

Joanna KEDRA ^{1,2}, Timothy RADSTAKE ³, Aridaman PANDIT ³, Xenofon BARALIAKOS ⁴, Francis BERENBAUM ⁵, Axel FINCKH ⁶, Bruno FAUTREL ^{1,2}, Tanja STAMM ⁷, David GOMEZ-CABRERO ⁸, Christian PRISTIPINO ⁹, Rémy CHOQUET ¹⁰, Hervé SERVY ¹¹, Simon STONES ^{12,13}, Gerd BURMESTER ¹⁴, Laure GOSSEC ^{1,2}

1. Sorbonne Université, Institut Pierre Louis d'Epidémiologie et de Santé Publique, INSERM UMR S1136, Paris, France. 2. AP-HP, Pitié Salpêtrière Hospital, Department of rheumatology, Paris, France. 3. Dept. of Rheumatology, Clinical Immunology and Laboratory for Translational Immunology, University Medical Center Utrecht, The Netherlands. 4. Rheumazentrum Ruhrgebiet Herne, Ruhr-University Bochum, Germany. 5. AP-HP, Saint-Antoine Hospital, Department of rheumatology, Paris, France. 6. Division of Rheumatology, Geneva University Hospital, Switzerland. 7. Section for Outcomes Research, Center for Medical Statistics, Informatics, and Intelligent Systems, Medical University of Vienna, Austria. 8. Translational Bioinformatics Unit, Navarra Biomed, Departamento de Salud-Universidad Pública de Navarra, Pamplona, Navarra, Spain. 9. Ospedale San Filippo Neri, Rome, Italy. 10. INSERM U1142, Orange Health, France. 11. Sanoia e-health services, France. 12. School of Healthcare, University of Leeds, Leeds, United Kingdom. 13. Young PARE, Zurich, Switzerland. 14. Department of Rheumatology and Clinical Immunology, Charité – University Medicine Berlin, Germany.

BACKGROUND

- Big data are defined as data sets that are too large or complex for traditional data-processing application software to adequately deal with. ⁽¹⁾
- Artificial Intelligence (AI) includes various statistical techniques which can deal with big data. ⁽²⁾
- The current use of these concepts in publications related to rheumatic and musculoskeletal diseases (RMDs) is unknown.

OBJECTIVE

To assess the current use of big data and AI in the field of RMDs.

METHODS

- A systematic literature review (SLR) was performed in PubMed MEDLINE in November 2018, with key words referring to big data, artificial intelligence and RMDs.
- All original reports published in English were analyzed.
- A mirror literature review was also performed outside of RMDs on the same number of articles, in a retrochronological way.
- The number of data analyzed, data sources and statistical methods used (traditional statistics, AI or both) were collected. The analysis compared findings within and beyond the field of RMDs.

Research algorithm in RMDs

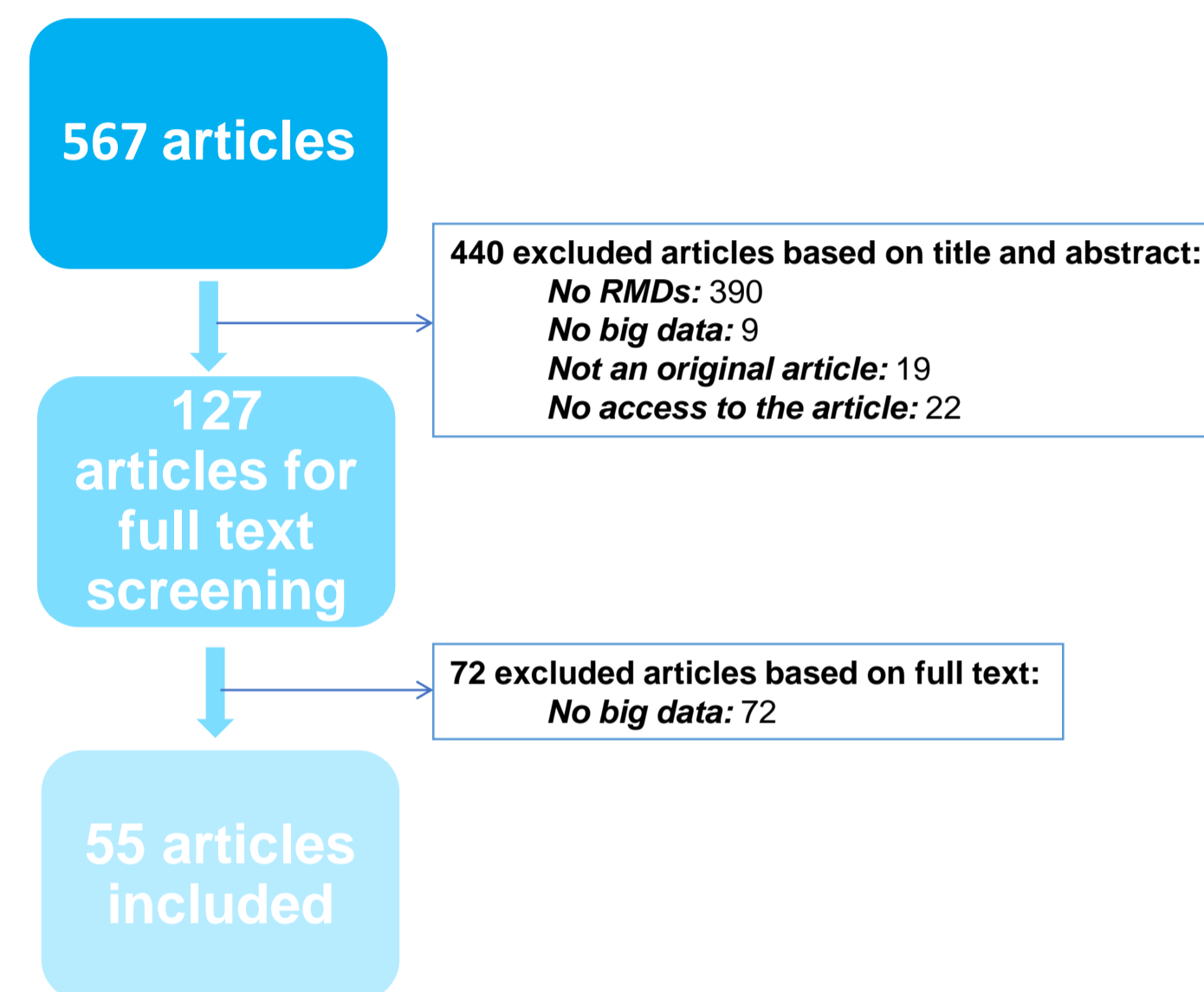
Research algorithm out of RMDs

- Line 1:** ("big data" [All Fields] OR "Artificial Intelligence"[MeSH Terms]) AND
- Line 2:** ("musculoskeletal diseases" [MeSH Terms] OR "musculoskeletal diseases" [All Fields] OR "rheumatology"[MeSH Terms] OR "rheumatology"[All Fields])

- Line 1:** ("big data" [All Fields] OR "Artificial Intelligence"[MeSH Terms]) NOT
- Line 2:** ("musculoskeletal diseases" [MeSH Terms] OR "musculoskeletal diseases" [All Fields] OR "rheumatology"[MeSH Terms] OR "rheumatology"[All Fields])

RESULTS

Figure 1. Flow-chart of the SLR in RMDs



Overall description:

- Flow-chart:** of 567 articles relating to RMDs, 55 met the inclusion criteria and were analyzed, as well as 55 articles in other medical fields. (Figure 1)
- Mean year of publication:** 2014 for the RMDs SLR, 2018 or 2019 in the mirror non-RMD review. (Table 1)
- Most represented fields in RMDs:** inflammatory joint diseases (N=22, 40%) and osteoarthritis (N=16, 29%).
- Most represented medical fields outside of RMDs:** oncology (N=14, 25%), neurology (N=8, 15%), infectious diseases (N=6, 11%), ophthalmology (N=5, 9%) and psychiatry (N=5, 9%). (Figure 2)

Table 1. Description of 55 articles on big data in RMDs, and 55 articles for comparison outside RMDs

	RMDs	Other medical fields
Year of publication, mean (SD) [range]	2014 (4.6) [1992-2018]	2018 (0.4) [2018-2019]
Year of publication: last 5 years, %	72	100
Impact factor, mean (SD) [range]	3.8 (4.0) [0.35-23.3]	5.56 (9.8) [0.56-47.7]
Geographic origin of the first author, %		
North America	38	31
Europe	33	33
Asia	27	34
Oceania, South America and Africa	2	2
Clear definition of big data, %	4	13
Units of observation, mean (SD) [range]	1,142,000 (3,990,000) [5 – 25,000,000]	5,298,000 (23,909,000) [40 – 140,000,000]
Data points, mean (SD) [range]	746,000,000 (1,660,000,000) [2000 – 5,000,000,000]	9,149,000,000 (39,000,000,000) [100,000 – 200,000,000,000]

Big data definition (Table 1):

- Only two articles in the field of RMDs (4%) and seven articles out of the field of RMDs (13%) mentioned a clear definition of big data.
- The mean number of data points was 746 million [range 2000–5 billion] in RMDs, and 9.1 billion [range 100,000 – 200 billion] outside of RMDs.

Big data sources (Table 2):

- Data sources were varied, both in RMDs and in other medical fields.

Table 2. Big data sources in RMDs and outside RMDs

	RMDs	Other medical fields
Clinical data sources, %	47	31
Registries/cohorts	25	18
EHR	20	6
Claims databases	2	0
PGHD (sensors, ...)	0	7
Biological data (-omics), %	15	31
Imaging, %	29	29
Text-mining from publications, %	9	9

Big data analysis (Figure 3):

- Both traditional and AI methods were used to analyze big data (respectively 10 (18%) and 45 (82%) in RMDs and 8 (15%) and 47 (85%) out of RMDs).
- Machine learning was used in almost all AI papers (44/45 in RMDs and 47/47 outside of RMDs), and among machine learning methods, the most represented was artificial neural network (20/44 in RMDs, and 24/47 out of RMDs).

Figure 2. Description of the diseases

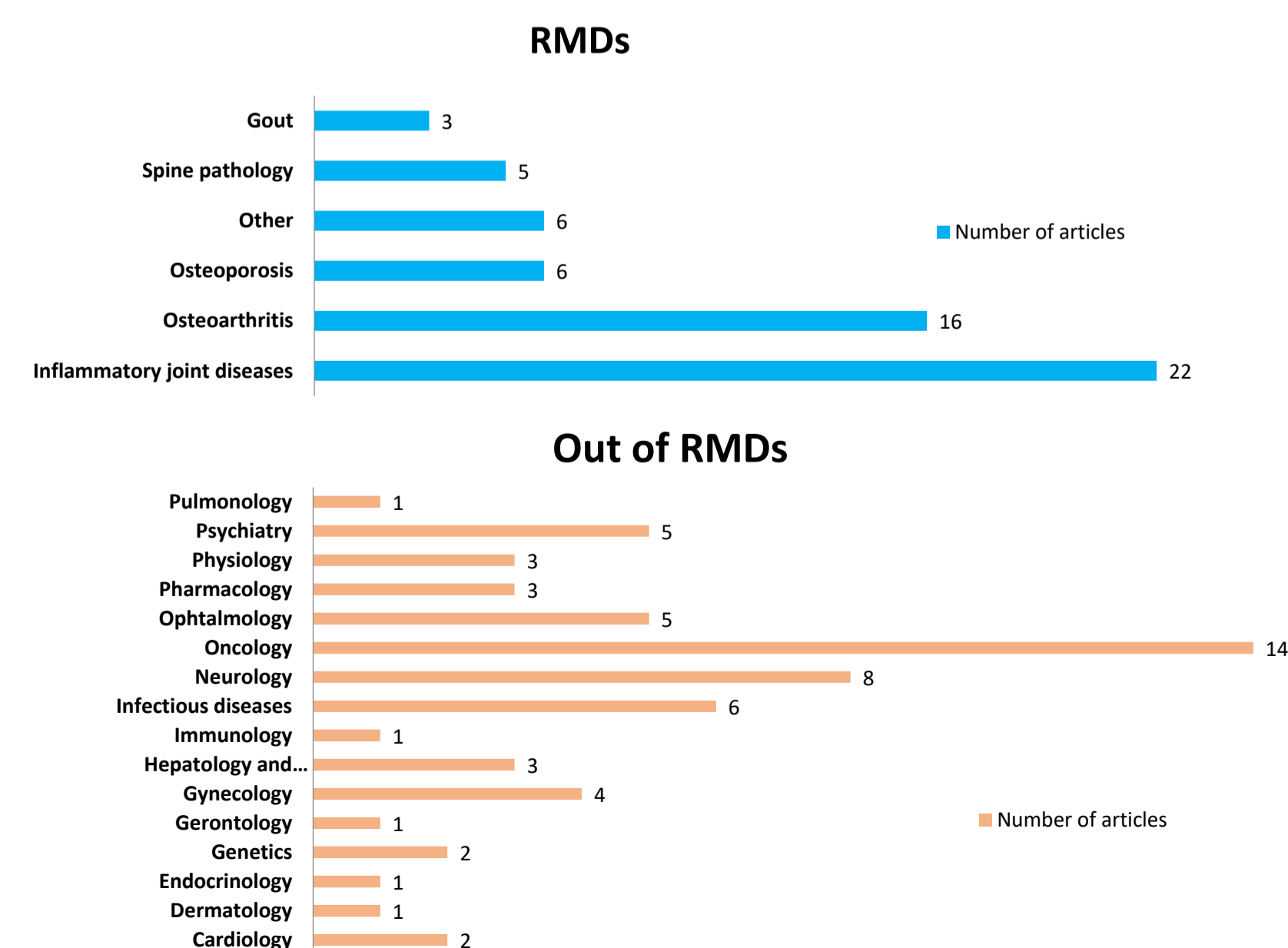
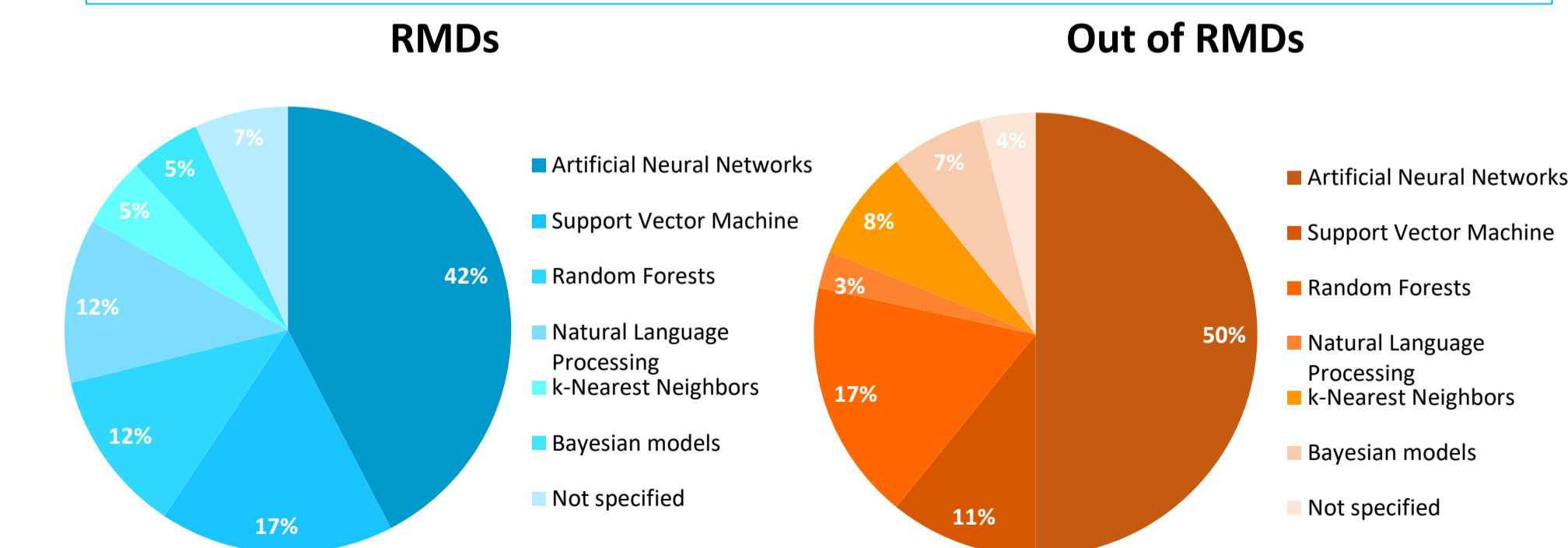


Figure 3. Distribution of machine learning methods



REFERENCES

(1) Cox M, Ellsworth D. Managing big data for scientific visualization. ACM SIGGRAPH '97 course #4, exploring gigabyte datasets in real-time: algorithms, data management, and time-critical design, ACM Digital Library, Anaheim, CA, US, Los Angeles, 1997:5-17. (2) Obermeyer Z, Emanuel EJ. Predicting the Future – Big Data, Machine Learning, and Clinical Medicine. *N Engl J Med*. 2016 Sep 29;375:1216-9. (3) Gossec L, Kedra J, Servy H, et al. European League Against Rheumatism points to consider for the use of big data in rheumatic and musculoskeletal diseases. *Annals of the Rheumatic Diseases* Published Online First: 22 June 2019. doi: 10.1136/annrheumdis-2019-21569 (4) Kedra J, Radstake T, Pandit A, et al. Current status of use of big data and artificial intelligence in RMDs: a systematic literature review informing EULAR recommendations. *RMD Open*. 2019; 5(2): e001004.

DISCLOSURES OF INTEREST

L Gossec has published a study for which Orange IMT (telecommunications company) performed machine learning analyses, without charge to the author. R Choquet and H Servy are employees of Clinical Research Organizations providing clinical research services including data science. This study was supported by EULAR.