

CA19110



European Cooperation in
Science and Technology

Virtual Mobility grant

**Bibliometric evaluation – Is modelling plasma for agriculture
keeping pace with Machine Learning trends**

Grantee name: Marija Puač

07/08/2024 to 31/08/2024

Introduction

Machine Learning (ML), as well as Deep Learning (DL) and Artificial Intelligence (AI) seem as hot topic of 21st century but the birth of ML and Neural Networks originates eight decades ago when neuroscientists Walter Pitts and Warren McCulloch presented their mathematical model of neural network in 1943. Through the decades after the 1940s, the evolution of ML included many notable milestones. The terms *Artificial Intelligence* and *Machine Learning* were introduced in 1950s when Turing, Samuel, McCarthy, Minsk, Edmonds and Newell presented the first artificial neural network and the Turing Test. And today, release of GPT in 2018 by the OpenAI is now affecting all of our lives.

Giant steps that AI took can be only possible with scientific society involvement, apart from the huge amounts of resources. And it is not surprising that many of areas of research are now oriented towards models based on ML algorithms. So, is there a place for plasma science, particularly for plasma application for agriculture and can it lead towards sustainable food production more quickly than without ML involvement?

One of the possible solutions for aforementioned problem of food shortage can be found in low-temperature plasmas that have versatile use in food production, from agriculture through food processing industry. In agriculture the seed treatments are used to improve yields, increase plants' size and robustness and reduction (or elimination) the need for antifungal agents. All of this makes low temperature plasma an ideal green agenda instrument which can significantly reduce chemical usage with improved quality and safer food.

Number of scientific papers that deal with plasma application for agriculture is exponentially growing in recent decades due to its importance for industry as well as for scientific community. That makes a good resource of data necessary for development of viable models that can change (speed up) path towards the industrial application of plasma in agriculture.

This report is written as a bibliometric analysis of the scientific publications that include ML (in addition with AI, Neural Networks (NN) and DL) to model plasma experiments, such as seeds and plants treatments (including fungi that can harm them) and plasma activated water (PAW), whether they use atmospheric or low pressure plasmas. Since the area of research is relatively new, short

review of conference papers and organized workshops or seminars that are available online will be also presented. Further attention will be given to the researchers who are pioneers of ML application to the plasma models.

Selection of database and methodology

Currently the most popular databases, with highest number of entries, are Google Scholar, Microsoft Academics, Crossref, Web of Science, PubMed, ScienceDirect, Dimensions, Scopus and ScienceOpen. There are also scientific paper repositories but they are usually focused on publications that are dealing with more specific field of research, like biology or economy.

Database was chosen to fulfill following requirements:

- Free to use – availability for everyone enables subsequent addition of data in the future; author did not have access to repositories that are maintained by subscriptions
- Up to date – necessary for bibliometric analysis to be viable is for database to be updated in real time
- Multidisciplinary – plasma treatments very often include scientist with different areas of expertise
- Rich metadata – enables high quality search in order to include and exclude everything that is necessary to obtain meaningful data at the end
- Possibility of data export – locally stored data makes data processing, creating of tables and graphs, much easier.

After testing aforementioned databases conclusion is that Google Scholar is the best base for author search, giving the comprehensive analysis of author's metrics supported by the graphical presentations. As for search that can include and exclude numerous keywords the best choice is Dimensions that had more than 147 million records. All further analysis was done by using this two databases.

Agriculture plasma and Machine Learning (AI, DL and NNs)

Requesting the data has to be done through a query. Dimensions support the search of 3 categories: Full data, Title and abstract, DOI. Search was done in Title and abstract category. The most important part of this analysis is to create proper query which will include the highest number of relevant entries and equally important is to exclude the most entries that are not important for this research, the most obvious example is to include word “plasma”, describing ionized gas consisting of positive ions and free electrons, but NOT include “plasma” that is colorless fluid part of blood. It is important to note that perfect query does not exist and that certain number of false positive and false negative will always appear. After some trial and error testing of what the query should be it was concluded that results of the query should fulfill following conditions:

- Main focus have two parts:
 - o First is on Machine Learning (ML), Artificial Intelligence (AI), Deep Learning (DL) and Neural Networks (NN).
 - o Second is on plasma treatments, not chemicals or laser treatments.
- Then to include seeds and plasma activated water (PAW) and soil remediation, not treatments of material surfaces (textile, etc.).
- To include all plasma sources that are being used for treatments.

Seeds and plants that are being used in researches, and included in query, are following (taken from *Shelar et al. RSC Adv.*, 2022, **12**, 10467):

Broccoli OR cotton OR pepper OR lentil OR soybean OR barley OR “Norway spruce” OR “chili pepper” OR wheat OR cucumber OR tomato OR Arabidopsis OR radish OR sunflower OR “Thuringian mallow” OR pea OR zucchini OR maze OR eggplant OR pumpkin OR grape OR artichoke OR Asparagus OR rapeseed OR mustard OR basil OR “Melisa officinalis” OR “mung bean” OR lettuce OR Dianthus OR “sticky bean” OR pine OR Moringa OR lupine OR “Catharanthus roseus” OR hyssop OR “black gram” OR rice OR hemp OR coriander OR clover OR ajwain OR peanut OR mulungu OR mimosa OR quinoa OR coneflower.

Plasma sources that are being used in researches, and included in query, are following:

“Corona discharge” OR atmospheric OR “low pressure” OR dbd OR “RF plasma” OR “RF inductive” OR “RF capacitive” OR ccp OR “ccp RF” OR “plasma jet” OR “plasma flashlight” OR “arc discharge” OR “gliding arc” OR “ccp glow” OR iccp OR “iccp RF” OR “needle to plain” OR “microwave plasma” OR “microcorona”.

And to EXCLUDE every kind of plasma induced spectrometry (ICP-MS, ICP-AES etc.).

Result of query obtained from Dimensions included 50 publications, 3 datasets 9 grants and 10 patents. Due to small number of publications further analysis was done by hand.

First ever paper (*Dettmar et al., Computer Chem. Vol. 20, No 2, pp 261, 1996*) that included artificial neural network (ANN) for the analysis was published in 1996, and scientists were using ANN called Dystal to classify orange juice products each characterized by HPLC and inductively coupled plasma spectroscopy measurements. As this research do not check all the requirements of this bibliometric analysis, it is excluded from the analysis but mentioned here as a reminder when did the scientists related to the plasma and food started to think about advance algorithms application to their area of research.

The main problem with obtained results were inclusion of publications dealing with plasma induced spectrometry (while it was excluded in the query but still appearing as a result). And those that are investigating plasma treatments of cotton and wool fabrics, also unacceptable for the analysis.

After data cleaning, only 8 publications (2 conferences and 6 papers) left and 1 book chapter:

1. Kumar *et al.* “Effect of cold plasma processing on physicochemical and nutritional quality attributes of kiwifruit juice” *Journal of Food Science* **28** 4 1533 (2023)
2. Chen *et al.* “A specific fine-grained identification model for plasma-treated rice growth using multiscale shortcut convolutional neural network” *Mathematical Bioscience and Engineering* **20** 6 10223 (2023)
3. Shelar *et al.* “Emerging cold plasma treatment and machine learning prospects for seed priming: a step towards sustainable food production” *RSC Advances* **12** 10467 (2022)
4. Ozdemir *et al.* “Machine Learning to Predict Oxidative Strength of Cold Atmospheric Plasma Activated Water via Paper-Based Sensor” *Medical Technologies Congress (TIPTEKNO), Turkey* (2022)

DOI: 10.1109/TIPTEKNO56568.2022.9960229

5. Rashvand *et al.* “Prediction of CO₂ and ethylene produced in-packaged apricot under cold plasma treatment by machine learning approach” *Journal of Food Process Engineering* **46** 9 e14418 (2023)
6. Wang *et al.* “Plasma rice yield prediction based on Bi-LSTM model” *Second International Conference on Electronic Information Engineering, Big Data and Computer Science (EIBDCT 2023)I*
DOI: <https://doi.org/10.1117/12.2674801>
7. Neiduzak *et al.* “Digitally manufactured air plasma-on-water reactor for nitrate production” *Plasma Sources Science and Technologies* **31** 3 035016 (2022)
8. Chutia *et al.* “Influence of cold plasma voltage and time on quality attributes of tender coconut water (*Cocos nucifera* L.) and degradation kinetics of its blended beverage” *Journal of Food Processing and Preservation* **45** 4 e15372 (2021)

Book chapter:

1. Donde, R. *et al.* (2021) “Artificial Intelligence and Machine Learning in Rice Research”
In: Gupta, M.K., Behera, L. (eds) *Applications of Bioinformatics in Rice Research*.
Springer, Singapore. https://doi.org/10.1007/978-981-16-3997-5_12

All of these publications were published in last 3 years, while 8 of them are directly connected with the food processing and only 1 is published in journal dedicated to plasma science, journal *Plasma Sources Science and Technology*.

Because of small number of results and in order to understand how much are new computational algorithms represented in areas close to this bibliometric analysis, similar analysis were done for two research areas: first, agriculture and new algorithms, and second, plasma medicine and new algorithms.

Agriculture and Machine Learning (AI, DL and NNs)

Query is formed in following way:

“Machine Learning” OR “Artificial Intelligence” OR “Deep Learning” OR “Neural Networks”

AND

Agriculture

Overall numbers are: 18,111 publications and more than 203,000 citations. Detailed search was done for the last decade from 2014 to 2024, number of publications is 17,670 with 200,204 citations and with citation mean by publication of 11.21. Due to number of publications, analysis was done by the Dimensions base, reviewed by author and presented here.

Distribution of published documents for every year in 2014-2024 range is shown in Figure 1. The year of 2014 was chosen as point in time when interest for this area was very low and can be seen as beginning of rapid increase in number of publications in years that followed. Total number of publications is 17,670 which is 97.565% of publications of all time. Slope of presented curve is exponential and there is a research maximum of 4,912 publications in 2023. The results in 2024 are still unknown which means that exponential increase is possible in 2024 as well.

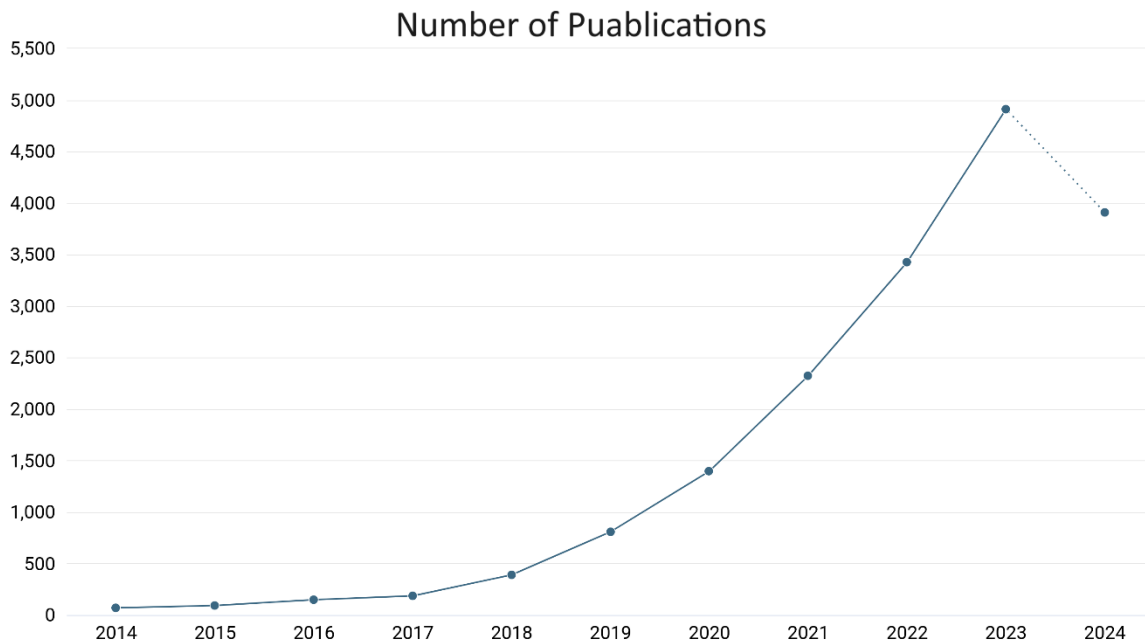


Figure 1. Number of documents published per year during 2014-2024 period.

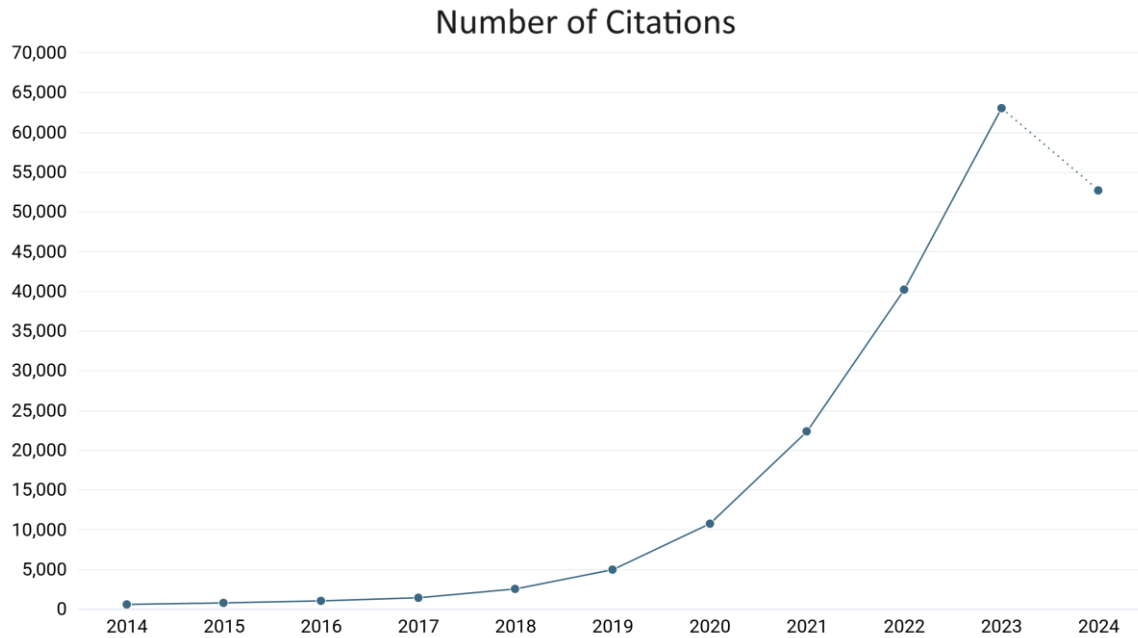


Figure 2. Total citations of documents per year during 2014-2024 period.

Number of citations per year is presented in Figure 2. Total number is 200,204 citations for selected 10 years period. The mean of number of citations per publication is 11.33.

Having in mind multidisciplinary nature of this field distribution of documents by subject area is presented in Figure 3. Obviously, each document can be categorized in more than one subject area. The majority of them (9,904) belong to the Informational and Computing Science, which is expected due to new computational algorithm that are being used. Second most used category is Agricultural, Veterinary and Food Science with 4,403 publications, than Earth Science (3,390), Engineering (2,229), Biological Science (1,283) and Environmental Science (1,252) and so on.

Number of publications in each research category

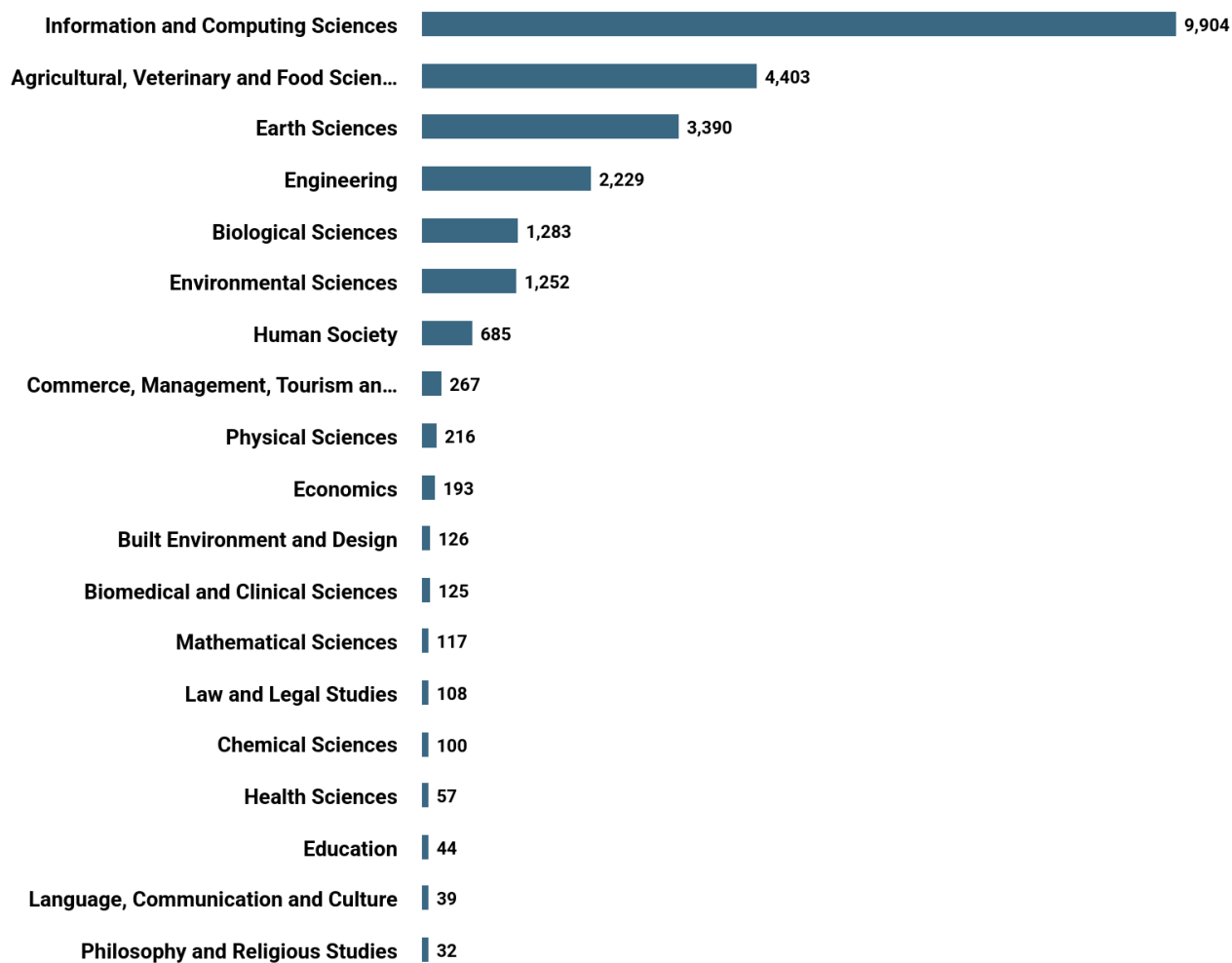


Figure 3. Distribution of documents by subject area.

Table 1. The most prolific authors, their organization, country and number of papers with citation number.

Researcher name	Organization	Country	Number of papers	Citations	Citations Mean
Vinay Kukreja	Chitkara University	India	124	653	5.24
Rishabh Sharma			49	137	2.80
Shiva Mehta			44	316	7.18
Arshleen Kaura			23	0	-
Ravinesh Chand Deo	University of Southern Queensland	Australia	19	1729	91

Zaher Mudher Yaseen	King Fahd University of Petroleum and Minerals	Saudi Arabia	19	701	36.89
Adel Hafiane	University of Orleans	France	18	979	54.39
Raphael Canals			15	974	64.93
Muhammed Attique Khan	HITEC University	Pakistan	17	393	23.12
Satvik Vats	Graphic Era Hill University	India	16	86	5.38
Saraju Prasad Mohanty	University of North Texas	USA	14	213	15.21
Ahmed Elbeltagi	Mansoura University	Egypt	14	272	19.43

Table 1 presents the most prolific authors in the field, their organizations, country and number of publications with number of citations. The most connected researcher is Vinay Kukreja from Chitkara University in India. This author has the wider net of co-authorship (Figure 4a) with 12 co-authorship links and 194 co-authorships within 124 publications (Figure 4b). Countries that are important in this field are: Australia, Saudi Arabia, France, Pakistan, USA and Egypt.

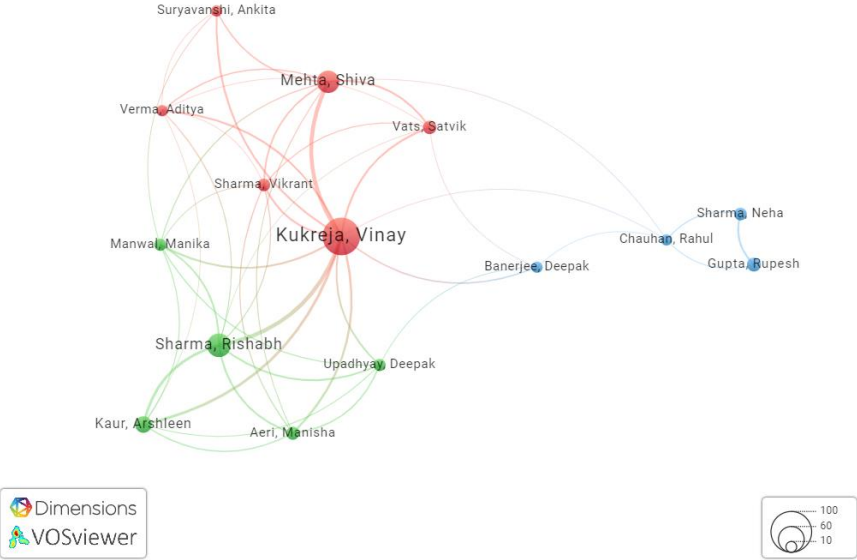


Figure 4 a). Analysis of co-authorship showing 15 researchers with highest number of co-authorship. Figure taken from Dimensions.

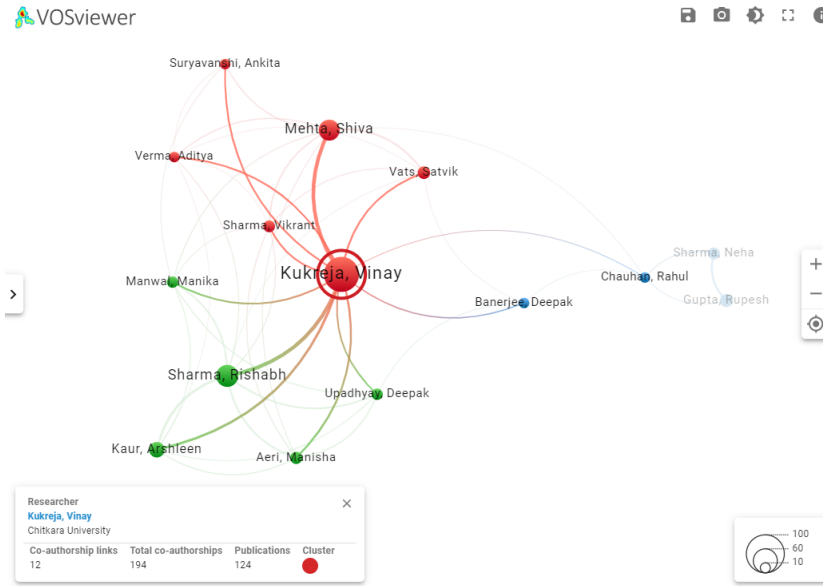


Figure 4 b). Analysis of co-authorship for biggest cluster Vinay Kukreja. Figure taken from Dimensions.

There is a difference between most prolific and most influential authors. Influence of an author can be measured through number of citations, h-index and i10-index. Publications with more than 1,000 citations are presented in Table 2a. The “oldest” paper is from 2017 and the one with the most citations of 3547 (at a time when this report was written) is from the 2018.

Table 2 a). Publications with more than 1000 citations.

Publication	Authors	Year	Number of citations
Deep learning in agriculture: A survey	Andreas Kamilaris, Francesc X. Prenafeta-Boldu	2018	3547
Machine Learning: Algorithms, Real-World Applications and Research Directions	Iqbal H. Sarker	2021	3368
Machine Learning in Agriculture: A Review	Konstatinos G. Liakos, Patrizia Busato, Dimitrios Moshou, Simon Pearson, Dionysis Bochtis	2018	2582

A Robust Deep-Learning-Based Detector for Real-Time Tomato Plant Diseases and Pests Recognition	Alvaro Fuentes, Sook Yoon, Sang Cheol Kim, Dong Sun Parka	2017	1427
Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review	Anna Chlingaryan, Salah Sukkarieh, Brett Whelan	2018	1221
Soil carbon debt of 12,000 years of human land use	Jonathan Sanderman, Tomislav Hengal Gregory J. Fiske	2017	1110
Review on Convolutional Neural Networks (CNN) in vegetation remote sensing	Teja Kattenborn, Jens Leitloff, Felix Schiefer, Stefan Hinz	2021	1080

Authors of the most cited papers that have the most citations (search was done in Google Scholar) with their h-index and i10-index, are presented in Table 2b.

Table 2 b). The most influential researchers.

Researcher	Organization	Number of citations	h-index, i10-index
Iqbal H Sarker	University of Technology Sydney, Australia	11703	40, 80
Dionysis Bochtis	Aarhus University, Denmark	11083	51, 126
Francesc X. Prenafeta-Boldu	Institute of Agrifood Research and Technology Barcelona, Spain	10018	32, 58
Andreas Kamalaris	University of Twente, Netherlands	8639	27, 55

There are some discrepancies between the list of countries obtained in this analysis and those that can be found as most advanced in AI implementation in agriculture (USA, China and Israel). To understand this it is needed more thorough investigation which is not in the scope of this bibliometric analysis, but author thinks that discrepancy is worth mentioning.

Plasma Medicine and Machine Learning (AI, DL and NNs)

Query was formed in the following way:

"Machine learning" OR "neural network" OR "artificial intelligence" OR "deep learning"

AND

"Plasma medicine" OR "plasma treatment" OR "plasma treated"

NOT seed NOT seeds NOT fruit (to exclude agriculture).

Result of query obtained from Dimensions included 78 publications, 2 datasets 10 grants and 5 patents. Due to small number of publications further analysis was done by hand.

After cleaning the data, 24 publications left with 130 citations in total which leads to 5.42 citations per publication.

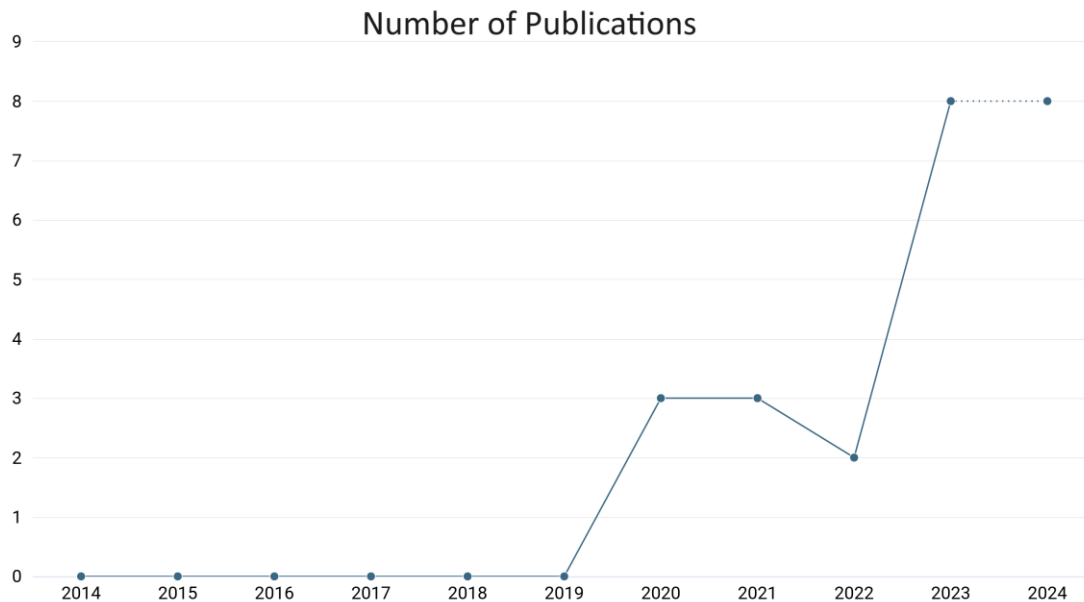


Figure 5. Number of documents published per year during 2014-2024 period.

Distribution of published documents for every year in 2014-2024 range is shown in Figure 5. First publications were published in 2020. By comparing number of publications in Figure 5 and number of citations in Figure 6 it is obvious that plasma medicine is new area of research with increasing interest.

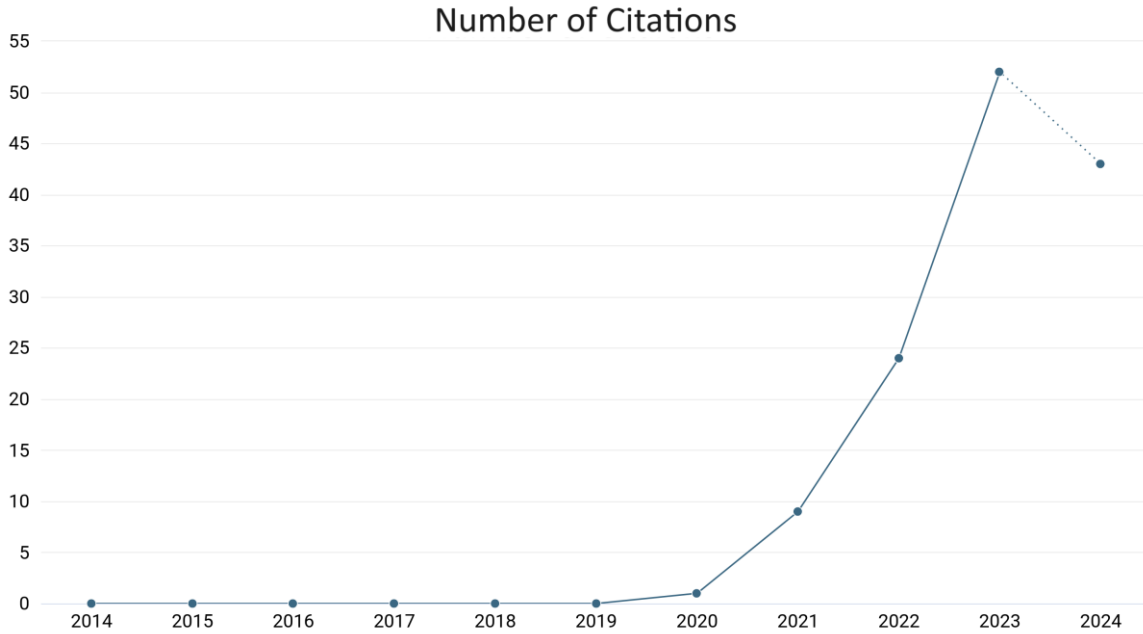


Figure 6. Total citations of documents per year during 2014-2024 period.

Due to number of publications and time span of only 5 years, significant statistical analysis cannot be done. On the other hand, it is expected that the majority of documents belong to engineering and computing science (Figure 7) which will eventually lead to increased number of publications in biomedical and clinical journals when researchers apply the designed models to their research.

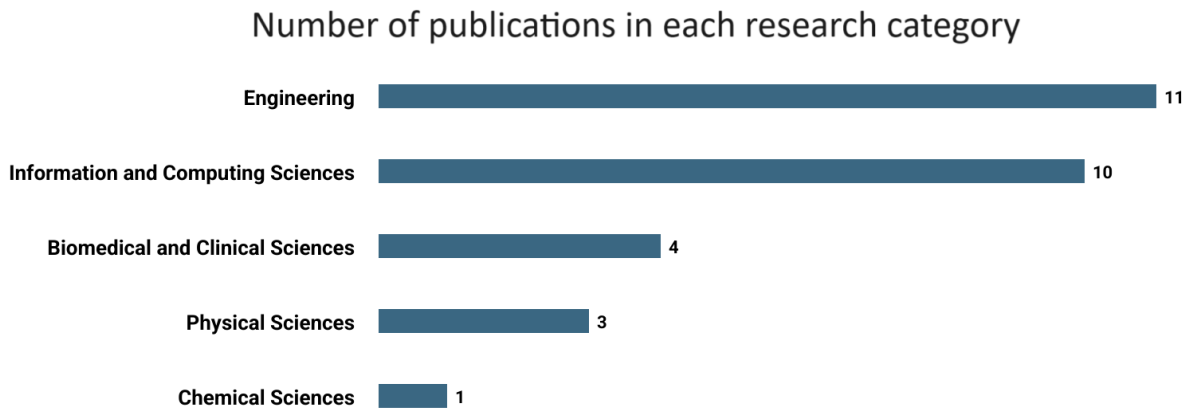


Figure 7. Distribution of documents by subject area.

Table 3. The most prolific authors, their organization, country and number of papers with citation number.

Researcher name	Organization	Country	Number of papers	Citations	Citations Mean
Utku Kursat Ercan	Izmir Katip Celebi University	Turkey	7	30	4.29
Mehmet Akif Ozdemir			7	30	4.29
Gizem Dilara Ozdemir			7	30	4.29
Ali Mesbah	University of California, Berkeley	USA	4	73	18.25
Onan Guren	Izmir Katip Celebi University	Turkey	4	30	7.5
Angelo D Bonzanini	University of California, Berkeley	USA	3	69	23
David B Graves	Princeton University	USA	3	69	23
Li Lin	George Washington University	USA	2	15	7.5
Michael Keidar	George Washington University	USA	2	15	7.5

The most prolific authors are presented in Table 3. It includes 9 authors from 2 countries, Turkey and United States of America. To better understand areas of plasma medicine that are applying new computational algorithms to their research, in Table 4 are presented researchers with areas of expertise (only areas associated with plasma medicine) and countries they come from. Utku Kursat Ercan from Turkey is using new computational algorithms to predict the antimicrobial activity and efficacy (bacteria are treated by plasma activated water (PAW) or direct cold atmospheric plasma), surface properties of biomaterials and oxidative strength of PAW. Ali Mesbah's (Berkeley, USA) area of research is personalized plasma medicine, including safe dose delivery.

Figure 8a presents co-authorship network of 9 researchers. Author with most connections is Ali Mesbah (Figure 8b) from Berkeley in USA with 8 co-authorship links, 13 total co-authorships in 4 publications. Followed by David B Graves and Angelo D Bonzanini.

Table 4. Pioneers, researchers who are applying new algorithms to their research

Researcher name	Organization	Country	Area of research
Utku Kursat Ercan	Izmir Katip Celebi University	Turkey	Wound healing, carcinoma cell and resistant bacterial strains treatments, biomaterials
Ali Mesbah	University of California, Berkeley	USA	Applied mathematics, analysis and control of complex chemical and biomolecular systems

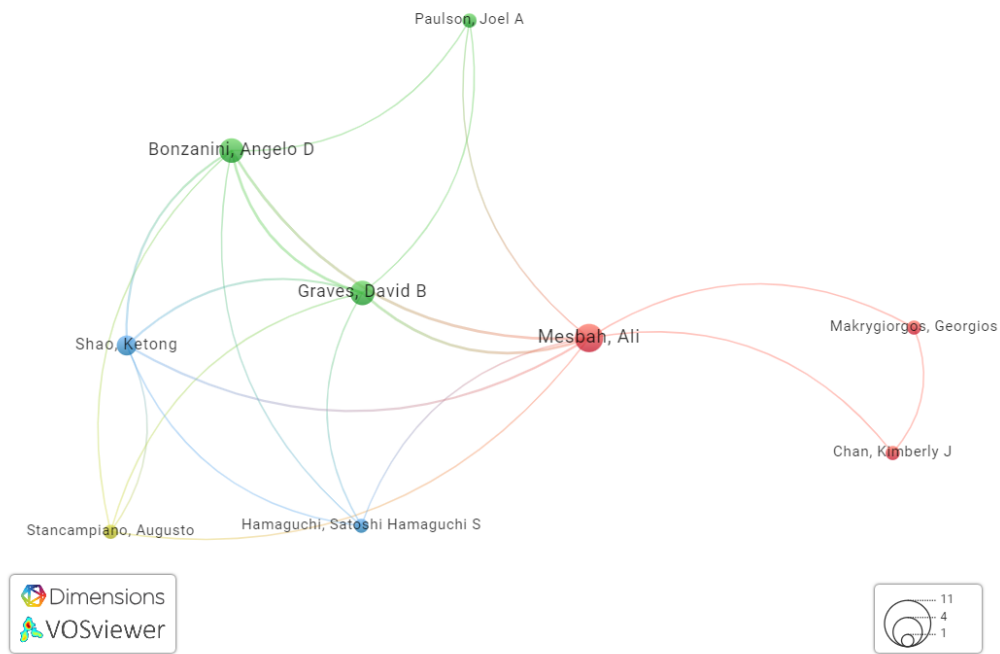


Figure 8 a). Analysis of co-authorship showing 9 researchers with highest number of co-authorship. Figure taken from Dimensions.

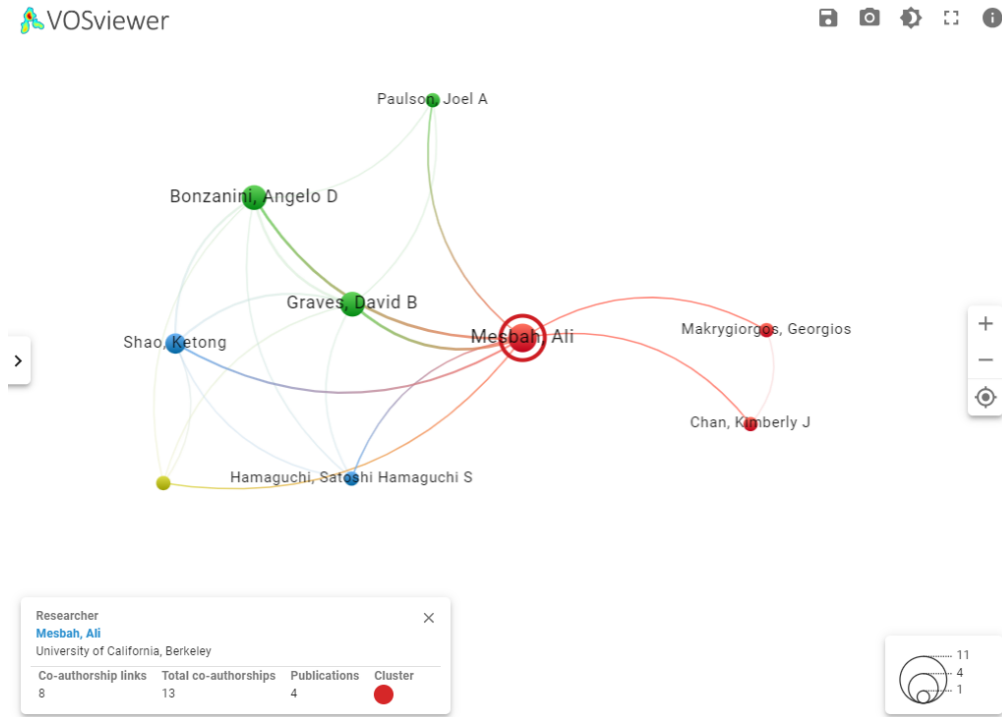


Figure 8 b). Analysis of co-authorship for biggest cluster Ali Mesbah. Figure taken from Dimensions.

Scientific conferences, workshops, journal special issues and seminars

1. Plasma conference with the most publications connected to the ML is Gas Electronic Conference (GEC). Table 5 presents number of publications by year with first publication appearing in 2017.

Table 5. GEC conference and ML connected publications by year

Year	Number of publications
2017	1
2018	3
2019	2
2020	14
2021	8
2022	4
2023	12

2. In 2020 at the GEC conference was held workshop on AI and ML:

Artificial Intelligence & Machine Learning in Plasma Science and Beyond –

Satoshi Hamaguchi, Osaka University, Japan

3. Journal Special Issues

a. Special Issue on Machine Learning, Data Science, and Artificial Intelligence in Plasma Research

IEEE Transactions on Plasma Science

b. Special Issue on Data Driven Plasma Science

Journal of Physics D: Applied Science

4. Online available seminars: *Data driven plasma science*, Satoshi Hamaguchi

LINK: https://www-ppl.eng.osaka-u.ac.jp/JSPS_Core/seminars.html

5. Open source code from the Lisbon group (<https://nprime.tecnico.ulisboa.pt/loki/>)

LoKI: <https://github.com/IST-Lisbon/LoKI/>

Conclusion

This report is written as an effort to draw attention on a new computational algorithms (ML, DL, AI and NN), their potential and to define the position of the plasma agriculture in the fast-growing field of ML. Database that was used, Dimensions, gave less than 10 results with only one strictly connected to the plasma agriculture. It shows that ML in plasma agriculture is at the beginning, with shown interest for that to be changed.

Comparison of ML (AI, DL and NN) usage in agricultural plasma with two scientific fields, agriculture and plasma medicine, was made. Agriculture came a long way in last 10 years. AI is widely used to analyze chemical composition of soil samples to determine which nutrients are lacking, identify or even predict crop diseases, identify optimal patterns and nutrient application times while predicting the optimal mix of agronomic products etc. That is all supported with more than 18,000 publications with more than 200,000 citations.

On the other hand, plasma medicine is showing more steady increase in number of publications, than plasma in agriculture. Few research groups with different areas of expertise can be singled out. Reason for that is probably availability of data, due to numerous experiments and repetitions required for plasma usage on humans.

The advantages of application of new computational algorithms is not questionable. But, do we have required amount of data available for those algorithms to be trained and then used? All of this leads to conclusion that more systematic approach to experiment design is needed so the data obtained correspond requirements of the new algorithms. To understand plasma, so it can be modeled, emphasis has to be on the experiment and it's diagnostic. And then the attention can be drawn to the outcome and the plasma effect on treated object.

Considering all that is presented in this bibliometric analysis, this report can be used as a motive for new experimental setups designed from different perspective, to provide necessary data for plasma modeling. Also, it can be helpful for scientific networking (future actions and projects) and education guidelines in the field of ML and AI.