 search terms): big data, massive data, mass data, large data, macro data, metadata Data mining (3 search terms): data mining, data processing, text mining Machine learning (8 search terms): machine learning, artificial intelligence, robotics, deep learning, neural networks, internet of things, IoT, business intelligence
Healthcare concept:	Healthcare (2 search terms): healthcare, health care
Platforms:	EBSCOhost, ProQuest and Scopus
Search strategy:	Boolean operators OR within search terms for each concept (i.e. analytics OR analysis of data) AND across concepts (i.e. analytics AND healthcare) Search field: Abstract (EBSCOhost, ProQuest and Scopus) Publications present in academic journals or conference materials Publications written in a language other than English
Exclusion criteria:	Exclude: Duplicate publications Publications not related with the topic or that did not address data science fields exclusively on healthcare engineering systems Publications for which an electronic file is not available

Table II. Systematic literature review search protocol utilized

Results

To obtain a comprehensive perspective of the published literature of data science, data analytics and machine learning applied into healthcare engineering systems, this section presents results of analyses conducted to address the research questions posed earlier.

Publications characteristics

a. Publications trend

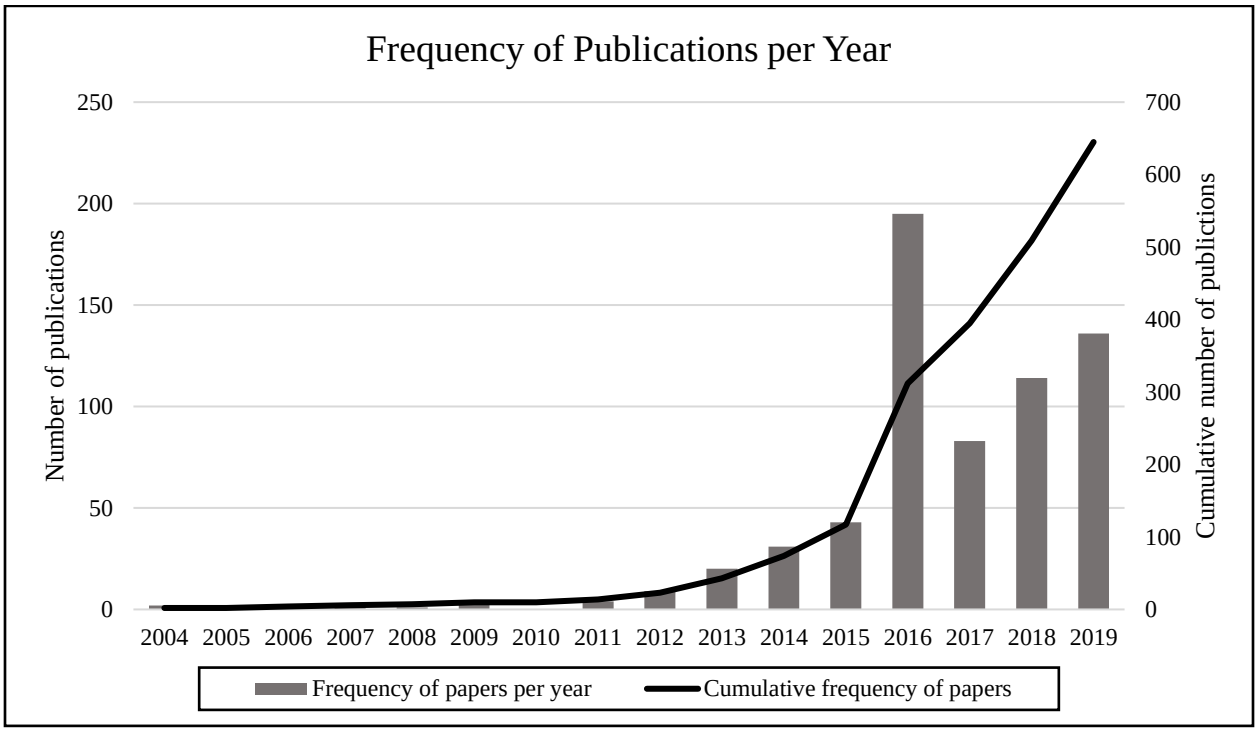

Figure 1. Frequency of publications per year

Trends analyses are useful for visualizing trends in the frequency of publications over time to determine the extent to which the frequency is changing. When conducting a SLR, publication rate is one of the multiple analyses often used to evaluate publication trends. Figure 1 consists in a chart corresponding to the frequency of publications per year; the following findings can be observed from it. The first paper focusing on data science, data analytics and machine learning applied into healthcare engineering systems was published in 2004 – thus, this particular research area spans only 15 years and appears to be relatively young. Second, from 2004 to 2010 the number of publications fluctuated between zero and three and does not seem to demonstrate an increasing trend. Third, as suggested by the cumulative frequency line, the publication trend started to increase after the year 2011, being 2016 the year with the highest number of publications (195 papers), up to date. For purposes of this analysis and considering that the publication set included papers published until end of June 2019, the last column corresponding to the frequency of published papers in 2019 was doubled to keep consistency within the data.

b. Publications' outlet type

Source type consists of analyzing the type of outlets that are publishing the work in this research area. From the publications0 outlet type analyses conducted, it is interesting to highlight the fact that 410 (or 71.18 per cent) of the papers were published in academic journals, 95 (or 16.49 per cent) in conference proceedings, 35 (or 6.07 percent) in reviews and 25 (or 4.34 per cent) in features. The rest were published in comparative studies, editorials, news or notes, respectively.

c. Top publications sources

A total of 346 publication outlets were identified from the set of 576 publications. From those outlets, the most frequently used were academic journals such as the *Journal of Medical Systems* (33), *PLoS One* (32), *BMC Medical Informatics and Decision Making* (13), *International Journal of Advanced Research in Computer Science* (12), *BMC Bioinformatics* (11), *Journal of Big Data* (11), *Computers in Biology* and *Medicine* (10) and *Journal of Medical Internet Research* (10). On the other hand, the conference proceedings authors most frequently published in were the *18th IEEE International Conference on e-Health Networking, Application and Services*, *IEEE 1st International Conference on Connected Health: Applications, Systems and Engineering Technologies*, *2016 IEEE International Conference on Healthcare Informatics*, *2016 IEEE International Conference on Mobile Services* and *2016 6th International Conference - Cloud System and Big Data Engineering*, all with two publications each, respectively.

d. Publications outlets' study fields

In addition, an analysis was conducted to identify the publications outlets' main study fields, according to SJR – Scimago Journal and Country Rank, in order to determine to which research field this topic would fit better. According to the results of the analysis, the publication outlets' main study fields were medicine (138), health informatics (101), information systems (82), computer science applications (67), computer networks and communications (61), biochemistry, genetics and molecular biology (55), health information management (48), electrical and electronic engineering (43), agricultural and biological sciences (37) and hardware and architecture (34). One interesting finding is the fact that most of the publication outlets' study fields could be associated in three main fields: health, computer science and information systems. Finally, an analysis of the journals' impact factor quartiles was conducted to identify their ranks on their respective categories. The analysis results showed that 42% where in Q1, 39% in Q2, 15% in Q3 and 4% in Q4, which suggests that most of the journals where the authors are publishing their works are highly ranked on their respective study field.

Authors characteristics

a. New authors' contribution

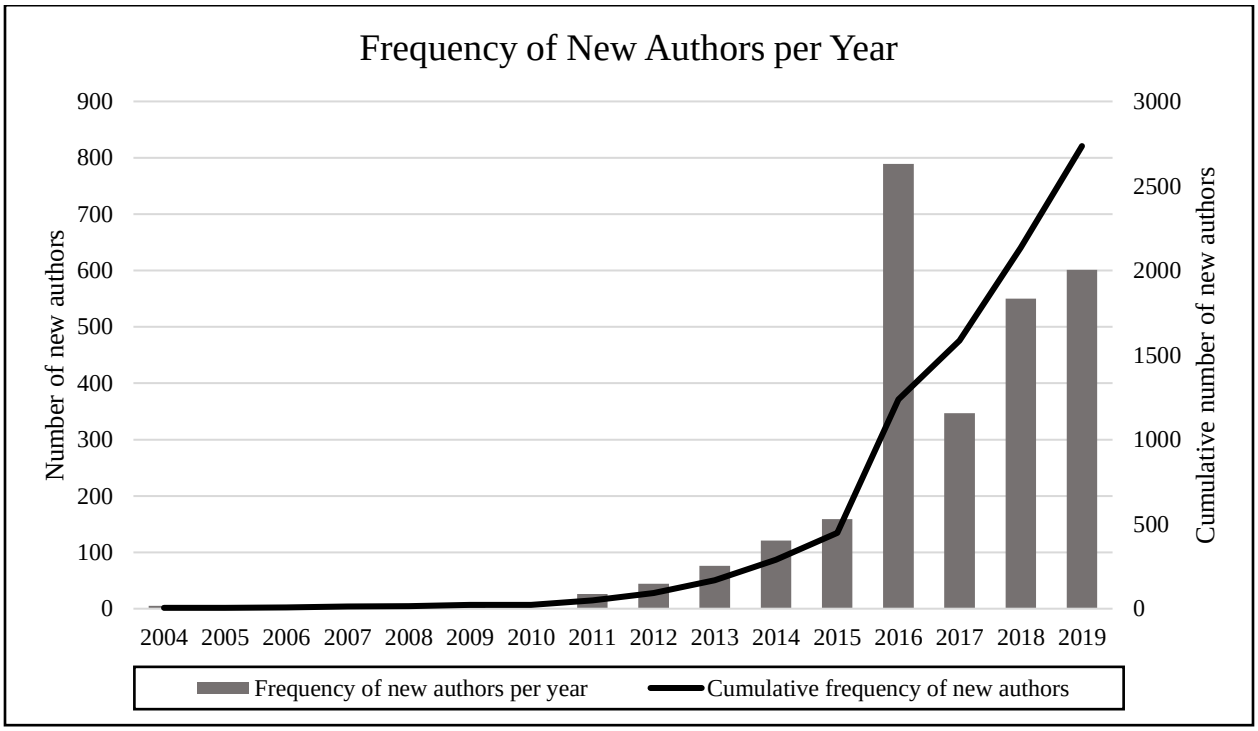

Figure 2. New authors per year

An analysis of the frequency of new authors publishing in this research area was conducted, as shown in Figure 2. The graph suggests an increasing trend on the number of new authors publishing in this research area, being the year 2016 the one with the highest introduction of new authors; further, the cumulative frequency seems to support the ability of this research area to attract new authors. For purposes of this analysis and considering that the publication set included papers published until end of June 2019, the last column corresponding to the frequency of new authors in 2019 was doubled to keep consistency within the data.

One of the characteristics of diversity of authors that is commonly investigated is the authors' affiliation country; this analysis allows determining to what extent author interest is concentrated primarily in a geographical region or dispersed around the world. The 2402 unique authors on both publication sets represent a total of 51 different countries. The countries with the highest number of authors were the USA (34.6 per cent), China (15.2 per cent), India (7.5 per cent), United Kingdom (6.3 per cent) and Australia (5.8 per cent). Other countries represented South Korea, Canada, Germany, Italy and Spain with less than 4 per cent each. Therefore, this research area, while attracting interest from authors around the world representing all continents, is concentrated primarily in five countries accounting for most of the authors (69.4 per cent).

b. Authors collaborations

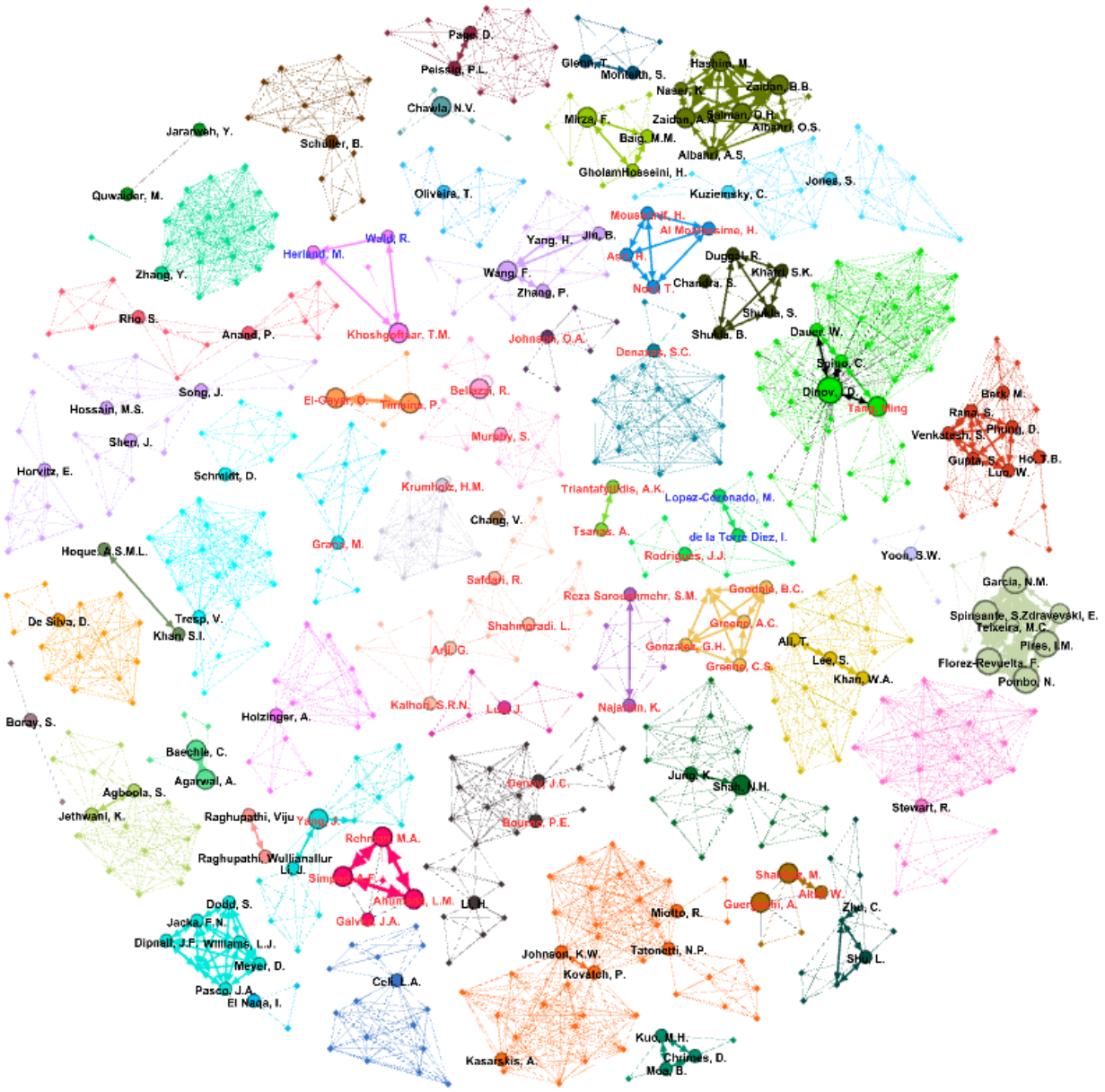

Figure 3. Co-author network for both publications sets

Collaboration between authors was analyzed using a social network created in Gephi to visualize direct and indirect interactions between authors and study groups; the Fruchterman Reingold clustering algorithm was applied since it was the one that allowed the best visualization. On Figure 3, the authors' names were color-coded: authors with a blue font appeared exclusively on the theoretical publications set, authors with a black font appeared exclusively on the case study publications set and authors with a red font appeared on both sets. For this figure, the size of the nodes represented the number of publications per author while the width

of the connecting line between nodes represented the total number of publications between two given authors. The authors with the highest number of publications were I. Dinov, Francisco Florez-Revuelta, Nuno Garcia, Ivan Pires, Nuno Pombo and S. Spisante, with four publications each, respectively. The large number of authors that have published more than a single paper suggests that this research area represents the main research focus for multiple authors. In the same way, Figure 3 illustrates the formation of multiple study groups, which confirms that diffusion of knowledge is occurring through collaboration.

c. Distribution of number of authors per publication

A study of the distribution of number of authors per publication was performed to get an insight about the way in which this research field is being studied (i.e. individually or in groups). Out of the 576 results, only 53 of them (or 9.20% of the analyzed publications) were written by a single author, while the other 523 publications were written in groups between 2 and 22 authors, being the group of 3 authors the one with the highest frequency with 117 publications (or 20.31% of the analyzed publications).With this analysis, it can be inferred that it is most likely for authors to study this research field in groups rather than individually, which strengthens the fact that diffusion of knowledge is occurring through collaboration.

Content characteristics

a. Data science fields applied into healthcare systems

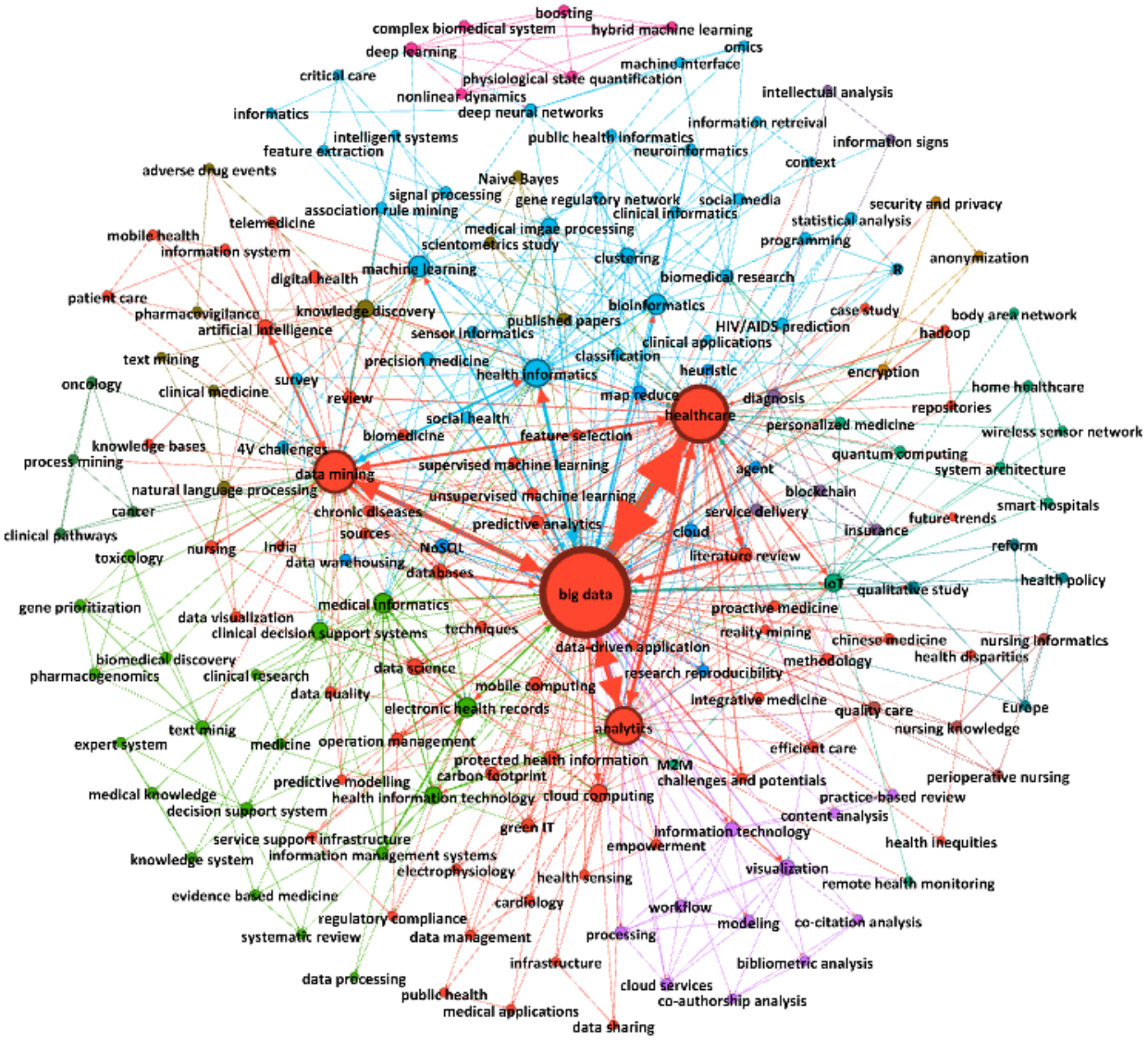

Figure 4. Keywords count network – application papers

In order to identify and analyze the top data science fields and machine learning algorithms applied into healthcare systems, as well as their concurrence relationship, a social network with the keywords from both publication sets was created using Gephi. The Fruchterman Reingold clustering algorithm was applied since it was the one that allowed the best visualization. Similarly, the size of the nodes represented the keyword's count frequency while the width of the connecting lines between nodes represented the total number of times their appeared together in a publication. The top five data science fields applied into healthcare systems were big data, machine learning, data mining, decision support systems and Internet of Things. On the other hand, the top machine learning and learning algorithms applied were cloud computing, decision

tree, neural networks, Naïve Bayes classifier, support vector machines and association rule. On interesting finding is the fact that the top machine learning algorithms applied into healthcare systems are classification and clustering algorithms, which suggests an idea of the purposes behind their applications.

b. Top medical departments where data science is being studied and applied

Identifying the top medical departments where data science and machine learning algorithms have been applied allows making inferences about the application fields' sizes, and thus, the degree to which they have been explored. According to an analysis conducted on this study, ontology, mental health, health services, elderly healthcare, epidemiology, omics, behavioral health, drug development, genomics and intensive care units turned to be the most popular medical departments where data science and machine learning algorithms have been applied (in descending order), opening road for further research in other medical departments.

c. Top diseases being addressed through data science approaches

Similarly, an analysis of the top diseases being addressed through data science and machine learning algorithms was conducted. One interesting finding is the fact that most of the diseases approached can be classified into three main groups: cancer (e.g. breast cancer), heart diseases (e.g. cardiovascular disease and strokes) and diabetes (e.g. diabetes type 2), which are all top leading causes of Americans' deaths and disabilities and leading drivers of the United States' $3.5 trillion in annual healthcare costs according to the National Center for Chronic Disease Prevention and Health Promotion (2019). Other diseases and medical disorders frequently studied and addressed through data science and machine learning algorithms were HIV, asthma and depression, respectively.

d. Study approach on theoretical publications set

Research area	Studies/analyses performed	Publication reference[1]
Big Data	Diverse uses and applications	2 3 5 7 9 18 22 25 26 28 31 32 33 35 36 40 41 42 43 44 45 46 47 48 49 54 55 56 61 62 72 85 86 95 107 108
	Implementation challenges/barriers/limitations	2 8 9 24 26 27 32 33 34 35 36 38 49 51 52 54 56 57 58 62 63 67 85 94 99
	Strengths, weaknesses, opportunities and/or threats	37 51 52 58 61 62 63 67 70 85 94 96 107
	Implementation advantages/benefits/promises	2 24 32 54 55 57 86 96 112
	Techniques	10 27 39 46 49 70 94
	Capabilities	13 27 31 45 70 78
	Systematic literature review	1 48 50 51 91
	Sources	10 33 70
	Characteristics	66 70
	Proposed model/framework	13
	Trends/future directions	108
	Others	79
Data Mining	Diverse uses and applications	11 15 17 19 75 76 77 89 101 103 109
	Techniques	76 77 89
	Strengths, weaknesses, opportunities and/or threats	102 111
	Systematic literature review	74 76 106
	Algorithms	75
	Characteristics	89
	Implementation advantages/benefits/promises	89
Healthcare Analytics	Diverse uses and applications	4 11 16 97
	Implementation challenges	4,97
	Perspectives	97
	Guidelines	97
Internet of Things	Diverse uses and applications	90
	Architecture, algorithms and applications	87
	Opportunities	90
	Systematic literature review	110
	Implementation challenges/barriers/limitations	90
Machine Learning	Diverse uses and applications	18 19 21 27 31 38 78 104
	Advantages and disadvantages	48
	Implementation challenges/barriers/limitations	93
Clinical Decision Support Systems	Architecture, algorithms and applications	69
	Implementation challenges/barriers/limitations	93
	Systematic/literature review	68
Medical Information Technologies	Architecture, algorithms and applications	83
	Trends/future directions	82
E-health	Diverse uses and applications	45 47 60 82
	Proposed model/framework	12 47 60 82
Others	Miscellaneous	6 23 71 80 81 88 92 98 100 115 116

Table III. Study approach of theoretical publications per paper

[1] Refer to Appendix 1 for references.

e. Application objectives on case study publications set

Application objective	Publication reference[2]
Prediction	1 2 6 7 9 12 14 25 27 35 46 51 52 78 81 92 105 111 112 120 121 126 138 144 145 160 121 174 178 189 191 194 201 205 251 262 265 266 276 282 288 291 304 306 328 330 331 332 336 341 353 356 359 375 384 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 407 408 409 410 411 412 413 414 415 416 417 418 427 429 436 438 439 444 449 474 480 481 485 489 490 496 497
Classification	21 22 25 36 39 40 80 93 117 184 245 250 267 307 325 326 342 354 358 362 367 371 373 380 382 451 452 478 495
Decision making	28 101 142 169 179 202 230 231 232 233 268 286 339 364 368 374 383 424 453 471 498
Data mining	15 17 103 216 218 219 220 345 346 347 350 351 370 386 422 423 487 494
Identification	16 90 129 182 214 252 274 298 299 300 301 302 310 333 463 502
Research	68 102 109 124 125 154 243 385 308 311 313 323 349 428 434 458
Diagnosis	10 53 72 88 91 95 115 130 188 261 295 296 297 317 318
Detection	19 29 31 41 55 140 143 211 226 239 275 466 479 483
Case study	77 152 165 209 235 263 271 337 343 430 435 468
Data analysis	106 247 281 289 419 454 456 464 469 500
Framework	42 49 175 176 180 208 237 287 293 340
Monitoring	26 44 56 79 97 99 100 278 294 431
Discovery	54 85 161 210 248 433 475 484
Data managing	18 57 64 213 221 240 437
Modeling	4 13 149 222 255 338
Pattern analysis	66 164 167 348 378 379
Association	118 139 196 322 387
Clustering	60 193 246 303 315
Data processing	3 69 98 357 440
Data visualization	91 224 229 361 388
Systematic review	73 74 75 123 447
Extraction	20 32 170 199
Forecasting	107 236 273 491
Comparative study	186 200 366
Data handling	116 283 284
Investigation	96 127 462
Optimization	225 292 372
Simulation	197 204 365
Assessment	137 181
Automation	83 141
Case management	134 467
Data integration	59 61
Exploration	37 499
Improvement	482 503
Prioritization	146 147
Screening	82 110
Stratification	8 376
Text mining	450 465
Translation	33 473
Miscellaneous	5 11 24 30 34 43 46 50 62 65 67 70 71 76 84 86 89 104 108 114 119 122 128 131 132 133 135 136 148 150 151 153 155 156 158 159 162 163 166 168 172 177 183 185 187 190 192 195 203 212 223 234 238 242 244 249 254 256 258 259 260 269 272 277 279 280 305 309 312 314 316 319 321 327 329 335 344 352 360 363 369 377 381 385 389 421 425 426 432 441 442 445 455 457 460 461 470 472 488 493 501

Table IV. Application objectives on the case study publications set per paper

[2] Refer to Appendix 2 for references.

Two different tables where constructed to classify the publications on both sets. While Table III classified the publications on the theoretical set based on their research area and study/analysis performed, Table IV classified the publications on the case study application set based on their application objective. As suggested previously on Figure 4 and displayed on Table III, most of the research of the publications on the theoretical set focused on big data, which is highly correlated with the amount of data generated daily by the healthcare industry. Similarly, as suggested on Figure 4 and displayed on Table IV, the application purposes of machine learning algorithms were mainly for prediction (e.g. readmissions prediction, disease prediction, fraud prediction, adverse event prediction, medical outcomes predictions) classification (i.e. based on the patients' treats and characteristics) and decision making (e.g. type of surgery, drugs and recovery process), outlining the significant role of predictive analytics in healthcare systems.

f. New emerging research lines

In order to understand the direction towards this research field is moving, a qualitative study was performed on the theoretical publications set to identify the new emerging research lines. These included: (1) the creation of algorithms and big data analytics technologies to address data privacy, data security and data traceability concerns; (2) improved understanding of the ethical, societal and economic implications of applying data analytics and machine learning algorithms in healthcare organizational decision-making; (3) big data and machine learning algorithms in conjunction with evidence-based medicine practices; (4) integration of multiple databases with different data structures; (5) big data applied into level molecular data (i.e. the atomic scale); (6) applications related with social media investigation; (7) addressing information loss in data preprocessing and cleaning steps; and (8) data analysis and automation for non-experts.

Conclusions and future research

This work contributes to the literature in three different ways: First, the authors with the highest frequencies of publications in this research area were identified. This information can aid other researches to connect with current researchers who have published works related with data science, data analytics and machine learning applied into healthcare engineering systems. Second, authors' social networks (i.e. study groups) were identified. Similarly, this information can aid other researches to identify study groups of authors who may have similar research lines within this research area; which becomes significantly relevant after taking into consideration its quick growing and expansion between academics and professionals. Third, the analysis of the publications using the three dimensions that were examined here offers the opportunity to obtain a more complete perspective from a publication set. From the perspective of publication characteristics, it can be strongly stated that the field of data science, data analytics and machine learning applied into healthcare engineering systems has passed the introduction phase and could be considered as a growing field: a growing trend in the number of publications, as well as the new research areas evolving and applications being explored, support this premise. In the same way, evidence from analyzing author characteristics suggest this research field as a growing field: a growing number of new authors contributing to this research area, the diversity of authors' nationalities and the formation of multiple study groups (i.e. research collaborations efforts between them). In the last dimension, content characteristics, the main medical departments and diseases/medical disorders that have been addressed through data science, data analytics and machine learning algorithms suggest the direction towards this research field is going and the new research lines that might evolve from it. Fourth, the papers on publication sets were classified based on their research objectives, approaches and purposes. This could aid new and current researches to identify publications related to a more specific field within the research area and with a particular objective. Finally, new research lines were identified. Future work is needed to determine to what extent these new research lines have been researched and why (i.e. limited technology, economic constraints, overall impact, lack of information). This could potentially inspire new and current researches to go more in depth into this fields to generate and share knowledge new knowledge to contribute to the study of this research field.

References

Alonso, S.G., de la Torre Diez, I., Rodrigues, J., Hamrioui, S. and Lopez-Coronado, M. (2017), "A systematic review of techniques and sources of big data in the healthcare sector", *Journal of Medical Systems*, Vol. 41 No. 11, pp. 1-9.

Conway, D. (2013), "The data science Venn diagram", available at: http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram (accessed 07 July 2019).

Emmert-Streib, F., Moutari, S. and Dehmer, M. (2016), "The process of analyzing data is the emergent feature of data science", *Front. Genet.*, Vol. 7 No. 12, pp. 1-4.

Hansen, M.M., Miron-Shatz, T., Lau, A.Y.S. and Paton, C. (2014), "Big data in science and healthcare: a review of recent literature and perspectives", *Yearb Med Inform*, Vol. 9 No. 1, pp. 21-26.

Higgins, J. and Green, S. (2011), "Chapter 4: Guide to the concepts of a Cochrane protocol and review", in Higgins, J. and Green, S., Cochrane Handbook for Systematic Reviews of Interventions, John Wiley & Sons Ltd., England, pp. 51-79.

Islam, S., Hasan, M., Wang, X., Germack, H.D. and E-Alam, N. (2018), "A systematic review on healthcare analytics: application and theoretical perspective of data mining", *Healthcare (Basel)*, Vol. 6 No. 2, pp .1-43.

Keathley, H., Gonzalez Aleu, F., Cardenas Orlandini, P., Van Aken, E., Deschamps, F. and Leite, L.R. (2013), "Maturity assessment of performance measurement implementation success factor failure", in *American Society for Engineering Management 2013 International Annual Conference, Minneapolis, MN, 2-5 October*.

Keathley, H., Van Aken, E.M., Gonzalez Aleu, F., Deschamps, F., Letens, G. and Cardenas Orlandini, P. (2016), "Assessing the maturity of a research area: bibliometric review and proposed framework", *Scientometrics*, Vol. 109 No. 2, pp. 927-951.

Lefebvre, C., Manheimer, E. and Glanville, J. (2011). "Chapter 6: searching for studies", available at: www.cochrane-handbook.org (accessed 22 July 2019).

Luo, J., Wu, M., Gopukumar, D. and Zhao, Y. (2016), "Big data application in biomedical research and health care: a literature review", *Biomedical Informatics Insights*, Vol. 8 No. 1, pp. 1-10.

Malik, M.M., Abdallah, S. and Ala'raj, M. (2018), "Data mining and predictive analytics applications for the delivery of healthcare services: a systematic literature review", Annals of Operations Research, Vol. 270 No. 1-2, pp. 287-312.

Mehta, N. and Pandit, A. (2018), "Concurrence of big data analytics and healthcare: a systematic review", International Journal of Medical Informatics, Vol. 114 No. 1, pp. 57-65.

Nambiar, R., Sethi, A., Bhardwaj, R. and Vargheese, R. (2013), "A look at challenges and opportunities of big data analytics in healthcare" in *Institute of Electrical and Electronics Engineers 2013 International Conference on big Data*.

National Center for Chronic Disease Prevention and Health Promotion. (2019), "Chronic Diseases in America", available at: https://www.cdc.gov/chronicdisease/resources/infographic/chronic-diseases.htm (accessed 25 October 2019).

Provost, F. and Fawcett, T. (2013). "Data science and its relationship to big data and data-driven decision making", Big Data, Vol. 1 No. 1, pp. 51-59.

Raghupathi, W. and Raghupathi, V. (2014), "Big data analytics in healthcare: promise and potential", *Health Information Science and Systems*, Vol. 2 No. 3, pp. 1-10.

Tranfield, D., Denyer, D. and Smart, P. (2003), “Towards a methodology for developing evidence-informed management knowledge by means of systematic review”, British Journal of Management, Vol. 14 No. 1, pp. 207-222.

Appendix 1.

[1] Liao, H., Tang, M., Luo, L., Li, C., Chiclana, F. and Zeng, X.J., 2018. A bibliometric analysis and visualization of medical big data research. *Sustainability*, *10*(1), p.166.

[2] Todor, R.D. and Anastasiu, C.V., 2018. A FUTURE TREND IN HEALTHCARE: THE USE OF BIG DATA. *Bulletin of the Transilvania University of Brasov. Economic Sciences. Series V*, *11*(1), pp.119-124.

[3] Luo, J., Wu, M., Gopukumar, D. and Zhao, Y., 2016. Big data application in biomedical research and health care: a literature review. *Biomedical informatics insights*, *8*, pp.BII-S31559.

[4] Simpao, A.F., Ahumada, L.M., Gálvez, J.A. and Rehman, M.A., 2014. A review of analytics and clinical informatics in health care. *Journal of medical systems*, *38*(4), p.45.

[5] Herland, M., Khoshgoftaar, T.M. and Wald, R., 2014. A review of data mining using big data in health informatics. *Journal of Big data*, *1*(1), pp.1-35.

[6] Wang, S., Ren, F. and Lu, H., 2018, May. A review of the application of natural language processing in clinical medicine. In *2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA)* (pp. 2725-2730). IEEE.

[7] Chauhan, R. and Jangade, R., 2016, January. A robust model for big healthcare data analytics. In *2016 6th International Conference-Cloud System and Big Data Engineering (Confluence)* (pp. 221-225). IEEE.

[8] Cyganek, B., Graña, M., Krawczyk, B., Kasprzak, A., Porwik, P., Walkowiak, K. and Woźniak, M., 2016. A survey of big data issues in electronic health record analysis. *Applied Artificial Intelligence*, *30*(6), pp.497-520.

[9] Kamble, S.S., Gunasekaran, A., Goswami, M. and Manda, J., 2019. A systematic perspective on the applications of big data analytics in healthcare management. *International Journal of Healthcare Management*, *12*(3), pp.226-240.

[10] Alonso, S.G., de la Torre Díez, I., Rodrigues, J.J., Hamrioui, S. and López-Coronado, M., 2017. A systematic review of techniques and sources of big data in the healthcare sector. *Journal of medical systems*, *41*(11), p.183.

[11] Islam, M.S., Hasan, M.M., Wang, X. and Germack, H.D., 2018, June. A systematic review on healthcare analytics: Application and theoretical perspective of data mining. In *Healthcare* (Vol. 6, No. 2, p. 54). Multidisciplinary Digital Publishing Institute.

[12] Amato, F., Cozzolino, G., Maisto, A., Mazzeo, A., Moscato, V., Pelosi, S., Picariello, A., Romano, S. and Sansone, C., 2015, September. ABC: A knowledge Based Collaborative framework for e-health. In *2015 IEEE 1st International Forum on Research and Technologies for Society and Industry Leveraging a better tomorrow (RTSI)* (pp. 258-263). IEEE.

[13] Wang, Y., Kung, L., Wang, W.Y.C. and Cegielski, C.G., 2018. An integrated big data analytics-enabled transformation model: Application to health care. *Information & Management*, *55*(1), pp.64-79.

[15] Zhang, Y., Guo, S.L., Han, L.N. and Li, T.L., 2016. Application and exploration of big data mining in clinical medicine. *Chinese medical journal*, *129*(6), p.731.

[16] Rueckel, D. and Koch, S., 2017. Application Areas of Predictive Analytics in Healthcare.

[17] Wilson, A.M., Thabane, L. and Holbrook, A., 2004. Application of data mining techniques in pharmacovigilance. *British journal of clinical pharmacology*, *57*(2), pp.127-134.

[18] Sughasiny, M. and Rajeshwari, J., 2018, August. Application Of Machine Learning Techniques, Big Data Analytics In Health Care Sector–A Literature Survey. In *2018 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC) I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC), 2018 2nd International Conference on* (pp. 741-749). IEEE.

[19] Shahid, N., Rappon, T. and Berta, W., 2019. Applications of artificial neural networks in health care organizational decision-making: A scoping review. *PloS one*, *14*(2), p.e0212356.

[20] Altaf, W., Shahbaz, M. and Guergachi, A., 2017. Applications of association rule mining in health informatics: a survey. *Artificial Intelligence Review*, *47*(3), pp.313-340.

[21] Triantafyllidis, A.K. and Tsanas, A., 2019. Applications of machine learning in real-life digital health interventions: Review of the literature. *Journal of medical Internet research*, *21*(4), p.e12286.

[22] Ahmad, F. and Tripathi, M.M., 2018. APPROACHES OF BIG DATA IN HEALTHCARE: A CRITICAL REVIEW. *International Journal of Advanced Research in Computer Science*, *9*(Special Issue 2), p.122.

[23] Radick, L., 2017. Artificial Intelligence in Healthcare: The Current, Compelling Wave of Interest. *Healthcare Executive*, *32*(5), pp.20-22.

[24] Kuiler, E.W., 2015. Big Data Adoption in the Health Care Domain: Challenges and Perspectives. *Washington Academy of Sciences. Journal of the Washington Academy of Sciences*, *101*(3), p.11.

[25] Raghupathi, W. and Raghupathi, V., 2014. Big data analytics in healthcare: promise and potential. *Health information science and systems*, *2*(1), p.3.

[26] Das, N., Das, L., Rautaray, S.S. and Pandey, M., 2018. Big data analytics for medical applications. *International Journal of Modern Education and Computer Science*, *11*(2), p.35.

[27] Kashyap, H., Ahmed, H.A., Hoque, N., Roy, S. and Bhattacharyya, D.K., 2016. Big data analytics in bioinformatics: architectures, techniques, tools and issues. *Network Modeling Analysis in Health Informatics and Bioinformatics*, *5*(1), p.28.

[28] Belle, A., Thiagarajan, R., Soroushmehr, S.M., Navidi, F., Beard, D.A. and Najarian, K., 2015. Big data analytics in healthcare. *BioMed research international*, *2015*.

[31] Ramesh, D., Suraj, P. and Saini, L., 2016, January. Big data analytics in healthcare: A survey approach. In *2016 International Conference on Microelectronics, Computing and Communications (MicroCom)* (pp. 1-6). IEEE.

[32] Milenkovic, M.J., Vukmirovic, A. and Milenkovic, D., 2019. Big data analytics in the health sector: challenges and potentials. *Management: Journal of Sustainable Business and Management Solutions in Emerging Economies*, *24*(1), pp.23-33.

[33] Rumsfeld, J.S., Joynt, K.E. and Maddox, T.M., 2016. Big data analytics to improve cardiovascular care: promise and challenges. *Nature Reviews Cardiology*, *13*(6), p.350.

[34] Bellazzi, R., 2014. Big data and biomedical informatics: a challenging opportunity. *Yearbook of medical informatics*, *23*(01), pp.08-13.

[35] Wang, W. and Krishnan, E., 2014. Big data and clinicians: a review on the state of the science. *JMIR medical informatics*, *2*(1), p.e1.

[36] Salas-Vega, S., Haimann, A. and Mossialos, E., 2015. Big data and health care: challenges and opportunities for coordinated policy development in the EU. *Health Systems & Reform*, *1*(4), pp.285-300.

[37] Collins, B., 2016. Big data and health economics: strengths, weaknesses, opportunities and threats. *Pharmacoeconomics*, *34*(2), pp.101-106.

[38] Krumholz, H.M., 2014. Big data and new knowledge in medicine: the thinking, training, and tools needed for a learning health system. *Health Affairs*, *33*(7), pp.1163-1170.

[39] Topaz, M. and Pruinelli, L., 2017. Big Data and nursing: implications for the future. *Stud Health Technol Inform*, *232*, pp.165-171.

[40] Westra, B.L. and Peterson, J.J., 2016. Big data and perioperative nursing. *AORN journal*, *104*(4), pp.286-292.

[41] Celi, L.A., Moseley, E., Moses, C., Ryan, P., Somai, M., Stone, D. and Tang, K.O., 2014. From pharmacovigilance to clinical care optimization. *Big data*, *2*(3), pp.134-141.

[42] Kansagra, A.P., John-Paul, J.Y., Chatterjee, A.R., Lenchik, L., Chow, D.S., Prater, A.B., Yeh, J., Doshi, A.M., Hawkins, C.M., Heilbrun, M.E. and Smith, S.E., 2016. Big data and the future of radiology informatics. *Academic radiology*, *23*(1), pp.30-42.

[43] Simpao, A.F., Ahumada, L.M. and Rehman, M.A., 2015. Big data and visual analytics in anaesthesia and health care. *British journal of anaesthesia*, *115*(3), pp.350-356.

[44] Luo, J., Wu, M., Gopukumar, D. and Zhao, Y., 2016. Big data application in biomedical research and health care: a literature review. *Biomedical informatics insights*, *8*, pp.BII-S31559.

[45] Liu, W. and Park, E.K., 2014, February. Big data as an e-health service. In *2014 International Conference on Computing, Networking and Communications (ICNC)* (pp. 982-988). IEEE.

[46] Patel, J.A. and Sharma, P., 2014, August. Big data for better health planning. In *2014 International Conference on Advances in Engineering & Technology Research (ICAETR-2014)* (pp. 1-5). IEEE.

[47] Andreu-Perez, J., Poon, C.C., Merrifield, R.D., Wong, S.T. and Yang, G.Z., 2015. Big data for health. *IEEE journal of biomedical and health informatics*, *19*(4), pp.1193-1208.

[48] Pashazadeh, A. and Navimipour, N.J., 2018. Big data handling mechanisms in the healthcare applications: A comprehensive and systematic literature review. *Journal of biomedical informatics*, *82*, pp.47-62.

[49] Hong, L., Luo, M., Wang, R., Lu, P., Lu, W. and Lu, L., 2018. Big data in health care: Applications and challenges. *Data and Information Management*, *2*(3), pp.175-197.

[50] de la Torre Díez, I., Cosgaya, H.M., Garcia-Zapirain, B. and López-Coronado, M., 2016. Big data in health: a literature review from the year 2005. *Journal of medical systems*, *40*(9), p.209.

[51] Alexandru, A.G., Radu, I.M. and Bizon, M.L., 2018. Big Data in Healthcare-Opportunities and Challenges. *Informatica Economica*, *22*(2).

[53] Asri, H., Mousannif, H., Al Moatassime, H. and Noel, T., 2015, June. Big data in healthcare: Challenges and opportunities. In *2015 International Conference on Cloud Technologies and Applications (CloudTech)* (pp. 1-7). IEEE.

[54] Dash, S., Shakyawar, S.K., Sharma, M. and Kaushik, S., 2019. Big data in healthcare: management, analysis and future prospects. *Journal of Big Data*, *6*(1), p.54.

[55] Li, G.Z. and Liu, B.Y., 2015. Big data is essential for further development of integrative medicine. *Chinese journal of integrative medicine*, *21*(5), pp.323-331.

[56] Benhlima, L., 2018. Big data management for healthcare systems: Architecture, requirements, and implementation. *Advances in bioinformatics*, *2018*.

[57] Schaeffer, C., Booton, L., Halleck, J., Studeny, J. and Coustasse, A., 2017. Big data management in US hospitals: benefits and barriers. *The health care manager*, *36*(1), pp.87-95.

[58] Zhang, X., Pérez-Stable, E.J., Bourne, P.E., Peprah, E., Duru, O.K., Breen, N., Berrigan, D., Wood, F., Jackson, J.S., Wong, D.W. and Denny, J., 2017. Big data science: opportunities and challenges

to address minority health and health disparities in the 21st century. *Ethnicity & disease*, *27*(2), p.95.

[60] Suciu, G., Suciu, V., Martian, A., Craciunescu, R., Vulpe, A., Marcu, I., Halunga, S. and Fratu, O., 2015. Big data, internet of things and cloud convergence–an architecture for secure e-health applications. *Journal of medical systems*, *39*(11), p.141.

[61] Makumbi, V., 2017. Big data's impact on the U.S. medical industry.

[62] Abouelmehdi, K., Beni-Hessane, A. and Khaloufi, H., 2018. Big healthcare data: preserving security and privacy. *Journal of Big Data*, *5*(1), p.1.

[65] Lopes, P., Silva, L.B. and Oliveira, J.L., 2015. Challenges and opportunities for exploring patient-level data. *BioMed research international*, *2015*.

[66] Wan, K. and Alagar, V., 2016, August. Characteristics and classification of big data in health care sector. In *2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)* (pp. 1439-1446). IEEE.

[67] Bellazzi, R., Masseroli, M., Murphy, S., Shabo, A. and Romano, P., 2012. Clinical Bioinformatics: challenges and opportunities.

[68] Yang, J., Kang, U. and Lee, Y., 2016. Clinical decision support system in medical knowledge literature review. *Information Technology and Management*, *17*(1), pp.5-14.

[69] Fan, A., Lin, D. and Tang, Y., 2017. Clinical Decision Support Systems for Comorbidity: Architecture, Algorithms, and Applications. *International journal of telemedicine and applications*, *2017*.

[70] Fang, R., Pouyanfar, S., Yang, Y., Chen, S.C. and Iyengar, S.S., 2016. Computational health informatics in the big data age: a survey. *ACM Computing Surveys (CSUR)*, *49*(1), p.12.

[71] Wong, Z.S.Y., Nøhr, C., Kuziemsky, C.E., Leung, E. and Youhua (Frank) Chen, 2017, July. Context Sensitive Health Informatics: Delivering 21st Century Healthcare-Building a Quality-and-Efficiency Driven System. In *CSHI* (pp. 1-5).

[72] Marino, S., Xu, J., Zhao, Y., Zhou, N., Zhou, Y. and Dinov, I.D., 2018. Controlled feature selection and compressive big data analytics: Applications to biomedical and health studies. *PloS one*, *13*(8), p.e0202674.

[74] Malik, M.M., Abdallah, S. and Ala'raj, M., 2018. Data mining and predictive analytics applications for the delivery of healthcare services: a systematic literature review. *Annals of Operations Research*, *270*(1-2), pp.287-312.

[75] Yoo, I., Alafaireet, P., Marinov, M., Pena-Hernandez, K., Gopidi, R., Chang, J.F. and Hua, L., 2012. Data mining in healthcare and biomedicine: a survey of the literature. *Journal of medical systems*, *36*(4), pp.2431-2448.

[76] Anand, T., Pal, R. and Dubey, S.K., 2016, March. Data mining in healthcare informatics: Techniques and applications. In *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)* (pp. 4023-4029). IEEE.

[77] Das, A.K., Kedia, A., Sinha, L., Goswami, S., Chakrabarti, T. and Chakrabarti, A., Data mining techniques in Indian Healthcare: A Short.

[78] Adam, N.R., Wieder, R. and Ghosh, D., 2017. Data science, learning, and applications to biomedical and health sciences. *Annals of the New York Academy of Sciences*, *1387*(1), pp.5-11.

[79] Hogle, L.F., 2016. Data-intensive resourcing in healthcare. *BioSocieties*, *11*(3), pp.372-393.

[80] Zemouri, R., Zerhouni, N. and Racoceanu, D., 2019. Deep Learning in the Biomedical Applications: Recent and Future Status. *Applied Sciences*, *9*(8), p.1526.

[81] Akay, A. and Hess, H., 2019. Deep Learning: Current and Emerging Applications in Medicine and Technology. *IEEE journal of biomedical and health informatics*, *23*(3), pp.906-920.

[82] Yang, J.J., Li, J., Mulder, J., Wang, Y., Chen, S., Wu, H., Wang, Q. and Pan, H., 2015. Emerging information technologies for enhanced healthcare. *Computers in industry*, *69*, pp.3-11.

[83] Lu, J. and Keech, M., 2015, September. Emerging technologies for health data analytics research: a conceptual architecture. In *2015 26th International Workshop on Database and Expert Systems Applications (DEXA)* (pp. 225-229). IEEE.

[85] Janke, A.T., Overbeek, D.L., Kocher, K.E. and Levy, P.D., 2016. Exploring the potential of predictive analytics and big data in emergency care. *Annals of emergency medicine*, *67*(2), pp.227-236.

[86] Aziz, H.A., 2016. Handling big data in modern healthcare. *Laboratory medicine*, *47*(4), pp.e38-e41.

[87] Hassanalieragh, M., Page, A., Soyata, T., Sharma, G., Aktas, M., Mateos, G., Kantarci, B. and Andreescu, S., 2015, June. Health monitoring and management using Internet-of-Things (IoT) sensing with cloud-based processing: Opportunities and challenges. In *2015 IEEE International Conference on Services Computing* (pp. 285-292). IEEE.

[88] Alexandre, B., 2018. How healthcare organizations can become information-driven.

[89] Tekieh, M.H. and Raahemi, B., 2015, August. Importance of data mining in healthcare: a survey. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015* (pp. 1057-1062). ACM.

[90] Singh, P., 2018. INTERNET OF THINGS BASED HEALTH MONITORING SYSTEM: OPPORTUNITIES AND CHALLENGES. *International Journal of Advanced Research in Computer Science*, *9*(1).

[91] Soleimani-Roozbahani, F., Ghatari, A.R. and Radfar, R., 2019. Knowledge discovery from a more than a decade studies on healthcare Big Data systems: a scientometrics study. *Journal of Big Data*, *6*(1), pp.1-15.

[92] El-Gayar, O. and Nawar, N., 2014. Leveraging Advanced Analytics to Generate Dynamic Medical Systematic Reviews.

[93] Johnson, A.E., Ghassemi, M.M., Nemati, S., Niehaus, K.E., Clifton, D.A. and Clifford, G.D., 2016. Machine learning and decision support in critical care. *Proceedings of the IEEE. Institute of Electrical and Electronics Engineers*, *104*(2), p.444.

[94] Dinov, I.D., 2016. Methodological challenges and analytic opportunities for modeling and interpreting Big Healthcare Data. *Gigascience*, *5*(1), pp.s13742-016.

[95] Swain, A.K., 2016. Mining big data to support decision making in healthcare. *Journal of Information Technology Case and Application Research*, *18*(3), pp.141-154.

[96] El-Gayar, O. and Timsina, P., 2014, January. Opportunities for business intelligence and big data analytics in evidence-based medicine. In *2014 47th Hawaii International Conference on System Sciences* (pp. 749-757). IEEE.

[97] Hahn, T., Nierenberg, A.A. and Whitfield-Gabrieli, S., 2017. Predictive analytics in mental health: applications, guidelines, challenges and perspectives. *Molecular psychiatry*, *22*(1), p.37.

[98] Kurniati, A.P., Johnson, O., Hogg, D. and Hall, G., 2016, October. Process mining in oncology: A literature review. In *2016 6th International Conference on Information Communication and Management (ICICM)* (pp. 291-297). IEEE.

[99] Huang, T., Lan, L., Fang, X., An, P., Min, J. and Wang, F., 2015. Promises and challenges of big data computing in health sciences. *Big Data Research*, *2*(1), pp.2-11.

[100] Thilakanathan, D., Zhao, Y., Chen, S., Nepal, S., Calvo, R.A. and Pardo, A., 2014, November. Protecting and Analysing Health Care Data on Cloud. In *2014 Second International Conference on Advanced Cloud and Big Data* (pp. 143-149). IEEE.

[101] Gonzalez, G.H., Tahsin, T., Goodale, B.C., Greene, A.C. and Greene, C.S., 2015. Recent advances and emerging applications in text and data mining for biomedical discovery. *Briefings in bioinformatics*, *17*(1), pp.33-42.

[102] Lesley, W.S., 2015. Risks and opportunities of data mining the electronic medical record. *Physician leadership journal*, *2*(4), p.40.

[103] Herland, M., Khoshgoftaar, T.M. and Wald, R., 2013, December. Survey of clinical data mining applications on big data in health informatics. In *2013 12th International Conference on Machine Learning and Applications* (Vol. 2, pp. 465-472). IEEE.

[105] Gavrishchaka, V., Senyukova, O. and Koepke, M., 2019. Synergy of physics-based reasoning and machine learning in biomedical applications: towards unlimited deep learning with limited data. *Advances in Physics: X*, *4*(1), p.1582361.

[106] Fallah, M. and Niakan Kalhori, S.R., 2017. Systematic review of data mining applications in patient-centered mobile-based information systems. *Healthcare informatics research*, *23*(4), pp.262-270.

[107] Zhu, R., Han, S., Su, Y., Zhang, C., Yu, Q. and Duan, Z., 2019. The application of big data and the development of nursing science: A discussion paper. *International Journal of Nursing Sciences*, *6*(2), pp.229-234.

[108] Austin, C. and Kusumoto, F., 2016. The application of Big Data in medicine: current implications and future directions. *Journal of Interventional Cardiac Electrophysiology*, *47*(1), pp.51-59.

[109] Belginova, S., Uvaliyeva, I. and Rustamov, S., 2019. The Application of Data Mining Methods for the Process of Diagnosing Diseases. *Journal of Theoretical and Applied Information Technology*, *97*(7).

[110] Ahmadi, H., Arji, G., Shahmoradi, L., Safdari, R., Nilashi, M. and Alizadeh, M., 2018. The application of internet of things in healthcare: a systematic literature review and classification. *Universal Access in the Information Society*, pp.1-33.

[111] Mantas, J., 2017. The hazards of data mining in healthcare. *Informatics Empowers Healthcare Transformation*, *238*, p.80.

[112] Rowh, M., 2019. The impact of big data on medical decisions. *Managed Healthcare Executive*.

[115] Denaxas, S.C., Asselbergs, F.W. and Moore, J.H., 2016. The tip of the iceberg: challenges of accessing hospital electronic health record data for biological data mining.

[116] Godbole, N.S. and Lamb, J., 2015, October. Using data science & big data analytics to make healthcare green. In *2015 12th International Conference & Expo on Emerging Technologies for a Smarter World (CEWIT)* (pp. 1-6). IEEE.

Appendix 2.

[1] Corani, G., Magli, C., Giusti, A., Gianaroli, L. and Gambardella, L.M., 2013. A Bayesian network model for predicting pregnancy after in vitro fertilization. *Computers in biology and medicine*, *43*(11), pp.1783-1792.

[2] Genc, O. and Dag, A., 2015. A Bayesian network-based data analytical approach to predict velocity distribution in small streams. *Journal of Hydroinformatics*, *18*(3), pp.466-480.

[3] Tan, J., Xiong, T., Miao, H., Sun, R. and Wu, M., 2018, April. A case study of medical big data processing: Data mining for the hyperuricemia. In *2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)* (pp. 196-201). IEEE.

[4] Quwaider, M. and Jararweh, Y., 2016. A cloud supported model for efficient community health awareness. *Pervasive and Mobile Computing*, *28*, pp.35-50.

[5] Ji, Z., Ganchev, I., O'Droma, M., Zhang, X. and Zhang, X., 2014. A cloud-based X73 ubiquitous mobile healthcare system: design and implementation. *The Scientific World Journal*, *2014*.

[6] Zaveri, S.H. and Joshi, N., 2017. A Comparative Study of Data Analysis Techniques in the domain of Medicative care for Disease Predication. *International Journal of Advanced Research in Computer Science*, *8*(3).

[7] Hao, F. and Blair, R.H., 2016. A comparative study: classification vs. user-based collaborative filtering for clinical prediction. *BMC medical research methodology*, *16*(1), p.172.

[8] De Meulder, B., Lefaudeux, D., Bansal, A.T., Mazein, A., Chaiboonchoe, A., Ahmed, H., Balaur, I., Saqi, M., Pellet, J., Ballereau, S. and Lemonnier, N., 2018. A computational framework for complex disease stratification from multiple large-scale datasets. *BMC systems biology*, *12*(1), p.60.

[9] Celi, L.A., Galvin, S., Davidzon, G., Lee, J., Scott, D. and Mark, R., 2012. A database-driven decision support system: customized mortality prediction. *Journal of personalized medicine*, *2*(4), pp.138-148.

[10] Vandewiele, G., De Backere, F., Lannoye, K., Berghe, M.V., Janssens, O., Van Hoecke, S., Keereman, V., Paemeleire, K., Ongenae, F. and De Turck, F., 2018. A decision support system to follow up and diagnose primary headache patients using semantically enriched data. *BMC medical informatics and decision making*, *18*(1), p.98.

[11] Cai, X., Dong, S. and Hu, J., 2019. A deep learning model incorporating part of speech and self-matching attention for named entity recognition of Chinese electronic medical records. *BMC medical informatics and decision making*, *19*(2), p.65.

[12] Tran, T., Luo, W., Phung, D., Gupta, S., Rana, S., Kennedy, R.L., Larkins, A. and Venkatesh, S., 2014. A framework for feature extraction from hospital medical data with applications in risk prediction. *BMC bioinformatics*, *15*(1), p.425.

[13] Bhattacharya, A., Tiwari, M.K. and Harding, J.A., 2012. A framework for ontology based decision support system for e-learning modules, business modeling and manufacturing systems. *Journal of Intelligent Manufacturing*, *23*(5), pp.1763-1781.

[14] Baechle, C. and Agarwal, A., 2017. A framework for the estimation and reduction of hospital readmission penalties using predictive analytics. *Journal of Big Data*, *4*(1), p.37.

[15] Rao, A.R. and Clarke, D., 2016, October. A fully integrated open-source toolkit for mining healthcare big-data: architecture and applications. In *2016 IEEE International Conference on Healthcare Informatics (ICHI)* (pp. 255-261). IEEE.

[16] Pooja, M.R. and Pushpalatha, M.P., 2015, December. A hybrid decision support system for the identification of asthmatic subjects in a cross-sectional study. In *2015 International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT)* (pp. 288-293). IEEE.

[17] Pasanisi, S. and Paiano, R., 2018. A hybrid information mining approach for knowledge discovery in cardiovascular disease (CVD). *Information*, *9*(4), p.90.

[18] Elhoseny, M., Abdelaziz, A., Salama, A.S., Riad, A.M., Muhammad, K. and Sangaiah, A.K., 2018. A hybrid model of internet of things and cloud computing to manage big data in health services applications. *Future generation computer systems*, *86*, pp.1383-1394.

[19] Yan, K., You, X., Ji, X., Yin, G. and Yang, F., 2016, October. A hybrid outlier detection method for health care big data. In *2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom)(BDCloud-SocialCom-SustainCom)* (pp. 157-162). IEEE.

[20] Luo, L., Li, L., Hu, J., Wang, X., Hou, B., Zhang, T. and Zhao, L.P., 2016. A hybrid solution for extracting structured medical information from unstructured data in medical records via a double-reading/entry system. *BMC medical informatics and decision making*, *16*(1), p.114.

[21] Patil, M.A., Patil, R.B., Krishnamoorthy, P. and John, J., 2016, August. A machine learning framework for auto classification of imaging system exams in hospital setting for utilization optimization. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 2423-2426). IEEE.

[22] Huo, D., Kou, B., Zhou, Z. and Lv, M., 2019. A machine learning model to classify aortic dissection patients in the early diagnosis phase. *Scientific reports*, *9*(1), p.2701.

[23] Golas, S.B., Shibahara, T., Agboola, S., Otaki, H., Sato, J., Nakae, T., Hisamitsu, T., Kojima, G., Felsted, J., Kakarmath, S. and Kvedar, J., 2018. A machine learning model to predict the risk of 30-day readmissions in patients with heart failure: a retrospective analysis of electronic medical records data. *BMC medical informatics and decision making*, *18*(1), p.44.

[24] Van Belle, V.M., Van Calster, B., Timmerman, D., Bourne, T., Bottomley, C., Valentin, L., Neven, P., Van Huffel, S., Suykens, J.A. and Boyd, S., 2012. A mathematical model for interpretable clinical decision support with applications in gynecology. *PloS one*, *7*(3), p.e34312.

[25] Jain, S. and Gangwar, M., 2017. A Mining Analysis over Psychiatric Database for Mental Health Classification. *i-Manager's Journal on Computer Science*, *5*(1), p.7.

[26] Li, S., Pham, H.T., Karunarathne, M.S., Lee, Y.S., Ekanayake, S.W. and Pathirana, P.N., 2015. A mobile cloud computing framework integrating multilevel encoding for performance monitoring in telerehabilitation. *Mathematical Problems in Engineering*, *2015*.

[27] Halawa, F., Al-Hihi, S., Shen, W. and Won, D., 2017. A Model-Based Approach of Data Analysis and Prediction in Chronic Kidney Diseases (CKD). In *IIE Annual Conference. Proceedings* (pp. 1018-1023). Institute of Industrial and Systems Engineers (IISE).

[28] Wimmer, H., Yoon, V.Y. and Sugumaran, V., 2016. A multi-agent system to support evidence based medicine and clinical decision making via data sharing and data privacy. *Decision Support Systems*, *88*, pp.51-66.

[29] Zhang, Y., Sun, Y., Phillips, P., Liu, G., Zhou, X. and Wang, S., 2016. A multilayer perceptron based smart pathological brain detection system by fractional Fourier entropy. *Journal of medical systems*, *40*(7), p.173.

[30] Chun, J.N., Lim, J.M., Kang, Y., Kim, E.H., Shin, Y.C., Kim, H.G., Jang, D., Kwon, D., Shin, S.Y., So, I. and Jeon, J.H., 2014. A network perspective on unraveling the role of TRP channels in biology and disease. *Pflügers Archiv-European Journal of Physiology*, *466*(2), pp.173-182.

[31] Subrahmanyam, M.S.L.B., Kumar, V.V. and Reddy, B.E., 2018. A New Algorithm for Skew Detection of Telugu Language Document based on Principle-axis Farthest Pairs Quadrilateral (PFPQ). *International Journal of Image, Graphics & Signal Processing*, *10*(3).

[32] Sadikin, M., Fanany, M.I. and Basaruddin, T., 2016. A new data representation based on training data characteristics to extract drug name entity in medical text. *Computational intelligence and neuroscience*, *2016*, p.6.

[33] Connolly, B., Cohen, K.B., Santel, D., Bayram, U. and Pestian, J., 2017. A nonparametric Bayesian method of translating machine learning scores to probabilities in clinical decision support. *BMC bioinformatics*, *18*(1), p.361.

[34] Ebrahimi, A.P., Ashlaghi, A.T. and Rad, M.M., 2013. A novel AIDS/HIV intelligent medical consulting system based on expert systems. *Journal of education and health promotion*, *2*.

[35] Gupta, P.K., Maharaj, B.T. and Malekian, R., 2017. A novel and secure IoT based cloud centric architecture to perform predictive analysis of users activities in sustainable health centres. *Multimedia Tools and Applications*, *76*(18), pp.18489-18512.

[36] Shobeetha, M. and Khadir, A.S.A., 2017. A Novel Framework for Classification of heart attack records from abc optimized Cloud Storage. *International Journal of Advanced Research in Computer Science*, *8*(9).

[37] Balakrishnan, M. and Radha, B., 2017. A Novel Healthcare Model to Explore Cloud Data Space Digitally for Health Improvement in Rural Areas. *International Journal of Advanced Research in Computer Science*, *8*(7).

[39] Kutlu, H. and Avcı, E., 2019. A Novel Method for Classifying Liver and Brain Tumors Using Convolutional Neural Networks, Discrete Wavelet Transform and Long Short-Term Memory Networks. *Sensors*, *19*(9), p.1992.

[40] Meenakshi, K., Safa, M., Karthick, T. and Sivaranjani, N., 2017. A Novel Study of Machine Learning Algorithms for Classifying Health Care Data. *Research Journal of Pharmacy and Technology*, *10*(5), p.1429.

[41] William, W., Ware, A., Basaza-Ejiri, A.H. and Obungoloch, J., 2019. A pap-smear analysis tool (PAT) for detection of cervical cancer from pap-smear images. *Biomedical engineering online*, *18*(1), p.16.

[42] Mishu, M.M., 2019, January. A Patient Oriented Framework using Big Data & C-means Clustering for Biomedical Engineering Applications. In *2019 International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)* (pp. 113-115). IEEE.

[43] Kim, M.K., Ter Jung, H., Kim, S.D. and La, H.J., 2016, June. A personal health index system with IoT devices. In *2016 IEEE International Conference on Mobile Services (MS)* (pp. 174-177). IEEE.

[44] Alfian, G., Syafrudin, M., Ijaz, M., Syaekhoni, M., Fitriyani, N. and Rhee, J., 2018. A personalized healthcare monitoring system for diabetic patients by utilizing BLE-based sensors and real-time data processing. *Sensors*, *18*(7), p.2183.

[46] Valenzuela, M.L., Rozenblit, J.W. and Hamilton, A.J., 2014, July. A predictive analytics toolbox for medical applications. In *Proceedings of the 2014 Summer Simulation Multiconference* (p. 25). Society for Computer Simulation International.

[48] Birnhack, M., 2019. A Process-based Approach to Informational Privacy and the Case of Big Medical Data. *Theoretical Inquiries in Law*, *20*(1), pp.257-290.

[49] Ahmed, S., El Seddawy, A.I. and Nasr, M., 2019. A Proposed Framework for Detecting and Predicting Diseases through Business Intelligence Applications. *International Journal of Advanced Networking and Applications*, *10*(4), pp.3951-3957.

[50] Xu, F., Cui, W., Kong, Q., Tang, Z. and Dong, J., 2018. A real-world evidence study for distribution of traditional chinese medicine syndrome and its elements on respiratory disease. *Evidence-Based Complementary and Alternative Medicine*, *2018*.

[51] Qian, B., Wang, X., Cao, N., Li, H. and Jiang, Y.G., 2015. A relative similarity based method for interactive patient risk prediction. *Data Mining and Knowledge Discovery*, *29*(4), pp.1070-1093.

[52] Canchi, T., Kumar, S.D., Ng, E.Y.K. and Narayanan, S., 2015. A review of computational methods to predict the risk of rupture of abdominal aortic aneurysms. *BioMed research international*, *2015*.

[53] Petrellis, N., 2018. A Review of Image Processing Techniques Common in Human and Plant Disease Diagnosis. *Symmetry*, *10*(7), p.270.

[54] Raja, K., Patrick, M., Gao, Y., Madu, D., Yang, Y. and Tsoi, L.C., 2017. A review of recent advancement in integrating omics data with literature mining towards biomedical discoveries. *International journal of genomics*, *2017*.

[55] Petit, C., Bezemer, R. and Atallah, L., 2018. A review of recent advances in data analytics for post-operative patient deterioration detection. *Journal of clinical monitoring and computing*, *32*(3), pp.391-402.

[56] Nguyen, H.H., Mirza, F., Naeem, M.A. and Nguyen, M., 2017, April. A review on IoT healthcare monitoring applications and a vision for transforming sensor data into real-time clinical feedback. In *2017 IEEE 21st International Conference on Computer Supported Cooperative Work in Design (CSCWD)* (pp. 257-262). IEEE.

[57] Delussu, G., Lianas, L., Frexia, F. and Zanetti, G., 2016. A Scalable Data Access Layer to Manage Structured Heterogeneous Biomedical Data. *PloS one*, *11*(12), p.e0168004.

[59] Mezghani, E., Exposito, E., Drira, K., Da Silveira, M. and Pruski, C., 2015. A semantic big data platform for integrating heterogeneous wearable data in healthcare. *Journal of medical systems*, *39*(12), p.185.

[60] Huang, M., Chen, Y., Chen, B.W., Liu, J., Rho, S. and Ji, W., 2016. A semi-supervised privacy-preserving clustering algorithm for healthcare. *Peer-to-Peer Networking and Applications*, *9*(5), pp.864-875.

[61] Massaro, A., Maritati, V., Savino, N., Galiano, A., Convertini, D., De Fonte, E. and Di Muro, M., 2018. A Study of a Health Resources Management Platform Integrating Neural Networks and DSS Telemedicine for Homecare Assistance. *Information*, *9*(7), p.176.

[62] Gim, J., Lee, S. and Joo, W., 2018. A Study of Prescriptive Analysis Framework for Human Care Services Based On CKAN Cloud. *Journal of Sensors*, *2018*.

[63] Farooqui, N.A., 2018. A STUDY ON EARLY PREVENTION AND DETECTION OF BREAST CANCER USING THREE-MACHINE LEARNING TECHNIQUES. *International Journal of Advanced Research in Computer Science*, *9*(Special Issue 2), p.37.

[64] Suganthalakshmi, T., Priya, M.S. and Thirunavukkarasu, S.K., 2016. A Study on Health Management Information System (HMIS) with reference to Periyanaickenpalayam Government Hospital, Coimbatore. *Journal of Contemporary Research in Management*, *11*(2).

[65] Fang, Z., Fan, X. and Chen, G., 2014. A study on specialist or special disease clinics based on big data. *Frontiers of medicine*, *8*(3), pp.376-381.

[66] Ku, J.H., 2017. A Study on the Real-time Analysis of Abnormal Pattern using the Sensor Data in IoT Environments. *International Information Institute (Tokyo). Information*, *20*(2B), p.1277.

[67] Richesson, R.L., Sun, J., Pathak, J., Kho, A.N. and Denny, J.C., 2016. Clinical phenotyping in selected national networks: demonstrating the need for high-throughput, portable, and computational methods. *Artificial intelligence in medicine*, *71*, pp.57-61.

[68] Bisaso, K.R., Anguzu, G.T., Karungi, S.A., Kiragga, A. and Castelnuovo, B., 2017. A survey of machine learning applications in HIV clinical research and care. *Computers in biology and medicine*, *91*, pp.366-371.

[69] Sundari, P.S., Subaji, M. and Karthikeyan, J., 2017. A Survey on effective similarity Search Models and Techniques for Big data Processing in Healthcare System. *Research Journal of Pharmacy and Technology*, *10*(8), pp.2677-2684.

[70] Zhang, L., Fabbri, D., Lasko, T.A., Ehrenfeld, J.M. and Wanderer, J.P., 2018. A System for Automated Determination of Perioperative Patient Acuity. *Journal of medical systems*, *42*(7), p.123.

[71] Pustišek, M., 2017. A System for Multi-Domain Contextualization of Personal Health Data. *Journal of medical systems*, *41*(1), p.16.

[72] Perveen, S., Shahbaz, M., Keshavjee, K. and Guergachi, A., 2018. A systematic machine learning based approach for the diagnosis of non-alcoholic fatty liver disease risk and progression. *Scientific reports*, *8*(1), p.2112.

[73] Schuler, J., Hudson, M.L., Schwartz, D. and Samudrala, R., 2017. A systematic review of computational drug discovery, development, and repurposing for Ebola virus disease treatment. *Molecules*, *22*(10), p.1777.

[74] Baig, M.M., GholamHosseini, H., Moqeem, A.A., Mirza, F. and Lindén, M., 2017. A systematic review of wearable patient monitoring systems–current challenges and opportunities for clinical adoption. *Journal of medical systems*, *41*(7), p.115.

[75] Baig, M.M., Afifi, S., GholamHosseini, H. and Mirza, F., 2019. A Systematic Review of Wearable Sensors and IoT-Based Monitoring Applications for Older Adults–a Focus on Ageing Population and Independent Living. *Journal of medical systems*, *43*(8), p.233.

[76] Gim, J., Lee, J., Jang, Y., Jeong, D.H. and Jung, H., 2016. A trend analysis method for iot technologies using patent dataset with goal and approach concepts. *Wireless Personal Communications*, *91*(4), pp.1749-1764.

[77] Guo, J., Qian, K., Zhang, G., Xu, H. and Schuller, B., 2017. Accelerating Biomedical Signal Processing Using GPU: A Case Study of Snore Sound Feature Extraction. *Interdisciplinary Sciences: Computational Life Sciences*, *9*(4), pp.550-555.

[78] Mandal, I. and Sairam, N., 2012. Accurate prediction of coronary artery disease using reliable diagnosis system. *Journal of medical systems*, *36*(5), pp.3353-3373.

[79] Wu, T., Gao, C.C., Lin, J.S. and Zha, J.L., 2017. Active Monitoring of Adverse Drug Reactions with Neural Network Technology. *Chinese medical journal*, *130*(12), p.1498.

[80] Lee, G., Bucheli, D.E.R. and Madabhushi, A., 2016. Adaptive Dimensionality Reduction with Semi-Supervision (AdDReSS): Classifying Multi-Attribute Biomedical Data. *PloS one*, *11*(7), p.e0159088.

[81] Li, J., Liu, L.S., Fong, S., Wong, R.K., Mohammed, S., Fiaidhi, J., Sung, Y. and Wong, K.K., 2017. Adaptive Swarm Balancing Algorithms for rare-event prediction in imbalanced healthcare data. *PloS one*, *12*(7), p.e0180830.

[82] De Silva, D., Burstein, F., Jelinek, H.F. and Stranieri, A., 2015. Addressing the Complexities of Big Data Analytics in Healthcare: The Diabetes Screening Case. *Australasian Journal of Information Systems*, *19*.

[83] Timsina, P., Liu, J. and El-Gayar, O., 2016. Advanced analytics for the automation of medical systematic reviews. *Information Systems Frontiers*, *18*(2), pp.237-252.

[84] Suwinski, P., Ong, C., Ling, M.H., Poh, Y.M., Khan, A.M. and Ong, H.S., 2019. Advancing personalized medicine through the application of whole exome sequencing and big data analytics. *Frontiers in genetics*, *10*, p.49.

[85] Tafti, A.P., Badger, J., LaRose, E., Shirzadi, E., Mahnke, A., Mayer, J., Ye, Z., Page, D. and Peissig, P., 2017. Adverse drug event discovery using biomedical literature: a big data neural network adventure. *JMIR medical informatics*, *5*(4), p.e51.

[86] Chen, D., Wang, H., Sheng, L., Hueman, M.T., Henson, D.E., Schwartz, A.M. and Patel, J.A., 2016. An algorithm for creating prognostic systems for cancer. *Journal of medical systems*, *40*(7), p.160.

[88] Gunčar, G., Kukar, M., Notar, M., Brvar, M., Černelč, P., Notar, M. and Notar, M., 2018. An application of machine learning to haematological diagnosis. *Scientific reports*, *8*(1), p.411.

[89] Cespedes, M.I., McGree, J., Drovandi, C.C., Mengersen, K., Doecke, J.D., Fripp, J. and Alzheimer's Disease Neuroimaging Initiative, 2018. An efficient algorithm for estimating brain covariance networks. *PloS one*, *13*(7), p.e0198583.

[90] Mencattini, A., Mosciano, F., Comes, M.C., Di Gregorio, T., Raguso, G., Daprati, E., Ringeval, F., Schuller, B., Di Natale, C. and Martinelli, E., 2018. An emotional modulation model as signature for the identification of children developmental disorders. *Scientific reports*, *8*(1), p.14487.

[91] Raghupathi, W. and Raghupathi, V., 2018. An empirical study of chronic diseases in the United States: a visual analytics approach to public health. *International journal of environmental research and public health*, *15*(3), p.431.

[92] Majumder, A.K.M., ElSaadany, Y.A., Young, R. and Ucci, D.R., 2019. An Energy Efficient Wearable Smart IoT System to Predict Cardiac Arrest. *Advances in Human-Computer Interaction*, *2019*.

[93] Yang, L., Wang, K., Xu, C., Zhu, C. and Sun, Y., 2016, May. An incremental learning classification algorithm based on forgetting factor for eHealth networks. In *2016 IEEE International Conference on Communications (ICC)* (pp. 1-6). IEEE.

[94] Razavi, F., Tarokh, M.J. and Alborzi, M., 2019. An intelligent Alzheimer's disease diagnosis method using unsupervised feature learning. *Journal of Big Data*, *6*(1), p.32.

[95] Djatna, T., Hardhienata, M.K.D. and Masruriyah, A.F.N., 2018. An intuitionistic fuzzy diagnosis analytics for stroke disease. *Journal of Big Data*, *5*(1), p.35.

[96] Boddy, A., Hurst, W., Mackay, M., El Rhalibi, A., Baker, T. and Montañez, C.A.C., 2019. An Investigation into Healthcare-Data Patterns. *Future Internet*, *11*(2), p.30.

[97] Mora, H., Gil, D., Terol, R.M., Azorín, J. and Szymanski, J., 2017. An IoT-based computational framework for healthcare monitoring in mobile environments. *Sensors*, *17*(10), p.2302.

[98] Taher, N.C., Mallat, I., Agoulmine, N. and El-Mawass, N., 2019, April. An IoT-Cloud Based Solution for Real-Time and Batch Processing of Big Data: Application in Healthcare. In *2019 3rd International Conference on Bio-engineering for Smart Technologies (BioSMART)* (pp. 1-8). IEEE.

[99] Yang, Z., Zhou, Q., Lei, L., Zheng, K. and Xiang, W., 2016. An IoT-cloud based wearable ECG monitoring system for smart healthcare. *Journal of medical systems*, *40*(12), p.286.

[100] García, L., Tomás, J., Parra, L. and Lloret, J., 2019. An m-health application for cerebral stroke detection and monitoring using cloud services. *International Journal of Information Management*, *45*, pp.319-327.

[101] Piovesan, L., Molino, G. and Terenziani, P., 2014. An ontological knowledge and multiple abstraction level decision support system in healthcare. *Decision Analytics*, *1*(1), p.8.

[102] Girardi, D., Dirnberger, J. and Giretzlehner, M., 2015. An ontology-based clinical data warehouse for scientific research. *Safety in Health*, *1*(1), p.6.

[103] Ding, D., Wu, D. and Yu, F., 2016, August. An overview on cloud computing platform spark for Human Genome mining. In *2016 IEEE International Conference on Mechatronics and Automation* (pp. 2605-2610). IEEE.

[104] Toti, G., Vilalta, R., Lindner, P., Lefer, B., Macias, C. and Price, D., 2016. Analysis of correlation between pediatric asthma exacerbation and exposure to pollutant mixtures with association rule mining. *Artificial intelligence in medicine*, *74*, pp.44-52.

[105] Sahoo, P.K., Mohapatra, S.K. and Wu, S.L., 2016. Analyzing healthcare big data with prediction for future health condition. *IEEE Access*, *4*, pp.9786-9799.

[106] Santos, G.L., Endo, P.T., da Silva Lisboa, M.F.F., da Silva, L.G.F., Sadok, D., Kelner, J. and Lynn, T., 2018. Analyzing the availability and performance of an e-health system integrated with edge, fog and cloud infrastructures. *Journal of Cloud Computing*, *7*(1), p.16.

[107] Luo, L., Liao, C., Zhang, F., Zhang, W., Li, C., Qiu, Z. and Huang, D., 2018. Applicability of internet search index for asthma admission forecast using machine learning. *The International journal of health planning and management*, *33*(3), pp.723-732.

[108] Li, D., Park, H.W., Batbaatar, E., Munkhdalai, L., Musa, I., Li, M. and Ryu, K.H., 2018. Application of a mobile chronic disease health-care system for hypertension based on big data platforms. *Journal of Sensors*, *2018*.

[109] Cheung, K.S., Leung, W.K. and Seto, W.K., 2019. Application of Big Data analysis in gastrointestinal research. *World journal of gastroenterology*, *25*(24), p.2990.

[110] Liu, M.M., Wen, L., Liu, Y.J., Cai, Q., Li, L.T. and Cai, Y.M., 2018. Application of data mining methods to improve screening for the risk of early gastric cancer. *BMC medical informatics and decision making*, *18*(5), p.121.

[111] Alahmar, A., Mohammed, E. and Benlamri, R., 2018, August. Application of Data Mining Techniques to Predict the Length of Stay of Hospitalized Patients with Diabetes. In *2018 4th International Conference on Big Data Innovations and Applications (Innovate-Data)* (pp. 38-43). IEEE.

[112] Qi, W., Ratanatharathorn, A., Gevonden, M., Bryant, R., Delahanty, D., Matsuoka, Y., Olff, M., deRoon-Cassini, T., Schnyder, U., Seedat, S. and Laska, E., 2018. Application of data pooling to longitudinal studies of early post-traumatic stress disorder (PTSD): the International Consortium to Predict PTSD (ICPP) project. *European journal of psychotraumatology*, *9*(1), p.1476442.

[114] Woo, H., Kim, K., Cha, K., Lee, J.Y., Mun, H., Cho, S.J., Chung, J.I., Pyo, J.H., Lee, K.C. and Kang, M., 2019. Application of Efficient Data Cleaning Using Text Clustering for Semistructured Medical Reports to Large-Scale Stool Examination Reports: Methodology Study. *Journal of medical Internet research*, *21*(1), p.e10013.

[115] Murphy, D.R., Meyer, A.N., Sittig, D.F., Meeks, D.W., Thomas, E.J. and Singh, H., 2019. Application of electronic trigger tools to identify targets for improving diagnostic safety. *BMJ Qual Saf*, *28*(2), pp.151-159.

[116] Chen, J., Zhang, Y., Zhang, X., Cao, R., Chen, S., Huang, Q., Lu, X., Wan, X., Wu, X., Xu, C. and Xu, G., 2011. Application of L-EDA in metabonomics data handling: global metabolite profiling and potential biomarker discovery of epithelial ovarian cancer prognosis. *Metabolomics*, *7*(4), pp.614-622.

[117] Naraei, P., Abhari, A. and Sadeghian, A., 2016, December. Application of multilayer perceptron neural networks and support vector machines in classification of healthcare data. In *2016 Future Technologies Conference (FTC)* (pp. 848-852). IEEE.

[118] Altaf, W., Shahbaz, M. and Guergachi, A., 2017. Applications of association rule mining in health informatics: a survey. *Artificial Intelligence Review*, *47*(3), pp.313-340.

[119] Sajedi, H., 2018. Applications of data hiding techniques in medical and healthcare systems: a survey. *Network Modeling Analysis in Health Informatics and Bioinformatics*, *7*(1), p.6.

[120] Cruz, J.A. and Wishart, D.S., 2006. Applications of machine learning in cancer prediction and prognosis. *Cancer informatics*, *2*, p.117693510600200030.

[121] Islam, M.M., Wu, C.C., Poly, T.N., Yang, H.C. and Yu-Chuan (Jack) Li, 2018, January. Applications of Machine Learning in Fatty Live Disease Prediction. In *MIE* (pp. 166-170).

[122] Triantafyllidis, A.K. and Tsanas, A., 2019. Applications of machine learning in real-life digital health interventions: Review of the literature. *Journal of medical Internet research*, *21*(4), p.e12286.

[123] Brunson, J.C. and Laubenbacher, R.C., 2017. Applications of network analysis to routinely collected health care data: a systematic review. *Journal of the American Medical Informatics Association*, *25*(2), pp.210-221.

[124] Hyde, K.K., Novack, M.N., LaHaye, N., Parlett-Pelleriti, C., Anden, R., Dixon, D.R. and Linstead, E., 2019. Applications of supervised machine learning in autism spectrum disorder research: a review. *Review Journal of Autism and Developmental Disorders*, *6*(2), pp.128-146.

[125] Wu, C., Weng, Y., Jiang, Q., Guo, W. and Wang, C., 2016, August. Applied research on visual mining technology in medical data. In *2016 4th International Conference on Cloud Computing and Intelligence Systems (CCIS)* (pp. 229-233). IEEE.

[126] Sangi, M., Win, K.T., Shirvani, F., Namazi-Rad, M.R. and Shukla, N., 2015. Applying a novel combination of techniques to develop a predictive model for diabetes complications. *PloS one*, *10*(4), p.e0121569.

[127] Ou-Yang, C., Wulandari, C.P., Hariadi, R.A.R., Wang, H.C. and Chen, C., 2018. Applying sequential pattern mining to investigate cerebrovascular health outpatients' re-visit patterns. *PeerJ*, *6*, p.e5183.

[128] Rahman, M. and Mondal, S.K., 2013. Approach for Labeling the Class of Credit card Customers via Clustering Method in Data Mining. *International Journal of Advanced Research in Computer Science*, *4*(9).

[129] Pires, I., Garcia, N., Pombo, N., Flórez-Revuelta, F. and Spinsante, S., 2018. Approach for the development of a framework for the identification of activities of daily living using sensors in mobile devices. *Sensors*, *18*(2), p.640.

[130] Dharmasiri, N.D.K.G. and Vasanthapriyan, S., 2018, September. Approach to Heart Diseases Diagnosis and Monitoring through Machine Learning and iOS Mobile Application. In *2018 18th International Conference on Advances in ICT for Emerging Regions (ICTer)* (pp. 407-412). IEEE.

[131] Catania, L.J. and Nicolitz, E., 2018. Artificial intelligence and its applications in vision and eye care. *Advances in Ophthalmology and Optometry*, *3*(1), pp.21-38.

[132] Tack, C., 2019. Artificial intelligence and machine learning| applications in musculoskeletal physiotherapy. *Musculoskeletal Science and Practice*, *39*, pp.164-169.

[133] Muhsen, I.N., Elhassan, T. and Hashmi, S.K., 2018. Artificial Intelligence Approaches in Hematopoietic Cell Transplantation: A Review of the Current Status and Future Directions. *Turkish Journal of Hematology*, *35*(3), p.152.

[134] Bennett, C.C., 2019. Artificial intelligence for diabetes case management: The intersection of physical and mental health. *Informatics in Medicine Unlocked*, *16*, p.100191.

[135] Chang, H.Y., Jung, C.K., Woo, J.I., Lee, S., Cho, J., Kim, S.W. and Kwak, T.Y., 2019. Artificial intelligence in pathology. *Journal of pathology and translational medicine*, *53*(1), p.1.

[136] Krittanawong, C., Zhang, H., Wang, Z., Aydar, M. and Kitai, T., 2017. Artificial intelligence in precision cardiovascular medicine. *Journal of the American College of Cardiology*, *69*(21), pp.2657-2664.

[137] Spertus, J.V., T. Normand, S.L., Wolf, R., Cioffi, M., Lovett, A. and Rose, S., 2016. Assessing hospital performance after percutaneous coronary intervention using big data. *Circulation: Cardiovascular Quality and Outcomes*, *9*(6), pp.659-669.

[138] Moustakas, A., 2018. Assessing the predictive causality of individual based models using Bayesian inference intervention analysis: an application in epidemiology. *Stochastic environmental research and risk assessment*, *32*(10), pp.2861-2869.

[139] Marschollek, M., 2016. Associations between sensor-based physical activity behaviour features and health-related parameters. *Human movement science*, *45*, pp.1-6.

[140] Yuan, J., Holtz, C., Smith, T. and Luo, J., 2016. Autism spectrum disorder detection from semi-structured and unstructured medical data. *EURASIP Journal on Bioinformatics and Systems Biology*, *2017*(1), p.3.

[141] Schneeweiss, S., 2018. Automated data-adaptive analytics for electronic healthcare data to study causal treatment effects. *Clinical epidemiology*, *10*, p.771.

[142] Monteith, S. and Glenn, T., 2016. Automated decision-making and big data: concerns for people with mental illness. *Current psychiatry reports*, *18*(12), p.112.

[143] Leclere, B., Buckeridge, D.L., Boelle, P.Y., Astagneau, P. and Lepelletier, D., 2017. Automated detection of hospital outbreaks: A systematic review of methods. *PloS one*, *12*(4), p.e0176438.

[144] Vukicevic, A.M., Stojadinovic, M., Radovic, M., Djordjevic, M., Cirkovic, B.A., Pejovic, T., Jovicic, G. and Filipovic, N., 2016. Automated development of artificial neural networks for clinical purposes: application for predicting the outcome of choledocholithiasis surgery. *Computers in biology and medicine*, *75*, pp.80-89.

[145] Lin, C., Karlson, E.W., Canhao, H., Miller, T.A., Dligach, D., Chen, P.J., Perez, R.N.G., Shen, Y., Weinblatt, M.E., Shadick, N.A. and Plenge, R.M., 2013. Automatic prediction of rheumatoid arthritis disease activity from the electronic medical records. *PloS one*, *8*(8), p.e69932.

[146] Kalid, N., Zaidan, A.A., Zaidan, B.B., Salman, O.H., Hashim, M., Albahri, O.S. and Albahri, A.S., 2018. Based on real time remote health monitoring systems: A new approach for prioritization “large scales data” patients with chronic heart diseases using body sensors and communication technology. *Journal of medical systems*, *42*(4), p.69.

[147] Kalid, N., Zaidan, A.A., Zaidan, B.B., Salman, O.H., Hashim, M. and Muzammil, H., 2018. Based real time remote health monitoring systems: A review on patients prioritization and related" big

data" using body sensors information and communication technology. *Journal of medical systems*, *42*(2), p.30.

[148] Khennak, I. and Drias, H., 2017. Bat-inspired algorithm based query expansion for medical web information retrieval. *Journal of medical systems*, *41*(2), p.34.

[149] Srividya, M., Mohanavalli, S. and Bhalaji, N., 2018. Behavioral modeling for mental health using machine learning algorithms. *Journal of medical systems*, *42*(5), p.88.

[150] He, K.Y., Ge, D. and He, M.M., 2017. Big data analytics for genomic medicine. *International journal of molecular sciences*, *18*(2), p.412.

[151] Coates, J., Souhami, L. and El Naqa, I., 2016. Big data analytics for prostate radiotherapy. *Frontiers in oncology*, *6*, p.149.

[152] Christensen, J.H., Petersen, M.K., Pontoppidan, N.H. and Cremonini, M., 2018, November. Big Data Analytics in Healthcare: Design and Implementation for a Hearing Aid Case Study. In *2018 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)* (pp. 296-303). IEEE.

[153] Weider, D.Y., Gottumukkala, C., Senthailselvi, D.A., Maniraj, P. and Khonde, T., 2016, June. Big Data Analytics in Service Computing: HealthCare-Software As A Service. In *2016 IEEE International Conference on Mobile Services (MS)* (pp. 170-173). IEEE.

[154] Doubal, F.N., Ali, M., Batty, G.D., Charidimou, A., Eriksdotter, M., Hofmann-Apitius, M., Kim, Y.H., Levine, D.A., Mead, G., Mucke, H.A. and Ritchie, C.W., 2017. Big data and data repurposing-using existing data to answer new questions in vascular dementia research. *BMC neurology*, *17*(1), p.72.

[155] Ngiam, K.Y. and Khor, W., 2019. Big data and machine learning algorithms for health-care delivery. *The Lancet Oncology*, *20*(5), pp.e262-e273.

[156] Bibault, J.E., Giraud, P. and Burgun, A., 2016. Big data and machine learning in radiation oncology: state of the art and future prospects. *Cancer letters*, *382*(1), pp.110-117.

[158] Cremona, C., 2016. Big data and structural health monitoring. *IABSE Congr. Stock., no*.

[159] Dion, M., AbdelMalik, P. and Mawudeku, A., 2015. Big Data: Big Data and the Global Public Health Intelligence Network (GPHIN). *Canada Communicable Disease Report*, *41*(9), p.209.

[160] Yang, S., Njoku, M. and Mackenzie, C.F., 2014. 'Big data'approaches to trauma outcome prediction and autonomous resuscitation. *British Journal of Hospital Medicine*, *75*(11), pp.637-641.

[161] Baechle, C., Agarwal, A. and Zhu, X., 2017. Big data driven co-occurring evidence discovery in chronic obstructive pulmonary disease patients. *Journal of Big Data*, *4*(1), p.9.

[162] Monteith, S., Glenn, T., Geddes, J., Whybrow, P.C. and Bauer, M., 2016. Big data for bipolar disorder. *International journal of bipolar disorders*, *4*(1), p.10.

[163] Agoston, D.V. and Langford, D., 2017. Big Data in traumatic brain injury; promise and challenges. *Concussion*, *2*(4), p.CNC44.

[164] Federer, C., Yoo, M. and Tan, A.C., 2016. Big data mining and adverse event pattern analysis in clinical drug trials. *Assay and drug development technologies*, *14*(10), pp.557-566.

[165] Katsis, Y., Balac, N., Chapman, D., Kapoor, M., Block, J., Griswold, W.G., Huang, J., Koulouris, N., Menarini, M., Nandigam, V. and Ngo, M., 2017, July. Big data techniques for public health: a case study. In *Proceedings of the Second IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies* (pp. 222-231). IEEE Press.

[166] Waldman, S.A. and Terzic, A., 2016. Big data transforms discovery–utilization therapeutics continuum. *Clinical Pharmacology & Therapeutics*, *99*(3), pp.250-254.

[167] Liyanage, H., De Lusignan, S., Liaw, S.T., Kuziemsky, C., Mold, F., Krause, P., Fleming, D. and Jones, S., 2014. Big data usage patterns in the health care domain: A use case driven approach applied to the assessment of vaccination benefits and risks. *Yearbook of medical informatics*, *23*(01), pp.27-35.

[168] Rapp, R., 2016. Big data, small kids: Medico-scientific, familial and advocacy visions of human brains. *BioSocieties*, *11*(3), pp.296-316.

[169] Alemayehu, D. and Berger, M.L., 2016. Big Data: transforming drug development and health policy decision making. *Health services and outcomes research methodology*, *16*(3), pp.92-102.

[170] Behadada, O., Trovati, M., Chikh, M.A. and Bessis, N., 2016. Big data-based extraction of fuzzy partition rules for heart arrhythmia detection: a semi-automated approach. *Concurrency and Computation: Practice and Experience*, *28*(2), pp.360-373.

[171] Ow, G.S. and Kuznetsov, V.A., 2016. Big genomics and clinical data analytics strategies for precision cancer prognosis. *Scientific reports*, *6*, p.36493.

[172] Lv, Z., Chirivella, J. and Gagliardo, P., 2016. Bigdata oriented multimedia mobile health applications. *Journal of medical systems*, *40*(5), p.120.

[173] Lossio-Ventura, J.A., Jonquet, C., Roche, M. and Teisseire, M., 2016. Biomedical term extraction: overview and a new methodology. *Information Retrieval Journal*, *19*(1-2), pp.59-99.

[174] Wang, H. and Yoon, S.W., 2015, October. Breast cancer prediction using data mining method. In *IIE Annu. Conf. Expo 2015* (pp. 818-828).

[175] Zhang, G., Bie, R. and Sun, Y., 2016, June. Bring biomedical ontologies to personalized healthcare: a smart inquiry framework. In *2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)* (pp. 84-88). IEEE.

[176] Chawla, N.V. and Davis, D.A., 2013. Bringing big data to personalized healthcare: a patient-centered framework. *Journal of general internal medicine*, *28*(3), pp.660-665.

[177] Murphy, S., Goodson, A., Mendis, M., Murphy, M., Phillips, L., Wang, Y. and Herrick, C., 2016. BRINGING HEALTHCARE ANALYTICS TO WHERE BIG DATA RESIDES USING A DISTRIBUTED QUERY SYSTEM. *IADIS International Journal on Computer Science & Information Systems*, *11*(2).

[178] Kocbek, S., Kocbek, P., Stozer, A., Zupanic, T., Groza, T. and Stiglic, G., 2018. Building interpretable models for polypharmacy prediction in older chronic patients based on drug prescription records. *PeerJ*, *6*, p.e5765.

[179] Moustakas, A., 2018. Assessing the predictive causality of individual based models using Bayesian inference intervention analysis: an application in epidemiology. *Stochastic environmental research and risk assessment*, *32*(10), pp.2861-2869.

[180] Wozniak, J.M., Jain, R., Balaprakash, P., Ozik, J., Collier, N.T., Bauer, J., Xia, F., Brettin, T., Stevens, R., Mohd-Yusof, J. and Cardona, C.G., 2018. CANDLE/Supervisor: A workflow framework for machine learning applied to cancer research. *BMC bioinformatics*, *19*(18), p.491.

[181] Stuckey, T.D., Gammon, R.S., Goswami, R., Depta, J.P., Steuter, J.A., Meine III, F.J., Roberts, M.C., Singh, N., Ramchandani, S., Burton, T. and Grouchy, P., 2018. Cardiac Phase Space Tomography: A novel method of assessing coronary artery disease utilizing machine learning. *PloS one*, *13*(8), p.e0198603.

[182] Zhang, H., Ma, S., Feng, Z., Wang, D., Li, C., Cao, Y., Chen, X., Liu, A., Zhu, Z., Zhang, J. and Zhang, G., 2016. Cardiovascular disease chemogenomics knowledgebase-guided target

identification and drug synergy mechanism study of an herbal formula. *Scientific reports*, *6*, p.33963.

[183] Zhang, H., Ma, S., Feng, Z., Wang, D., Li, C., Cao, Y., Chen, X., Liu, A., Zhu, Z., Zhang, J. and Zhang, G., 2016. Cardiovascular disease chemogenomics knowledgebase-guided target identification and drug synergy mechanism study of an herbal formula. *Scientific reports*, *6*, p.33963.

[184] Kunwar, V., Chandel, K., Sabitha, A.S. and Bansal, A., 2016, January. Chronic Kidney Disease analysis using data mining classification techniques. In *2016 6th International Conference-Cloud System and Big Data Engineering (Confluence)* (pp. 300-305). IEEE.

[185] Al'Aref, S.J., Anchouche, K., Singh, G., Slomka, P.J., Kolli, K.K., Kumar, A., Pandey, M., Maliakal, G., van Rosendael, A.R., Beecy, A.N. and Berman, D.S., 2018. Clinical applications of machine learning in cardiovascular disease and its relevance to cardiac imaging. *European heart journal*, *40*(24), pp.1975-1986.

[186] Salcedo-Bernal, A., Villamil-Giraldo, M.P. and Moreno-Barbosa, A.D., 2016. Clinical Data Analysis: An opportunity to compare machine learning methods. *Procedia Computer Science*, *100*, pp.731-738.

[187] Roederer, A., Dimartino, J., Gutsche, J., Mullen-Fortino, M., Shah, S., Hanson, C.W. and Lee, I., 2016, June. Clinician-in-the-Loop Annotation of ICU Bedside Alarm Data. In *2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)* (pp. 229-237). IEEE.

[188] Hans, A., Kaur, S. and Singh, N., 2016. Cloud Based Data Mining Model for Asthma Diagnosis. *International Journal of Grid and Distributed Computing*, *9*(9), pp.317-326.

[189] Vukićević, M., Radovanović, S., Milovanović, M. and Minović, M., 2014. Cloud based metalearning system for predictive modeling of biomedical data. *The Scientific World Journal*, *2014*.

[190] Navale, V. and Bourne, P.E., 2018. Cloud computing applications for biomedical science: A perspective. PLoS computational biology, 14(6), p.e1006144.

[191] Mahmud, S., Iqbal, R. and Doctor, F., 2016. Cloud enabled data analytics and visualization framework for health-shocks prediction. Future Generation Computer Systems, 65, pp.169-181.

[192] Ho, C.S. and Lin, I.P., 2013. Cloud-based Bayesian inference for online people health status assessment. In Applied Mechanics and Materials (Vol. 303, pp. 2231-2234). Trans Tech Publications.

[193] Vaska, J.S. and Sowjanya, A.M., 2015. Clustering diabetics data Using M-CFICA. International Journal of Advanced Computer Research, 5(20), p.327.

[194] Elbattah, M. and Molloy, O., 2017, March. Clustering-Aided approach for predicting patient outcomes with application to Elderly Healthcare in Ireland. In Workshops at the Thirty-First AAAI Conference on Artificial Intelligence.

[195] Diamond, C.C., Mostashari, F. and Shirky, C., 2009. Collecting and sharing data for population health: a new paradigm. Health affairs, 28(2), pp.454-466.

[196] Belyi, E., Giabbanelli, P.J., Patel, I., Balabhadrapathruni, N.H., Abdallah, A.B., Hameed, W. and Mago, V.K., 2016. Combining association rule mining and network analysis for pharmacosurveillance. The Journal of Supercomputing, 72(5), pp.2014-2034.

[197] Ceglowski, R., Churilov, L. and Wasserthiel, J., 2007. Combining data mining and discrete event simulation for a value-added view of a hospital emergency department. Journal of the Operational Research Society, 58(2), pp.246-254.

[199] Gehrmann, S., Dernoncourt, F., Li, Y., Carlson, E.T., Wu, J.T., Welt, J., Foote Jr, J., Moseley, E.T., Grant, D.W., Tyler, P.D. and Celi, L.A., 2018. Comparing deep learning and concept extraction based methods for patient phenotyping from clinical narratives. PloS one, 13(2), p.e0192360.

[200] Papachristou, N., Miaskowski, C., Barnaghi, P., Maguire, R., Farajidavar, N., Cooper, B. and Hu, X., 2016, November. Comparing machine learning clustering with latent class analysis on cancer symptoms' data. In *2016 IEEE Healthcare Innovation Point-Of-Care Technologies Conference (HI-POCT)* (pp. 162-166). IEEE.

[201] Shahmoradi, L., Langarizadeh, M. and Pourmand, G., 2016. Comparing three data mining methods to predict kidney transplant survival. *Acta Informatica Medica*, *24*(5), p.322.

[202] Afzal, M., Hussain, M., Khan, W.A., Ali, T., Lee, S., Huh, E.N., Ahmad, H.F., Jamshed, A., Iqbal, H., Irfan, M. and Hydari, M.A., 2017. Comprehensible knowledge model creation for cancer treatment decision making. *Computers in biology and medicine*, *82*, pp.119-129.

[203] Liang, Y. and Kelemen, A., 2017. Computational dynamic approaches for temporal omics data with applications to systems medicine. *BioData mining*, *10*(1), p.20.

[204] Chang, V., 2018. Computational intelligence for medical imaging simulations. *Journal of medical systems*, *42*(1), p.10.

[205] Chicco, D. and Rovelli, C., 2019. Computational prediction of diagnosis and feature selection on mesothelioma patient health records. *PloS one*, *14*(1), p.e0208737.

[208] Nammour, F., Danas, K. and Mansour, N., 2016, September. CorporateMeasures: A clinical analytics framework leading to clinical intelligence. In *2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom)* (pp. 1-6). IEEE.

[209] He, Q., Zhu, X., Li, D., Wang, S., Shen, J. and Yang, Y., 2017, June. Cost-effective Big Data Mining in the Cloud: A Case Study with K-means. In *2017 IEEE 10th International Conference on Cloud Computing (CLOUD)* (pp. 74-81). IEEE.

[210] Lorberbaum, T., Sampson, K.J., Chang, J.B., Iyer, V., Woosley, R.L., Kass, R.S. and Tatonetti, N.P., 2016. Coupling data mining and laboratory experiments to discover drug interactions causing QT prolongation. *Journal of the American College of Cardiology*, *68*(16), pp.1756-1764.

[211] Lorberbaum, T., Sampson, K.J., Chang, J.B., Iyer, V., Woosley, R.L., Kass, R.S. and Tatonetti, N.P., 2016. Coupling data mining and laboratory experiments to discover drug interactions causing QT prolongation. *Journal of the American College of Cardiology*, *68*(16), pp.1756-1764.

[212] Chang, V., 2018. Data analytics and visualization for inspecting cancers and genes. *Multimedia Tools and Applications*, *77*(14), pp.17693-17707.

[213] Spinsante, S., Gambi, E., Montanini, L. and Raffaeli, L., 2015, December. Data management in ambient assisted living platforms approaching IoT: A case study. In *2015 IEEE Globecom Workshops (GC Wkshps)* (pp. 1-7). IEEE.

[214] Gupta, M., Solanki, V.K., Singh, V.K. and García-Díaz, V., 2018. 4Data Mining Approach of Accident Occurrences Identification with Effective Methodology and Implementation. *International Journal of Electrical & Computer Engineering (2088-8708)*, *8*.

[216] Banaee, H., Ahmed, M.U. and Loutfi, A., 2013. Data mining for wearable sensors in health monitoring systems: a review of recent trends and challenges. *Sensors*, *13*(12), pp.17472-17500.

[218] Ranjan, J., 2009. Data mining in pharma sector: benefits. *International journal of health care quality assurance*, *22*(1), pp.82-92.

[219] Abdar, M., Kalhori, S.R.N., Sutikno, T., Subroto, I.M.I. and Arji, G., 2015. Comparing Performance of Data Mining Algorithms in Prediction Heart Diseases. *International Journal of Electrical & Computer Engineering (2088-8708)*, *5*(6).

[220] Mani, A., Ravindran, R., Mannepalli, S., Vang, D., Luciw, P.A., Hogarth, M., Khan, I.H. and Krishnan, V.V., 2015. Data mining strategies to improve multiplex microbead immunoassay tolerance in a mouse model of infectious diseases. *PloS one*, *10*(1), p.e0116262.

[221] Vararuk, A., Petrounias, I. and Kodogiannis, V., 2008. Data mining techniques for HIV/AIDS data management in Thailand. *Journal of Enterprise Information Management*.

[222] Hsiao, K.F. and Rashvand, H.F., 2015. Data modeling mobile augmented reality: integrated mind and body rehabilitation. *Multimedia Tools and Applications*, *74*(10), pp.3543-3560.

[223] Almeida, H., Meurs, M.J., Kosseim, L. and Tsang, A., 2016. Data sampling and supervised learning for hiv literature screening. *IEEE transactions on nanobioscience*, *15*(4), pp.354-361.

[224] Segall, R.S. and Zhang, Q., 2006. Data visualization and data mining of continuous numerical and discrete nominal-valued microarray databases for bioinformatics. *Kybernetes*, *35*(10), pp.1538-1566.

[225] Kintsakis, A.M., Psomopoulos, F.E. and Mitkas, P.A., 2016. Data-aware optimization of bioinformatics workflows in hybrid clouds. *Journal of Big Data*, *3*(1), p.20.

[226] Ho, T.B., Le, L., Thai, D.T. and Taewijit, S., 2016. Data-driven approach to detect and predict adverse drug reactions. *Current pharmaceutical design*, *22*(23), pp.3498-3526.

[229] Robson, B. and Boray, S., 2016. Data-mining to build a knowledge representation store for clinical decision support. Studies on curation and validation based on machine performance in multiple choice medical licensing examinations. *Computers in biology and medicine*, *73*, pp.71-93.

[230] Feuerriegel, S., 2016. Decision support in healthcare: determining provider influence on treatment outcomes with robust risk adjustment. *Journal of Decision systems*, *25*(4), pp.371-390.

[231] Goh, W.P., Tao, X., Zhang, J. and Yong, J., 2016. Decision support systems for adoption in dental clinics: a survey. *Knowledge-Based Systems*, *104*, pp.195-206.

[232] van Wijk, Y., Halilaj, I., van Limbergen, E., Walsh, S., Lutgens, L., Lambin, P. and Vanneste, B.G.L., 2019. Decision Support Systems in Prostate Cancer Treatment: An Overview. *BioMed Research International*, *2019*.

[233] Mansour, A.M., 2018. Decision tree-based expert system for adverse drug reaction detection using fuzzy logic and genetic algorithm. *International Journal of Advanced Computer Research*, *8*(36), pp.110-128.

[235] Zhou, S.M., Fernandez-Gutierrez, F., Kennedy, J., Cooksey, R., Atkinson, M., Denaxas, S., Siebert, S., Dixon, W.G., O'Neill, T.W., Choy, E. and Sudlow, C., 2016. Defining disease phenotypes in primary care electronic health records by a machine learning approach: a case study in identifying rheumatoid arthritis. *PloS one*, *11*(5), p.e0154515.

[236] Kim, W., Won, J.H., Park, S. and Kang, J., 2015. Demand forecasting models for medicines through wireless sensor networks data and topic trend analysis. *International Journal of Distributed Sensor Networks*, *11*(9), p.907169.

[237] Kuo, M.H., Chrimes, D., Moa, B. and Hu, W., 2015, December. Design and construction of a big data analytics framework for health applications. In *2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity)* (pp. 631-636). IEEE.

[238] Jin, W. and Kim, D., 2018. Design and implementation of e-health system based on semantic sensor network using IETF YANG. *Sensors*, *18*(2), p.629.

[239] Gujral, S., Rathore, A. and Chauhan, S., 2017. Detecting and Predicting Diabetes Using Supervised Learning: An Approach towards Better Healthcare for Women. *International Journal of Advanced Research in Computer Science*, *8*(5).

[240] Shi, J., Zheng, M., Yao, L. and Ge, Y., 2018. Developing a healthcare dataset information resource (DIR) based on Semantic Web. *BMC medical genomics*, *11*(5), p.102.

[241] Corey, K.M., Kashyap, S., Lorenzi, E., Lagoo-Deenadayalan, S.A., Heller, K., Whalen, K., Balu, S., Heflin, M.T., McDonald, S.R., Swaminathan, M. and Sendak, M., 2018. Development and validation of machine learning models to identify high-risk surgical patients using automatically curated electronic health record data (Pythia): A retrospective, single-site study. *PLoS medicine*, *15*(11), p.e1002701.

[242] Wong, B., Ho, G.T. and Tsui, E., 2017. Development of an intelligent e-healthcare system for the domestic care industry. *Industrial Management & Data Systems*, *117*(7), pp.1426-1445.

[243] Nor, N.A.M., Taib, N.A., Saad, M., Zaini, H.S., Ahmad, Z., Ahmad, Y. and Dhillon, S.K., 2019. Development of electronic medical records for clinical and research purposes: the breast cancer module using an implementation framework in a middle income country-Malaysia. *BMC bioinformatics*, *19*(13), p.402.

[244] Khan, S.I. and Hoque, A.S.M.L., 2015, December. Development of national health data warehouse Bangladesh: Privacy issues and a practical solution. In *2015 18th International Conference on Computer and Information Technology (ICCIT)* (pp. 373-378). IEEE.

[245] Kumar, M.A. and Aroquiaraj, I.L., 2018. DIABETES DATA ANALYSIS USING MAPREDUCE AND CLASSIFICATION TECHNIQUES. *International Journal of Advanced Research in Computer Science*, *9*(Special Issue 1), p.49.

[246] Sharmila, K. and Manickam, S., 2016. Diagnosing diabetic dataset using Hadoop and K-means clustering techniques. *Indian Journal of Science and Technology*, *9*(40), pp.1-5.

[247] Lin, C., Song, Z., Song, H., Zhou, Y., Wang, Y. and Wu, G., 2016. Differential privacy preserving in big data analytics for connected health. *Journal of medical systems*, *40*(4), p.97.

[248] Gligorijevic, D., Stojanovic, J. and Obradovic, Z., 2016. Disease types discovery from a large database of inpatient records: A sepsis study. *Methods*, *111*, pp.45-55.

[249] Popovic, J.R., 2017. Distributed data networks: a blueprint for Big Data sharing and healthcare analytics. *Annals of the New York Academy of Sciences*, *1387*(1), pp.105-111.

[250] DencelinX, L. and Ramkumar, T., 2017. Distributed Machine Learning Algorithms to classify Protein secondary structures for Drug Design-A Survey. *Research Journal of Pharmacy and Technology*, *10*(9), pp.3173-3180.

[251] Christie, S.A., Conroy, A.S., Callcut, R.A., Hubbard, A.E. and Cohen, M.J., 2019. Dynamic multi-outcome prediction after injury: Applying adaptive machine learning for precision medicine in trauma. *PloS one*, *14*(4), p.e0213836.

[252] White, R.W., Wang, S., Pant, A., Harpaz, R., Shukla, P., Sun, W., DuMouchel, W. and Horvitz, E., 2016. Early identification of adverse drug reactions from search log data. *Journal of biomedical informatics*, *59*, pp.42-48.

[254] Chen, H., Martin, B., Daimon, C.M. and Maudsley, S., 2013. Effective use of latent semantic indexing and computational linguistics in biological and biomedical applications. *Frontiers in physiology*, *4*, p.8.

[255] Richter, A.N. and Khoshgoftaar, T.M., 2019. Efficient learning from big data for cancer risk modeling: A case study with melanoma. *Computers in biology and medicine*, *110*, pp.29-39.

[256] Sathya, S. and Sethukarasi, T., 2016, February. Efficient privacy preservation technique for healthcare records using big data. In *2016 International Conference on Information Communication and Embedded Systems (ICICES)* (pp. 1-6). IEEE.

[258] Sadrawi, M., Sun, W.Z., Ma, M., Yeh, Y.T., Abbod, M. and Shieh, J.S., 2018. Ensemble Genetic Fuzzy Neuro Model Applied for the Emergency Medical Service via Unbalanced Data Evaluation. *Symmetry*, *10*(3), p.71.

[259] Park, J.A., Kim, M. and Yoon, S., 2016. Evaluation of Large-scale Data to Detect Irregularity in Payment for Medical Services. *Methods of information in medicine*, *55*(03), pp.284-291.

[260] Gomathi, R.M., Joseph, O. and Gupta, V.K., 2016. Evidence Based Disease Analysis using Big Data. *RESEARCH JOURNAL OF PHARMACEUTICAL BIOLOGICAL AND CHEMICAL SCIENCES*, *7*(4), pp.352-358.

[261] Pasupathi, C. and Kalavakonda, V., 2016, February. Evidence Based health care system using Big Data for disease diagnosis. In *2016 2nd International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB)* (pp. 743-747). IEEE.

[262] Wang, Z., Wen, X., Lu, Y., Yao, Y. and Zhao, H., 2016. Exploiting machine learning for predicting skeletal-related events in cancer patients with bone metastases. *Oncotarget*, *7*(11), p.12612.

[263] Harjumaa, M., Saraniemi, S., Pekkarinen, S., Lappi, M., Similä, H. and Isomursu, M., 2016. Feasibility of digital footprint data for health analytics and services: an explorative pilot study. *BMC medical informatics and decision making*, *16*(1), p.139.

[265] Wu, L., Hu, Y., Liu, X., Zhang, X., Chen, W., Alan, S.L., Kellum, J.A., Waitman, L.R. and Liu, M., 2018. Feature Ranking in Predictive Models for Hospital-Acquired Acute Kidney Injury. *Scientific reports*, *8*(1), p.17298.

[266] El Moudden, I., Ouzir, M. and ElBernoussi, S., 2017. Feature selection and extraction for class prediction in dysphonia measures analysis: A case study on Parkinson's disease speech rehabilitation. *Technology and Health Care*, *25*(4), pp.693-708.

[267] Thippa Reddy, G. and Khare, N., 2016. FFBAT-optimized rule based fuzzy logic classifier for diabetes. In *International Journal of Engineering Research in Africa* (Vol. 24, pp. 137-152). Trans Tech Publications.

[268] Sharma, A., Hostetter, J., Morrison, J., Wang, K. and Siegel, E., 2016. Focused decision support: A data mining tool to query the prostate, lung, colorectal, and ovarian cancer screening trial dataset and guide screening management for the individual patient. *Journal of digital imaging*, *29*(2), pp.160-164.

[269] Singh, S., Bansal, A., Sandhu, R. and Sidhu, J., 2018. Fog computing and IoT based healthcare support service for dengue fever. *International Journal of Pervasive Computing and Communications*, *14*(2), pp.197-207.

[271] Gia, T.N., Jiang, M., Rahmani, A.M., Westerlund, T., Liljeberg, P. and Tenhunen, H., 2015, October. Fog computing in healthcare internet of things: A case study on ecg feature extraction. In *2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing* (pp. 356-363). IEEE.

[272] Paul, A., Pinjari, H., Hong, W.H., Seo, H.C. and Rho, S., 2018. Fog computing-based IoT for health monitoring system. *Journal of Sensors*, *2018*.

[273] Mavragani, A. and Ochoa, G., 2018. Forecasting AIDS prevalence in the United States using online search traffic data. *Journal of Big Data*, *5*(1), p.17.

[274] Pires, I., Garcia, N., Pombo, N. and Flórez-Revuelta, F., 2016. From data acquisition to data fusion: a comprehensive review and a roadmap for the identification of activities of daily living using mobile devices. *Sensors*, *16*(2), p.184.

[275] Dipnall, J.F., Pasco, J.A., Berk, M., Williams, L.J., Dodd, S., Jacka, F.N. and Meyer, D., 2016. Fusing data mining, machine learning and traditional statistics to detect biomarkers associated with depression. *PloS one*, *11*(2), p.e0148195.

[276] Johnson, D.E., 2013. Fusion of nonclinical and clinical data to predict human drug safety. *Expert review of clinical pharmacology*, *6*(2), pp.185-195.

[277] Vashistha, R., Dangi, A.K., Kumar, A., Chhabra, D. and Shukla, P., 2018. Futuristic biosensors for cardiac health care: an artificial intelligence approach. *3 Biotech*, *8*(8), p.358.

[278] Ali, F., El-Sappagh, S. and Kwak, D., 2019. Fuzzy Ontology and LSTM-Based Text Mining: A Transportation Network Monitoring System for Assisting Travel. *Sensors*, *19*(2), p.234.

[279] Navarro, F.C., Mohsen, H., Yan, C., Li, S., Gu, M., Meyerson, W. and Gerstein, M., 2019. Genomics and data science: an application within an umbrella. *Genome biology*, *20*(1), p.109.

[280] Tresp, V., Overhage, J.M., Bundschus, M., Rabizadeh, S., Fasching, P.A. and Yu, S., 2016. Going digital: a survey on digitalization and large-scale data analytics in healthcare. *Proceedings of the IEEE*, *104*(11), pp.2180-2206.

[281] Liu, J., Bier, E., Wilson, A., Guerra-Gomez, J.A., Honda, T., Sricharan, K., Gilpin, L. and Davies, D., 2016. Graph analysis for detecting fraud, waste, and abuse in healthcare data. *AI Magazine*, *37*(2), pp.33-46.

[282] Luo, W., Phung, D., Tran, T., Gupta, S., Rana, S., Karmakar, C., Shilton, A., Yearwood, J., Dimitrova, N., Ho, T.B. and Venkatesh, S., 2016. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. *Journal of medical Internet research*, *18*(12), p.e323.

[283] Almasoud, A.M., Al-Khalifa, H.S. and Al-Salman, A.S., 2019. Handling Big Data Scalability in Biological Domain Using Parallel and Distributed Processing: A Case of Three Biological Semantic Similarity Measures. *BioMed research international*, *2019*.

[284] Mirkes, E.M., Coats, T.J., Levesley, J. and Gorban, A.N., 2016. Handling missing data in large healthcare dataset: A case study of unknown trauma outcomes. *Computers in biology and medicine*, *75*, pp.203-216.

[285] McPadden, J., Durant, T.J., Bunch, D.R., Coppi, A., Price, N., Rodgerson, K., Torre Jr, C.J., Byron, W., Hsiao, A.L., Krumholz, H.M. and Schulz, W.L., 2019. Health Care and Precision Medicine Research: Analysis of a Scalable Data Science Platform. *Journal of medical Internet research*, *21*(4), p.e13043.

[286] Venkatraman, S., Sundarraj, R.P. and Seethamraju, R., 2015. Healthcare Analytics Adoption-Decision Model: A Case Study. In *PACIS* (p. 51).

[287] Hossain, M.S. and Muhammad, G., 2016. Healthcare big data voice pathology assessment framework. *iEEE Access*, *4*, pp.7806-7815.

[288] Lin, Y.K., Chen, H., Brown, R.A., Li, S.H. and Yang, H.J., 2017. Healthcare predictive analytics for risk profiling in chronic care: A Bayesian multitask learning approach. *Mis Quarterly*, *41*(2).

[289] Sahoo, S.S., Jayapandian, C., Garg, G., Kaffashi, F., Chung, S., Bozorgi, A., Chen, C.H., Loparo, K., Lhatoo, S.D. and Zhang, G.Q., 2013. Heart beats in the cloud: distributed analysis of electrophysiological 'Big Data'using cloud computing for epilepsy clinical research. *Journal of the American Medical Informatics Association*, *21*(2), pp.263-271.

[291] Kaur, A. and Arora, J., 2018. HEART DISEASE PREDICTION USING DATA MINING TECHNIQUES: A SURVEY. *International Journal of Advanced Research in Computer Science*, *9*(2).

[292] Shao, Y., Wang, K., Shu, L., Deng, S. and Deng, D.J., 2016. Heuristic optimization for reliable data congestion analytics in crowdsourced ehealth networks. *IEEE Access*, *4*, pp.9174-9183.

[293] Jain, P., Agarwal, A., Behara, R. and Baechle, C., 2019. HPCC based framework for COPD readmission risk analysis. *Journal of Big Data*, *6*(1), p.26.

[294] Jo, B., Khan, R. and Lee, Y.S., 2018. Hybrid Blockchain and internet-of-things network for underground structure health monitoring. *Sensors*, *18*(12), p.4268.

[295] Tomar, D. and Agarwal, S., 2015. Hybrid feature selection based weighted least squares twin support vector machine approach for diagnosing breast cancer, hepatitis, and diabetes. *Advances in Artificial Neural Systems*, *2015*, p.1.

[296] Manogaran, G., Varatharajan, R. and Priyan, M.K., 2018. Hybrid recommendation system for heart disease diagnosis based on multiple kernel learning with adaptive neuro-fuzzy inference system. *Multimedia tools and applications*, *77*(4), pp.4379-4399.

[297] Wiharto, W., Kusnanto, H. and Herianto, H., 2017. Hybrid system of tiered multivariate analysis and artificial neural network for coronary heart disease diagnosis. *International Journal of Electrical and Computer Engineering*, *7*(2), p.1023.

[298] Carson, N.J., Mullin, B., Sanchez, M.J., Lu, F., Yang, K., Menezes, M. and Le Cook, B., 2019. Identification of suicidal behavior among psychiatrically hospitalized adolescents using natural language processing and machine learning of electronic health records. *PloS one*, *14*(2), p.e0211116.

[299] Bansal, V., Poddar, A. and Ghosh-Roy, R., 2019. Identifying a Medical Department Based on Unstructured Data: A Big Data Application in Healthcare. *Information*, *10*(1), p.25.

[300] Hansen, P.W., Clemmensen, L., Sehested, T.S., Fosbøl, E.L., Torp-Pedersen, C., Køber, L., Gislason, G.H. and Andersson, C., 2016. Identifying Drug–Drug Interactions by Data Mining: A Pilot Study of Warfarin-Associated Drug Interactions. *Circulation: Cardiovascular Quality and Outcomes*, *9*(6), pp.621-628.

[301] Bahja, M. and Lycett, M., 2016, December. Identifying patient experience from online resources via sentiment analysis and topic modelling. In *Proceedings of the 3rd IEEE/ACM International Conference on Big Data Computing, Applications and Technologies* (pp. 94-99). ACM.

[302] Fernandes, A.C., Dutta, R., Velupillai, S., Sanyal, J., Stewart, R. and Chandran, D., 2018. Identifying suicide ideation and suicidal attempts in a psychiatric clinical research database using natural language processing. *Scientific reports*, *8*(1), p.7426.

[303] Rallapalli, S., Gondkar, R. and Ketavarapu, U.P.K., 2016. Impact of processing and analyzing healthcare big data on cloud computing environment by implementing hadoop cluster. *Procedia Computer Science*, *85*, pp.16-22.

[304] Duggal, R., Shukla, S., Chandra, S., Shukla, B. and Khatri, S.K., 2016. Impact of selected pre-processing techniques on prediction of risk of early readmission for diabetic patients in India. *International Journal of Diabetes in Developing Countries*, *36*(4), pp.469-476.

[305] Robson, B. and Boray, S., 2015. Implementation of a web based universal exchange and inference language for medicine: sparse data, probabilities and inference in data mining of clinical data repositories. *Computers in biology and medicine*, *66*, pp.82-102.

[306] Hassler, A.P., Menasalvas, E., García-García, F.J., Rodríguez-Mañas, L. and Holzinger, A., 2019. Importance of medical data preprocessing in predictive modeling and risk factor discovery for the frailty syndrome. *BMC medical informatics and decision making*, *19*(1), p.33.

[307] Senkamalavalli, R. and Bhuvaneswari, T., 2017. IMPROVED CLASSIFICATION OF BREAST CANCER DATA USING HYBRID TECHNIQUES. *International Journal of Advanced Research in Computer Science*, *8*(8).

[308] Woodbridge, J., Mortazavi, B., Bui, A.A. and Sarrafzadeh, M., 2016. Improving biomedical signal search results in big data case-based reasoning environments. *Pervasive and mobile computing*, *28*, pp.69-80.

[309] Avati, A., Jung, K., Harman, S., Downing, L., Ng, A. and Shah, N.H., 2018. Improving palliative care with deep learning. *BMC medical informatics and decision making*, *18*(4), p.122.

[310] Afzal, Z., Schuemie, M.J., van Blijderveen, J.C., Sen, E.F., Sturkenboom, M.C. and Kors, J.A., 2013. Improving sensitivity of machine learning methods for automated case identification from free-text electronic medical records. *BMC medical informatics and decision making*, *13*(1), p.30.

[311] Spiranovic, C., Matthews, A., Scanlan, J. and Kirkby, K.C., 2016. Increasing knowledge of mental illness through secondary research of electronic health records: opportunities and challenges. *Advances in Mental Health*, *14*(1), pp.14-25.

[312] Dayal, M. and Singh, N., 2016. Indian Health Care Analysis using Big Data Programming Tool. *Procedia Computer Science*, *89*, pp.521-527.

[313] Tilve, C.A., Ayora, A.P., Ruiz, C.R., Llamas, D.G., Carrajo, L.G., Blanco, F.G. and Vázquez, G.G., 2015. Integrating medical and research information: a big data approach. *Studies in health technology and informatics*, *210*, pp.707-711.

[314] Krempel, R., Kulkarni, P., Yim, A., Lang, U., Habermann, B. and Frommolt, P., 2018. Integrative analysis and machine learning on cancer genomics data using the Cancer Systems Biology Database (CancerSysDB). *BMC bioinformatics*, *19*(1), p.156.

[315] Xin, B., Xu, C., Wang, L., Dong, T., Zheng, C. and Wang, X., 2018, November. Integrative Clustering and Supervised Feature Selection for Clinical Applications. In *2018 15th International Conference on Control, Automation, Robotics and Vision (ICARCV)* (pp. 1316-1320). IEEE.

[316] Teow, K.L., El-Darzi, E., Foo, C., Jin, X. and Sim, J., 2012. Intelligent analysis of acute bed overflow in a tertiary hospital in Singapore. *Journal of medical systems*, *36*(3), pp.1873-1882.

[317] Li, X., Wang, H., He, H., Du, J., Chen, J. and Wu, J., 2019. Intelligent diagnosis with Chinese electronic medical records based on convolutional neural networks. *BMC bioinformatics*, *20*(1), p.62.

[318] Batool, H., Akram, M.U., Batool, F. and Butt, W.H., 2016. Intelligent framework for diagnosis of frozen shoulder using cross sectional survey and case studies. *SpringerPlus*, *5*(1), p.1840.

[319] Choi, J., Choi, C., Ko, H. and Kim, P., 2016. Intelligent Healthcare Service Using Health Lifelog Analysis. *Journal of medical systems*, *40*(8), p.188.

[321] Chrimes, D., Moa, B., Zamani, H. and Kuo, M.H., 2016, August. Interactive healthcare big data analytics platform under simulated performance. In *2016 IEEE 14th Intl Conf on Dependable, Autonomic and Secure Computing, 14th Intl Conf on Pervasive Intelligence and Computing, 2nd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech)* (pp. 811-818). IEEE.

[322] Dipnall, J.F., Pasco, J.A., Berk, M., Williams, L.J., Dodd, S., Jacka, F.N. and Meyer, D., 2016. Into the bowels of depression: unravelling medical symptoms associated with depression by applying

machine-learning techniques to a community based population sample. *PloS one*, *11*(12), p.e0167055.

[323] Wang, J.Y., Liu, C.S., Lung, C.H., Yang, Y.T. and Lin, M.H., 2017. Investigating spousal concordance of diabetes through statistical analysis and data mining. *PloS one*, *12*(8), p.e0183413.

[325] Hosseini, Z.Z. and Mohammadzadeh, M., 2016. Knowledge discovery from patients' behavior via clustering-classification algorithms based on weighted eRFM and CLV model: An empirical study in public health care services. *Iranian journal of pharmaceutical research: IJPR*, *15*(1), p.355.

[326] Rao, R.R. and Makkithaya, K., 2017. Learning from a Class Imbalanced Public Health Dataset: a Cost-based Comparison of Classifier Performance. *International Journal of Electrical and Computer Engineering*, *7*(4), p.2215.

[327] Eze, B., Kuziemsky, C., Lakhani, R. and Peyton, L., 2016. Leveraging cloud computing for systematic performance management of quality of care. *Procedia Computer Science*, *98*, pp.316-323.

[328] Morid, M.A., Sheng, O.R. and Abdelrahman, S., 2017. Leveraging Time Series Data in Similarity Based Healthcare Predictive Models: The Case of Early ICU Mortality Prediction.

[329] Wu, H., Yang, J.J. and Li, J., 2014, October. Low redundancy feature selection with grouped variables and its application to healthcare data. In *2014 IEEE International Conference on Big Data (Big Data)* (pp. 71-76). IEEE.

[330] Nikan, S., Gwadry-Sridhar, F. and Bauer, M., 2016, December. Machine learning application to predict the risk of coronary artery atherosclerosis. In *2016 International conference on computational science and computational intelligence (CSCI)* (pp. 34-39). IEEE.

[331] Pan, L., Liu, G., Lin, F., Zhong, S., Xia, H., Sun, X. and Liang, H., 2017. Machine learning applications for prediction of relapse in childhood acute lymphoblastic leukemia. *Scientific reports*, *7*(1), p.7402.

[332] Finkelstein, J. and cheol Jeong, I., 2017. Machine learning approaches to personalize early prediction of asthma exacerbations. *Annals of the New York Academy of Sciences*, *1387*(1), p.153.

[333] Jagga, Z. and Gupta, D., 2015. Machine learning for biomarker identification in cancer research–developments toward its clinical application. *Personalized medicine*, *12*(4), pp.371-387.

[335] Kubota, K.J., Chen, J.A. and Little, M.A., 2016. Machine learning for large-scale wearable sensor data in Parkinson's disease: Concepts, promises, pitfalls, and futures. *Movement disorders*, *31*(9), pp.1314-1326.

[336] Weiss, J.C., Natarajan, S., Peissig, P.L., McCarty, C.A. and Page, D., 2012. Machine learning for personalized medicine: Predicting primary myocardial infarction from electronic health records. *AI Magazine*, *33*(4), pp.33-33.

[337] Vellido, A., Ribas, V., Morales, C., Sanmartín, A.R. and Rodríguez, J.C.R., 2018. Machine learning in critical care: state-of-the-art and a sepsis case study. *Biomedical engineering online*, *17*(1), p.135.

[338] Wolff, P., Graña, M., Ríos, S.A. and Yarza, M.B., 2019. Machine Learning Readmission Risk Modeling: A Pediatric Case Study. *BioMed research international*, *2019*.

[339] De Silva, D., Ranasinghe, W., Bandaragoda, T., Adikari, A., Mills, N., Iddamalgoda, L., Alahakoon, D., Lawrentschuk, N., Persad, R., Osipov, E. and Gray, R., 2018. Machine learning to support social media empowered patients in cancer care and cancer treatment decisions. *PloS one*, *13*(10), p.e0205855.

[340] Karlsson, J. and Trelles, O., 2013. MAPI: a software framework for distributed biomedical applications. *Journal of biomedical semantics*, *4*(1), p.4.

[341] Dasgupta, D. and Chawla, N.V., 2016, October. MedCare: Leveraging Medication Similarity for Disease Prediction. In *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)* (pp. 706-715). IEEE.

[342] Manikandan, P., 2018. Medical Big Data Classification Using a Combination of Random Forest Classifier and K-Means Clustering. *International Journal of Intelligent Systems and Applications*, *10*(11), p.11.

[343] Sebaa, A., Chikh, F., Nouicer, A. and Tari, A., 2018. Medical big data warehouse: architecture and system design, a case study: improving healthcare resources distribution. *Journal of medical systems*, *42*(4), p.59.

[344] Tsai, C.W., Chiang, M.C., Ksentini, A. and Chen, M., 2016. Metaheuristic algorithms for healthcare: open issues and challenges. *Computers & Electrical Engineering*, *53*, pp.421-434.

[345] Concaro, S., Sacchi, L., Cerra, C. and Bellazzi, R., 2009, August. Mining administrative and clinical diabetes data with temporal association rules. In *MIE* (pp. 574-578).

[346] Sampathkumar, H., Chen, X.W. and Luo, B., 2014. Mining adverse drug reactions from online healthcare forums using hidden Markov model. *BMC medical informatics and decision making*, *14*(1), p.91.

[347] Mumini, O.O., Lingxue, R., Ivanov, K. and Wang, L., 2016, July. Mining brain features from schizophrenia studies with Shift-And pattern matching. In *2016 IEEE International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)* (pp. 157-163). IEEE.

[348] Wang, Y., Hou, W. and Wang, F., 2018. Mining co-occurrence and sequence patterns from cancer diagnoses in New York State. *PloS one*, *13*(4), p.e0194407.

[349] Jensen, P.B., Jensen, L.J. and Brunak, S., 2012. Mining electronic health records: towards better research applications and clinical care. *Nature Reviews Genetics*, *13*(6), p.395.

[350] Yang, F.C., Lee, A.J. and Kuo, S.C., 2016. Mining health social media with sentiment analysis. *Journal of medical systems*, *40*(11), p.236.

[351] Leary, A., Cook, R., Jones, S., Smith, J., Gough, M., Maxwell, E., Punshon, G. and Radford, M., 2016. Mining routinely collected acute data to reveal non-linear relationships between nurse staffing levels and outcomes. *BMJ open*, *6*(12), p.e011177.

[352] Wang, K., Shao, Y., Shu, L., Zhu, C. and Zhang, Y., 2016. Mobile big data fault-tolerant processing for ehealth networks. *IEEE Network*, *30*(1), pp.36-42.

[353] Gao, C., Sun, H., Wang, T., Tang, M., Bohnen, N.I., Müller, M.L., Herman, T., Giladi, N., Kalinin, A., Spino, C. and Dauer, W., 2018. Model-based and model-free machine learning techniques for diagnostic prediction and classification of clinical outcomes in Parkinson's disease. *Scientific reports*, *8*(1), p.7129.

[354] Masud, M.M. and Al Harahsheh, A.R., 2016, December. Mortality prediction of ICU patients using lab test data by feature vector compaction & classification. In *2016 IEEE International Conference on Big Data (Big Data)* (pp. 3404-3411). IEEE.

[356] Zhang, S., Liu, L., Li, H., Xiao, Z. and Cui, L., 2016, August. MTPGraph: A Data-Driven Approach to Predict Medical Risk Based on Temporal Profile Graph. In *2016 IEEE Trustcom/BigDataSE/ISPA* (pp. 1174-1181). IEEE.

[357] Ahmad, A., Paul, A., Din, S., Rathore, M.M., Choi, G.S. and Jeon, G., 2018. Multilevel data processing using parallel algorithms for analyzing big data in high-performance computing. *International Journal of Parallel Programming*, *46*(3), pp.508-527.

[358] Razzaghi, T., Roderick, O., Safro, I. and Marko, N., 2016. Multilevel weighted support vector machine for classification on healthcare data with missing values. *PloS one*, *11*(5), p.e0155119.

[359] Jin, B., Yang, H., Xiao, C., Zhang, P., Wei, X. and Wang, F., 2017, February. Multitask dyadic prediction and its application in prediction of adverse drug-drug interaction. In *Thirty-First AAAI Conference on Artificial Intelligence*.

[360] Quwaider, M. and Jararweh, Y., 2016. Multi-tier cloud infrastructure support for reliable global health awareness system. *Simulation Modelling Practice and Theory*, *67*, pp.44-58.

[361] Zhao, Y., Parvinzamir, F., Deng, Z., Wei, H., Zhao, X., Liu, E., Dong, F., Clapworthy, G., Lukoševičius, A., Marozas, V. and Kaldoudi, E., 2016. MyHealthAvatar and CARRE: case studies of interactive visualisation for internet-enabled sensor-assisted health monitoring and risk analysis. *IET Networks*, *5*(5), pp.114-121.

[362] Hickey, S.J., 2013. Naive Bayes classification of public health data with greedy feature selection. *Communications of the IIMA*, *13*(2), p.7.

[363] Chung, Y., Chen, Y.C. and Tsai, J.J., 2016, April. National Representatively Healthcare Database and the Application. In *2016 IEEE Second International Conference on Multimedia Big Data (BigMM)* (pp. 377-384). IEEE.

[364] Malak, J.S., Zeraati, H., Nayeri, F.S., Safdari, R. and Shahraki, A.D., 2019. Neonatal intensive care decision support systems using artificial intelligence techniques: A systematic review. *Artificial Intelligence Review*, *52*(4), pp.2685-2704.

[365] Johnson, O.A., Hall, P.S. and Hulme, C., 2016. NETIMIS: dynamic simulation of health economics outcomes using big data. *PharmacoEconomics*, *34*(2), pp.107-114.

[366] Zhang, X., Yuan, Z., Ji, J., Li, H. and Xue, F., 2016. Network or regression-based methods for disease discrimination: a comparison study. *BMC medical research methodology*, *16*(1), p.100.

[367] Gálvez, J.A., Jalali, A., Ahumada, L., Simpao, A.F. and Rehman, M.A., 2017. Neural network classifier for automatic detection of invasive versus noninvasive airway management technique based on respiratory monitoring parameters in a pediatric anesthesia. *Journal of medical systems*, *41*(10), p.153.

[368] Tan, J. and Wang, F., 2017. Non-Traditional Data Mining Applications in Taiwan National Health Insurance (NHI) Databases: A Hybrid Mining (HM) Case for the Framing of NHI Decisions. *International Journal of Healthcare Information Systems and Informatics (IJHISI)*, *12*(4), pp.31-51.

[369] Amin, M., Banos, O., Khan, W., Muhammad Bilal, H., Gong, J., Bui, D.M., Cho, S., Hussain, S., Ali, T., Akhtar, U. and Chung, T., 2016. On curating multimodal sensory data for health and wellness platforms. *Sensors*, *16*(7), p.980.

[370] Abar, O., Charnigo, R.J., Rayapati, A. and Kavuluru, R., 2016, October. On Interestingness Measures for Mining Statistically Significant and Novel Clinical Associations from EMRs. In *Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics* (pp. 587-594). ACM.

[371] Zhang, M.H., Ma, J.S., Shen, Y. and Chen, Y., 2016. Optimal classification for the diagnosis of duchenne muscular dystrophy images using support vector machines. *International journal of computer assisted radiology and surgery*, *11*(9), pp.1755-1763.

[372] He, H., Du, Z., Zhang, W. and Chen, A., 2016. Optimization strategy of Hadoop small file storage for big data in healthcare. *The Journal of Supercomputing*, *72*(10), pp.3696-3707.

[373] Tahmassebi, A., Gandomi, A.H., Schulte, M.H., Goudriaan, A.E., Foo, S.Y. and Meyer-Baese, A., 2018. Optimized naive-bayes and decision tree approaches for fmri smoking cessation classification. *Complexity*, *2018*.

[374] Partington, S.N., Papakroni, V. and Menzies, T., 2014. Optimizing data collection for public health decisions: a data mining approach. *BMC public health*, *14*(1), p.593.

[375] Chen, J.H., Podchiyska, T. and Altman, R.B., 2015. OrderRex: clinical order decision support and outcome predictions by data-mining electronic medical records. *Journal of the American Medical Informatics Association*, *23*(2), pp.339-348.

[376] Wiens, J., Guttag, J. and Horvitz, E., 2016. Patient risk stratification with time-varying parameters: a multitask learning approach. *The Journal of Machine Learning Research*, *17*(1), pp.2797-2819.

[377] Brown, S.A., 2016. Patient similarity: emerging concepts in systems and precision medicine. *Frontiers in physiology*, *7*, p.561.

[378] Liu, B., Li, J., Yang, J., Bi, J., Li, R. and Li, Y., 2016, June. Pattern Recognition for Large-Scale and Incremental Time Series in Healthcare. In *2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC)* (Vol. 2, pp. 653-658). IEEE.

[379] Pires, I.M., Garcia, N.M., Pombo, N., Flórez-Revuelta, F. and Spinsante, S., 2017. Pattern recognition techniques for the identification of Activities of Daily Living using mobile device accelerometer. *arXiv preprint arXiv:1711.00096*.

[380] Yuliastuti, G.E., Alfiyatin, A.N., Rizki, A.M., Hamdianah, A., Taufiq, H. and Mahmudy, W.F., 2018. Performance Analysis of Data Mining Methods for Sexually Transmitted Disease Classification. *International Journal of Electrical and Computer Engineering*, *8*(5), p.3933.

[381] Danjuma, K.J., 2015. Performance evaluation of machine learning algorithms in post-operative life expectancy in the lung cancer patients. *arXiv preprint arXiv:1504.04646*.

[382] Moreira, M.W., Rodrigues, J.J., Oliveira, A.M., Saleem, K. and Neto, A., 2016, December. Performance evaluation of predictive classifiers for pregnancy care. In *2016 IEEE Global Communications Conference (GLOBECOM)* (pp. 1-6). IEEE.

[383] Bal, M., Amasyali, M.F., Sever, H., Kose, G. and Demirhan, A., 2014. Performance evaluation of the machine learning algorithms used in inference mechanism of a medical decision support system. *The Scientific World Journal*, *2014*.

[384] Lee, J., 2016, February. Personalized mortality prediction for the critically ill using a patient similarity metric and bagging. In *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)* (pp. 332-335). IEEE.

[385] El Naqa, I., 2016. Perspectives on making big data analytics work for oncology. *Methods*, *111*, pp.32-44.

[386] Shameer, K., Perez-Rodriguez, M.M., Bachar, R., Li, L., Johnson, A., Johnson, K.W., Glicksberg, B.S., Smith, M.R., Readhead, B., Scarpa, J. and Jebakaran, J., 2018. Pharmacological risk factors associated with hospital readmission rates in a psychiatric cohort identified using prescriptome data mining. *BMC medical informatics and decision making*, *18*(3), p.79.

[387] Khader, N., Lashier, A. and Yoon, S.W., 2016. Pharmacy robotic dispensing and planogram analysis using association rule mining with prescription data. *Expert Systems with Applications*, *57*, pp.296-310.

[388] Baytas, I.M., Lin, K., Wang, F., Jain, A.K. and Zhou, J., 2016. Phenotree: Interactive visual analytics for hierarchical phenotyping from large-scale electronic health records. *IEEE Transactions on Multimedia*, *18*(11), pp.2257-2270.

[389] Wang, M., 2013. Predefined three tier business intelligence architecture in healthcare enterprise. *Journal of medical systems*, *37*(2), p.9928.

[390] Yang, C., Delcher, C., Shenkman, E. and Ranka, S., 2016, September. Predicting 30-day all-cause readmissions from hospital inpatient discharge data. In *2016 IEEE 18th International conference on e-Health networking, applications and services (Healthcom)* (pp. 1-6). IEEE.

[391] Min, H., Mobahi, H., Irvin, K., Avramovic, S. and Wojtusiak, J., 2017. Predicting activities of daily living for cancer patients using an ontology-guided machine learning methodology. *Journal of biomedical semantics*, *8*(1), p.39.

[392] Zhang, B., Ren, H., Huang, G., Cheng, Y. and Hu, C., 2019. Predicting blood pressure from physiological index data using the SVR algorithm. *BMC bioinformatics*, *20*(1), p.109.

[393] Johri, V., Srivastava, I. and Singh, V., 2014. Predicting child health using big data analytics. *International Journal of Advanced Research in Computer Science*, *5*(7).

[394] Hammond, R., Athanasiadou, R., Curado, S., Aphinyanaphongs, Y., Abrams, C., Messito, M.J., Gross, R., Katzow, M., Jay, M., Razavian, N. and Elbel, B., 2019. Predicting childhood obesity using electronic health records and publicly available data. *PloS one*, *14*(4), p.e0215571.

[395] Huddar, V., Desiraju, B.K., Rajan, V., Bhattacharya, S., Roy, S. and Reddy, C.K., 2016. Predicting complications in critical care using heterogeneous clinical data. *IEEE Access*, *4*, pp.7988-8001.

[396] Fiore, M., Gallo, C., Tsoukatos, E. and La Sala, P., 2017. Predicting consumer healthy choices regarding type 1 wheat flour. *British Food Journal*, *119*(11), pp.2388-2405.

[397] Choi, J. and Choi, W.J., 2018. Predicting Depression Among Community Residing Older Adults: A Use of Machine Learning Approch. *Studies in health technology and informatics*, *250*, pp.265-265.

[398] Tsao, H.Y., Chan, P.Y. and Su, E.C.Y., 2018. Predicting diabetic retinopathy and identifying interpretable biomedical features using machine learning algorithms. *BMC bioinformatics*, *19*(9), p.195.

[399] Ndour, C., Gbété, S.D., Bru, N., Abrahamowicz, M., Fauconnier, A., Traoré, M., Diop, A., Fournier, P. and Dumont, A., 2013. Predicting in-hospital maternal mortality in Senegal and Mali. *PloS one*, *8*(5), p.e64157.

[400] Pieszko, K., Hiczkiewicz, J., Budzianowski, P., Budzianowski, J., Rzeźniczak, J., Pieszko, K. and Burchardt, P., 2019. Predicting Long-Term Mortality after Acute Coronary Syndrome Using Machine Learning Techniques and Hematological Markers. *Disease markers*, *2019*.

[401] Huber, M., Kurz, C. and Leidl, R., 2019. Predicting patient-reported outcomes following hip and knee replacement surgery using supervised machine learning. *BMC medical informatics and decision making*, *19*(1), p.3.

[402] Topuz, K., Uner, H., Oztekin, A. and Yildirim, M.B., 2018. Predicting pediatric clinic no-shows: a decision analytic framework using elastic net and Bayesian belief network. *Annals of Operations Research*, *263*(1-2), pp.479-499.

[403] Tsai, J.T., Hou, M.F., Chen, Y.M., Wan, T.T., Kao, H.Y. and Shi, H.Y., 2013. Predicting quality of life after breast cancer surgery using ANN-based models: performance comparison with MR. *Supportive Care in Cancer*, *21*(5), pp.1341-1350.

[404] Wellner, B., Grand, J., Canzone, E., Coarr, M., Brady, P.W., Simmons, J., Kirkendall, E., Dean, N., Kleinman, M. and Sylvester, P., 2017. Predicting unplanned transfers to the intensive care unit: a

machine learning approach leveraging diverse clinical elements. *JMIR medical informatics*, *5*(4), p.e45.

[405] Singla, S.K., Kaur, R. and Bala, A., 2018. Prediction of Depression in Human Behaviour using Deep Learning Neural Network. *International Journal of Health and Economic Development*, *4*(2), pp.16-23.

[407] Ye, C., Fu, T., Hao, S., Zhang, Y., Wang, O., Jin, B., Xia, M., Liu, M., Zhou, X., Wu, Q. and Guo, Y., 2018. Prediction of incident hypertension within the next year: prospective study using statewide electronic health records and machine learning. *Journal of medical Internet research*, *20*(1), p.e22.

[408] Taylor, R.A., Pare, J.R., Venkatesh, A.K., Mowafi, H., Melnick, E.R., Fleischman, W. and Hall, M.K., 2016. Prediction of in-hospital mortality in emergency department patients with sepsis: a local big data–driven, machine learning approach. *Academic emergency medicine*, *23*(3), pp.269-278.

[409] Singh, S.K., Tiwari, S., Abidi, A.I. and Singh, A., 2017. Prediction of pain intensity using multimedia data. *Multimedia Tools and Applications*, *76*(18), pp.19317-19342.

[410] Prasad, V., Guerrisi, M., Dauri, M., Coniglione, F., Tisone, G., De Carolis, E., Cillis, A., Canichella, A., Toschi, N. and Heldt, T., 2017. Prediction of postoperative outcomes using intraoperative hemodynamic monitoring data. *Scientific reports*, *7*(1), p.16376.

[411] VAGHELA, K., 2019. Prediction of protein cellular localization site by using data mining techniques.

[412] Dinov, I.D., Heavner, B., Tang, M., Glusman, G., Chard, K., Darcy, M., Madduri, R., Pa, J., Spino, C., Kesselman, C. and Foster, I., 2016. Predictive big data analytics: a study of Parkinson's disease using large, complex, heterogeneous, incongruent, multi-source and incomplete observations. *PloS one*, *11*(8), p.e0157077.

[413] Kop, R., Hoogendoorn, M., Ten Teije, A., Büchner, F.L., Slottje, P., Moons, L.M. and Numans, M.E., 2016. Predictive modeling of colorectal cancer using a dedicated pre-processing pipeline on routine electronic medical records. *Computers in biology and medicine*, *76*, pp.30-38.

[414] Shameer, K., Johnson, K.W., Yahi, A., Miotto, R., Li, L.I., Ricks, D., Jebakaran, J., KOVATCH, P., Sengupta, P.P., GELIJNS, S. and Moskovitz, A., 2017. Predictive modeling of hospital readmission rates using electronic medical record-wide machine learning: a case-study using Mount Sinai heart failure cohort. In *PACIFIC SYMPOSIUM ON BIOCOMPUTING 2017* (pp. 276-287).

[415] Bosnyak, Z., Zhou, F.L., Jimenez, J. and Berria, R., 2019. Predictive modeling of hypoglycemia risk with basal insulin use in type 2 diabetes: use of machine learning in the LIGHTNING study. *Diabetes Therapy*, *10*(2), pp.605-615.

[416] Kim, K.H., Lee, S., Shim, J.B., Chang, K.H., Cao, Y., Choi, S.W., Jeon, S.H., Yang, D.S., Yoon, W.S., Park, Y.J. and Kim, C.Y., 2017. Predictive modelling analysis for development of a radiotherapy decision support system in prostate cancer: a preliminary study. *Journal of Radiotherapy in Practice*, *16*(2), pp.161-170.

[417] Duggal, R., Shukla, S., Chandra, S., Shukla, B. and Khatri, S.K., 2016. Predictive risk modelling for early hospital readmission of patients with diabetes in India. *International Journal of Diabetes in Developing Countries*, *36*(4), pp.519-528.

[418] Luo, G., 2016. PredicT-ML: a tool for automating machine learning model building with big clinical data. *Health information science and systems*, *4*(1), p.5.

[419] Raghupathi, V. and Raghupathi, W., 2017, March. Preventive healthcare: A neural network analysis of behavioral habits and chronic diseases. In *Healthcare* (Vol. 5, No. 1, p. 8). Multidisciplinary Digital Publishing Institute.

[421] Khan, S.I. and Hoque, A.S.L., 2016, January. Privacy and security problems of national health data warehouse: a convenient solution for developing countries. In *2016 International Conference on Networking Systems and Security (NSysS)* (pp. 1-6). IEEE.

[422] Abuwardih, L.A., Shatnawi, W.E. and Aleroud, A., 2016, July. Privacy preserving data mining on published data in healthcare: A survey. In *2016 7th International Conference on Computer Science and Information Technology (CSIT)* (pp. 1-6). IEEE.

[423] Ostherr, K., 2018. Privacy, Data Mining, and Digital Profiling in Online Patient Narratives. *Catalyst: Feminism, Theory, Technoscience*, *4*(1).

[424] Liu, X., Lu, R., Ma, J., Chen, L. and Qin, B., 2015. Privacy-preserving patient-centric clinical decision support system on naive Bayesian classification. *IEEE journal of biomedical and health informatics*, *20*(2), pp.655-668.

[425] Gong, Y., Fang, Y. and Guo, Y., 2016. Private data analytics on biomedical sensing data via distributed computation. *IEEE/ACM transactions on computational biology and bioinformatics*, *13*(3), pp.431-444.

[426] Boyaci, I., Chin, J., Zeid, A., Kamarthi, S.V., Agboola, S. and Jethwani, K., 2014. Product Platform Approach to Personalized Type 2 Diabetes Mellitus Management. In *IIE Annual Conference. Proceedings* (p. 2822). Institute of Industrial and Systems Engineers (IISE).

[427] Khan, S.S., Barve, R. and Kulkarni, U., 2018, October. Proposed model on Prediction and Analysis using application of Health care. In *2018 3rd International Conference on Communication and Electronics Systems (ICCES)* (pp. 340-345). IEEE.

[428] Wang, Y., Goh, W., Wong, L., Montana, G. and Alzheimer's Disease Neuroimaging Initiative, 2013. Random forests on Hadoop for genome-wide association studies of multivariate neuroimaging phenotypes. *BMC bioinformatics*, *14*(S16), p.S6.

[429] Xiao, C., Ma, T., Dieng, A.B., Blei, D.M. and Wang, F., 2018. Readmission prediction via deep contextual embedding of clinical concepts. *PloS one*, *13*(4), p.e0195024.

[430] Lorgelly, P.K., Doble, B., Knott, R.J. and Cancer 2015 Investigators, 2016. Realising the value of linked data to health economic analyses of cancer care: a case study of cancer 2015. *Pharmacoeconomics*, *34*(2), pp.139-154.

[431] Triantafyllopoulos, D., Korvesis, P., Mporas, I. and Megalooikonomou, V., 2016. Real-time management of multimodal streaming data for monitoring of epileptic patients. *Journal of medical systems*, *40*(3), p.45.

[432] Rathore, M.M., Ahmad, A., Paul, A., Wan, J. and Zhang, D., 2016. Real-time medical emergency response system: exploiting IoT and big data for public health. *Journal of medical systems*, *40*(12), p.283.

[433] Gonzalez, G.H., Tahsin, T., Goodale, B.C., Greene, A.C. and Greene, C.S., 2015. Recent advances and emerging applications in text and data mining for biomedical discovery. *Briefings in bioinformatics*, *17*(1), pp.33-42.

[434] Li, X., Jiang, Y., Chen, M. and Li, F., 2018. Research on iris image encryption based on deep learning. *EURASIP Journal on Image and Video Processing*, *2018*(1), p.126.

[435] Chen, Q., Ai, N., Liao, J., Shao, X., Liu, Y. and Fan, X., 2017. Revealing topics and their evolution in biomedical literature using Bio-DTM: a case study of ginseng. *Chinese medicine*, *12*(1), p.27.

[436] Cheng, Y., Wang, F., Zhang, P. and Hu, J., 2016, June. Risk prediction with electronic health records: A deep learning approach. In *Proceedings of the 2016 SIAM International Conference on Data Mining* (pp. 432-440). Society for Industrial and Applied Mathematics.

[437] Firdaus, A., Anuar, N.B., Ab Razak, M.F., Hashem, I.A.T., Bachok, S. and Sangaiah, A.K., 2018. Root exploit detection and features optimization: mobile device and blockchain based medical data management. *Journal of medical systems*, *42*(6), p.112.

[438] Van Poucke, S., Zhang, Z., Schmitz, M., Vukicevic, M., Vander Laenen, M., Celi, L.A. and De Deyne, C., 2016. Scalable predictive analysis in critically ill patients using a visual open data analysis platform. *PloS one*, *11*(1), p.e0145791.

[439] Feldman, K., Davis, D. and Chawla, N.V., 2015. Scaling and contextualizing personalized healthcare: A case study of disease prediction algorithm integration. *Journal of biomedical informatics*, *57*, pp.377-385.

[440] Thirunarayan, K. and Sheth, A.P., 2013, November. Semantics-empowered approaches to big data processing for physical-cyber-social applications. In *2013 AAAI Fall Symposium Series*.

[441] Atkins, A.S. and Alharbe, N., 2014. Sensor Technologies using ZigBee and RFID within the Cloud of Internet of Things in Healthcare Applications. *Technia*, *6*(2), p.923.

[442] Nienhold, D., Dornberger, R. and Korkut, S., 2016, October. Sensor-Based Tracking and Big Data Processing of Patient Activities in Ambient Assisted Living. In *2016 IEEE International Conference on Healthcare Informatics (ICHI)* (pp. 473-482). IEEE.

[444] Oliveira, T., Silva, A., Satoh, K., Julian, V., Leão, P. and Novais, P., 2018. Survivability Prediction of Colorectal Cancer Patients: A System with Evolving Features for Continuous Improvement. *Sensors*, *18*(9), p.2983.

[445] Lin, K., Xia, F., Wang, W., Tian, D. and Song, J., 2016. System design for big data application in emotion-aware healthcare. *IEEE Access*, *4*, pp.6901-6909.

[446] Han, S., Kim, K., Cha, E., Kim, K. and Shon, H., 2017. System framework for cardiovascular disease prediction based on big data technology. *Symmetry*, *9*(12), p.293.

[447] Albahri, O.S., Albahri, A.S., Mohammed, K.I., Zaidan, A.A., Zaidan, B.B., Hashim, M. and Salman, O.H., 2018. Systematic review of real-time remote health monitoring system in triage and priority-based sensor technology: Taxonomy, open challenges, motivation and recommendations. *Journal of medical systems*, *42*(5), p.80.

[448] Wang, H., Zhang, Q., So, H.Y., Kwok, A. and Wong, Z.S.Y., 2018. Temporal prediction of in-hospital falls using tensor factorisation. *BMJ innovations*, *4*(2), pp.75-83.

[449] Zhong, Z., Xu, T., Wang, F. and Tang, T., 2018. Text Case-Based Reasoning Framework for Fault Diagnosis and Predication by Cloud Computing. *Mathematical Problems in Engineering*, *2018*.

[450] Harpaz, R., Callahan, A., Tamang, S., Low, Y., Odgers, D., Finlayson, S., Jung, K., LePendu, P. and Shah, N.H., 2014. Text mining for adverse drug events: the promise, challenges, and state of the art. *Drug safety*, *37*(10), pp.777-790.

[451] Botsis, T., Nguyen, M.D., Woo, E.J., Markatou, M. and Ball, R., 2011. Text mining for the Vaccine Adverse Event Reporting System: medical text classification using informative feature selection. *Journal of the American Medical Informatics Association*, *18*(5), pp.631-638.

[452] Huang, Y.C., 2013. The application of data mining to explore association rules between metabolic syndrome and lifestyles. *Health Information Management Journal*, *42*(3), pp.29-36.

[453] Connor, J.P., Symons, M., Feeney, G.F.X., Young, R.M. and Wiles, J., 2007. The application of machine learning techniques as an adjunct to clinical decision making in alcohol dependence treatment. *Substance use & misuse*, *42*(14), pp.2193-2206.

[454] Turkington, R., Mulvenna, M., Bond, R., O. Neill, S. and Armour, C., 2018, July. The application of user event log data for mental health and wellbeing analysis. In *Proceedings of the 32nd International BCS Human Computer Interaction Conference 32* (pp. 1-14).

[455] Sonntag, D., Tresp, V., Zillner, S., Cavallaro, A., Hammon, M., Reis, A., Fasching, P.A., Sedlmayr, M., Ganslandt, T., Prokosch, H.U. and Budde, K., 2016. The clinical data intelligence project. *Informatik-Spektrum*, *39*(4), pp.290-300.

[456] Sibbritt, D. and Gibberd, R., 2004. The effective use of a summary table and decision tree methodology to analyze very large healthcare datasets. *Health care management science*, *7*(3), pp.163-171.

[457] Peng, H. and You, M., 2016, August. The Health Care Fraud Detection Using the Pharmacopoeia Spectrum Tree and Neural Network Analytic Contribution Hierarchy Process. In *2016 IEEE Trustcom/BigDataSE/ISPA* (pp. 2006-2011). IEEE.

[458] Kondylakis, H., Claerhout, B., Keyur, M., Koumakis, L., van Leeuwen, J., Marias, K., Perez-Rey, D., De Schepper, K., Tsiknakis, M. and Bucur, A., 2016. The INTEGRATE project: Delivering solutions for efficient multi-centric clinical research and trials. *Journal of biomedical informatics*, *62*, pp.32-47.

[460] Wu, J., Li, H., Cheng, S. and Lin, Z., 2016. The promising future of healthcare services: When big data analytics meets wearable technology. *Information & Management*, *53*(8), pp.1020-1033.

[461] Vuppalapati, C., Ilapakurti, A. and Kedari, S., 2016, March. The role of big data in creating sense ehr, an integrated approach to create next generation mobile sensor and wearable data driven electronic health record (ehr). In *2016 IEEE Second International Conference on Big Data Computing Service and Applications (BigDataService)* (pp. 293-296). IEEE.

[462] Tartarisco, G., Tonacci, A., Minciullo, P.L., Billeci, L., Pioggia, G., Incorvaia, C. and Gangemi, S., 2017. The soft computing-based approach to investigate allergic diseases: a systematic review. *Clinical and Molecular Allergy*, *15*(1), p.10.

[463] Ross, E.G., Shah, N.H., Dalman, R.L., Nead, K.T., Cooke, J.P. and Leeper, N.J., 2016. The use of machine learning for the identification of peripheral artery disease and future mortality risk. *Journal of vascular surgery*, *64*(5), pp.1515-1522.

[464] Coelho, J.R., Carriço, J.A., Knight, D., Martínez, J.L., Morrissey, I., Oggioni, M.R. and Freitas, A.T., 2013. The use of machine learning methodologies to analyse antibiotic and biocide susceptibility in Staphylococcus aureus. *PLoS One*, *8*(2), p.e55582.

[465] Hao, H. and Zhang, K., 2016. The voice of Chinese health consumers: a text mining approach to web-based physician reviews. *Journal of medical Internet research*, *18*(5), p.e108.

[466] Shen, Q. and Yang, R., 2015, December. Thompson-Tau Outlier Detection Method for Detecting Abnormal Data of Listed Pharmaceutical Companies in China. In *2015 8th International Symposium on Computational Intelligence and Design (ISCID)* (Vol. 1, pp. 379-382). IEEE.

[467] Kurbalija, V., Radovanović, M., Ivanović, M., Schmidt, D., von Trzebiatowski, G.L., Burkhard, H.D. and Hinrichs, C., 2014. Time-series analysis in the medical domain: A study of Tacrolimus administration and influence on kidney graft function. *Computers in biology and medicine*, *50*, pp.19-31.

[468] Ibanez-Sanchez, G., Fernandez-Llatas, C., Martinez-Millana, A., Celda, A., Mandingorra, J., Aparici-Tortajada, L., Valero-Ramon, Z., Munoz-Gama, J., Sepúlveda, M., Rojas, E. and Gálvez, V., 2019. Toward Value-Based Healthcare through Interactive Process Mining in Emergency Rooms: The Stroke Case. *International journal of environmental research and public health*, *16*(10), p.1783.

[469] Chen, H., Shen, J., Wang, L. and Song, J., 2016, June. Towards data analytics of pathogen-host protein-protein interaction: a survey. In *2016 IEEE International Congress on Big Data (BigData Congress)* (pp. 377-388). IEEE.

[470] Shi, X., Li, W., Song, J., Hossain, M.S., Rahman, S.M.M. and Alelaiwi, A., 2016. Towards interactive medical content delivery between simulated body sensor networks and practical data center. *Journal of medical systems*, *40*(10), p.214.

[471] Qi, J., Chen, L., Leister, W. and Yang, S., 2015, August. Towards knowledge driven decision support for personalized home-based self-management of chronic diseases. In *2015 IEEE 12Th Intl Conf on Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and its Associated Workshops (UIC-ATC-ScalCom)* (pp. 1724-1729). IEEE.

[472] Soroushmehr, S.R. and Najarian, K., 2016. Transforming big data into computational models for personalized medicine and health care. *Dialogues in clinical neuroscience*, *18*(3), p.339.

[473] Xu, J., Yang, P., Xue, S., Sharma, B., Sanchez-Martin, M., Wang, F., Beaty, K.A., Dehan, E. and Parikh, B., 2019. Translating cancer genomics into precision medicine with artificial intelligence: applications, challenges and future perspectives. *Human genetics*, *138*(2), pp.109-124.

[474] Silva, A.P.P.D., Oliveira, T.J.M., Neves, J. and Novais, P., 2016. Treating colon cancer survivability prediction as a classification problem. *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal*, *5*(1), pp.37-50.

[475] Rusanov, A., Miotto, R. and Weng, C., 2018. Trends in anesthesiology research: a machine learning approach to theme discovery and summarization. *JAMIA open*, *1*(2), pp.283-293.

[478] Sariyer, G., Taşar, C.Ö. and Cepe, G.E., 2019. Use of data mining techniques to classify length of stay of emergency department patients. *Bio-Algorithms and Med-Systems*, *15*(1).

[479] Park, E., Chang, H.J. and Nam, H.S., 2017. Use of machine learning classifiers and sensor data to detect neurological deficit in stroke patients. *Journal of medical Internet research*, *19*(4), p.e120.

[480] Tighe, P., Laduzenski, S., Edwards, D., Ellis, N., Boezaart, A.P. and Aygtug, H., 2011. Use of machine learning theory to predict the need for femoral nerve block following ACL repair. *Pain Medicine*, *12*(10), pp.1566-1575.

[481] Sanchez-Morillo, D., Fernandez-Granero, M.A. and Leon-Jimenez, A., 2016. Use of predictive algorithms in-home monitoring of chronic obstructive pulmonary disease and asthma: a systematic review. *Chronic respiratory disease*, *13*(3), pp.264-283.

[482] McTaggart, S., Nangle, C., Caldwell, J., Alvarez-Madrazo, S., Colhoun, H. and Bennie, M., 2018. Use of text-mining methods to improve efficiency in the calculation of drug exposure to support pharmacoepidemiology studies. *International journal of epidemiology*, *47*(2), pp.617-624.

[483] Pires, I.M., Garcia, N.M., Pombo, N. and Flórez-Revuelta, F., 2017. User Environment Detection with Acoustic Sensors Embedded on Mobile Devices for the Recognition of Activities of Daily Living. *arXiv preprint arXiv:1711.00124*.

[484] Sunmoo, Y.O.O.N. and Basirah, T.A.H.A., 2014. Using a data mining approach to discover behavior correlates of chronic disease: a case study of depression. *Studies in health technology and informatics*, *201*, p.71.

[485] Merali, Z.G., Witiw, C.D., Badhiwala, J.H., Wilson, J.R. and Fehlings, M.G., 2019. Using a machine learning approach to predict outcome after surgery for degenerative cervical myelopathy. *PloS one*, *14*(4), p.e0215133.

[487] Zhao, Y., Fesharaki, N.J., Liu, H. and Luo, J., 2018. Using data-driven sublanguage pattern mining to induce knowledge models: application in medical image reports knowledge representation. *BMC medical informatics and decision making*, *18*(1), p.61.

[488] Hamidi, H. and Fazeli, K., 2018. Using Internet of Things and biosensors technology for health applications. *IET Wireless Sensor Systems*, *8*(6), pp.260-267.

[489] Asri, H., Mousannif, H., Al Moatassime, H. and Noel, T., 2016. Using machine learning algorithms for breast cancer risk prediction and diagnosis. *Procedia Computer Science*, *83*, pp.1064-1069.

[490] Sakr, S., Elshawi, R., Ahmed, A., Qureshi, W.T., Brawner, C., Keteyian, S., Blaha, M.J. and Al-Mallah, M.H., 2018. Using machine learning on cardiorespiratory fitness data for predicting hypertension: The Henry Ford ExercIse Testing (FIT) Project. *PloS one*, *13*(4), p.e0195344.

[491] Fritz, B.A., Chen, Y., Murray-Torres, T.M., Gregory, S., Abdallah, A.B., Kronzer, A., McKinnon, S.L., Budelier, T., Helsten, D.L., Wildes, T.S. and Sharma, A., 2018. Using machine learning techniques to develop forecasting algorithms for postoperative complications: protocol for a retrospective study. *BMJ open*, *8*(4), p.e020124.

[493] Yala, A., Barzilay, R., Salama, L., Griffin, M., Sollender, G., Bardia, A., Lehman, C., Buckley, J.M., Coopey, S.B., Polubriaginof, F. and Garber, J.E., 2017. Using machine learning to parse breast pathology reports. *Breast cancer research and treatment*, *161*(2), pp.203-211.

[494] Chao-Wen, C., Yuh-Wen, C., Moussa, L. and Tzung-Hung, L., 2015. Using multi-objective affinity model for mining the rules of revisits within 72 hours for emergency department patients. *Multiple Criteria Decision Making*, *10*, pp.5-31.

[495] Balyan, R., Crossley, S.A., Brown III, W., Karter, A.J., McNamara, D.S., Liu, J.Y., Lyles, C.R. and Schillinger, D., 2019. Using natural language processing and machine learning to classify health literacy from secure messages: The ECLIPPSE study. *PloS one*, *14*(2), p.e0212488.

[496] Hernandez, I. and Zhang, Y., 2017. Using predictive analytics and big data to optimize pharmaceutical outcomes. *American Journal of Health-System Pharmacy*, *74*(18), pp.1494-1500.

[497] Jain, D. and Singh, V., 2016. Utilization of Data Mining Classification Approach for Disease Prediction: A Survey. *International Journal Education and Management Engineering*, *6*, pp.45-52.

[498] Stopa, B.M., Yan, S.C., Dasenbrock, H.H., Kim, D.H. and Gormley, W.B., 2019. Variance Reduction in Neurosurgical Practice: The Case for Analytics-Driven Decision Support in the Era of Big Data. *World neurosurgery*, *126*, pp.e190-e195.

[499] Hund, M., Böhm, D., Sturm, W., Sedlmair, M., Schreck, T., Ullrich, T., Keim, D.A., Majnaric, L. and Holzinger, A., 2016. Visual analytics for concept exploration in subspaces of patient groups. *Brain informatics*, *3*(4), p.233.

[500] Vaitsis, C., Nilsson, G. and Zary, N., 2014. Visual analytics in healthcare education: exploring novel ways to analyze and represent big data in undergraduate medical education. *PeerJ*, *2*, p.e683.

[501] Price, J., 2016. What can big data offer the pharmacovigilance of orphan drugs?. *Clinical therapeutics*, *38*(12), pp.2533-2545.

[502] Guo, L., Jin, B., Yao, C., Yang, H., Huang, D. and Wang, F., 2016. Which doctor to trust: a recommender system for identifying the right doctors. *Journal of medical Internet research*, *18*(7), p.e186.

[503] Sharma, D. and Shadabi, F., 2014, December. The potential use of multi-agent and hybrid data mining approaches in social informatics for improving e-Health services. In *2014 IEEE Fourth International Conference on Big Data and Cloud Computing* (pp. 350-354). IEEE.