



Leveraging Text Mining for Better Policy Making to Combat Climate Change: A Bibliometric Analysis Correlating Precipitation Changes with Agriculture in India

Nikita Ray^a, Sachin^b, Jhanak Saxena^c, Than Dar Aung^d and Manisha Goel^{e*}

^{abcde}Department of Biophysics, University of Delhi South Campus, New Delhi – 110021, India

Received: 09 August 2024; Accepted: 20 August 2024

India is widely considered as an agrarian economy, where agriculture defines the livelihood of a large population. However, our current agriculture practices are facing formidable challenges due to rapid progression in climate change. Climate change is leading to various unprecedented spatio-temporal alterations in weather patterns resulting in devastating events such as drought, heat waves, erratic and intense rainfall. The impact of climate change on agriculture is widely accepted, and the possible threats and its affects are continuously being studied. The idea is to gain enough understanding of various factors affecting the precipitation, to be able to predict changes in precipitation patterns as well as design mitigation strategies to minimize the impact of such disturbances, through effective policy making. However, policy making requires synthesis of disparate factors, involving not only detailed understanding of the individual component but also the correlation between them. Establishing such correlations can be challenging, given the diverse array of information available, especially if presented in an unstructured format. Scientific literature often explores a problem from varied facets, which can be highly useful but hard to correlate manually. However, text mining can serve as a useful approach to generate knowledge out of such varigated information pool in minimal time. Text mining is an evidence-based method for knowledge generation, which works through quantification of information present in texts, helping in understanding the correlation between various key terms. The present study is an extension of our goal at Semantic Climate's Climate knowledge hunt hackathon organised on 26-27th February 2024 (https://semanticclimate.github.io/p/en/past-events/climate_knowledge_hunt_Feb24/). Here we intended to leverage the text mining strategy to extract pertinent information from literature sources, such as scientific papers, related to precipitation patterns in India. The study relies solely on the extraction of published data and no ground data related to agriculture or precipitation has been used directly. The key findings emerging from the information corpus through text mining will have potential to direct the policy makers towards comprehending the gaps created in our agricultural practices due to climate change. This understanding will allow policy makers to devise more effective and potent solutions to combat the effects of climate change on Indian agriculture, thereby safeguarding the broader socio-economic fabric of the country.

Keywords: Bibliometry, Climate Change, Policy, Text mining, precipitation, Agriculture

Introduction

India is a developing country where income of a major portion of the population is dependent on agriculture. Agricultural output relies on various interdependent factors such as the type of soil, quality of seed used, precipitation and water availability, pests and the overall climate of the geographical location⁸. Climate change therefore remains an important variable. Climate change is generally characterized by an increase in the global average surface temperature⁴. However, it is common knowledge that climate change is leading to various spatio-temporal alterations in weather patterns

leading to several types of devastating events such as excess rainfall, heatwaves and drought¹¹. These unprecedented changes temper with various factors coupled to the agricultural yield by altering the quality of soil, modulating the precipitation index and eventually disturbing the mineral cycles in the soil. Thus, climate change is not a single variable associated linearly with crop yield, instead it is a network of correlated factors that together control the outcome.

Due to various social, and economic limitations, it is challenging for developing countries to absorb the shock of variabilities resulting from the climate change. Recently, the effort to tackle climate change has gained momentum and there is an increased co-operation and research in climate sciences globally.

(Corresponding Author)

E-mail: *manishagoel@south.du.ac.in

Finding practical solutions to overcome the adverse effects of climate change are the ultimate goal of every organisation working on climate change impacts. Intergovernmental Panel on Climate Change (IPCC) is one of such body of United Nation which assesses the science related to climate change (<https://www.ipcc.ch/>). It reports the state of knowledge of climate change and also provides access to a huge amount of data dedicated to various variables associated to climate. There are three working groups of IPCC e.g. working group (WG) I, II and III which look into the physical scientific basis, vulnerability with adaptations and its impact and mitigation of climate change, respectively. IPCC also contains special reports having various chapters, one of which is regarding the effects of climate change on the food system¹².

In order to deduce new strategies and solutions to combat the harmful impact of climate change, a large number of studies have been carried out in recent years. This has generated a huge corpus of data in the form of reports and statistics. This data corpus is a potent source of information in providing solutions to this issue. However, relying solely on human intervention for extracting useful information from such huge corpus is an extensively time-consuming process due to its large size and complexity of unstructured data. Such unstructured data can be converted to structured form to extract knowledge using text mining approaches. Text mining is a process of data mining where informative patterns are extracted from mosaic pool of text and presented in a structured form, which is helpful in gaining insights¹.

In this present case study, two literature resources: PubMed¹³ and PubMed Central (PMC)⁹ are used for extracting information in the form of abstracts and fulltext articles. An abstract of an article generally contains the essence of that article as it contains both the premise as well as the novel findings of the study. The abstract of most articles is freely available for the scientific articles of PubMed, while the fulltext of freely available articles can be taken from PMC. Here, we aimed to extract important words termed “keywords” from abstracts (in the first step) and then look for the significance of extracted keywords in the fulltext articles. While retrieval of abstract from PubMed is a one step process, retrieving fulltext article from PMC is cumbersome process as one can only retrieve one fulltext article at a time and only in *.pdf* format. This step therefore requires some level of computational programming skills and tools to

optimize this step. However, bulk retrieval and conversion of *.pdf* files to readable file format of all fulltext articles for any query phrases at PMC can be done by *pygetpapers*² and *pyamihtmlx*⁴, respectively.

For the current pilot study, we chose to limit our text mining-based data collation to the topic of “climate change-based change in precipitation and its effect on agriculture” in the context of only our country, India. A preliminary study on this objective was conducted at Climate knowledge hunt hackathon organised by team of Semantic climate on 26-27th February, 2024⁵. The extension of that study starts by compiling all the abstracts related to climate change affecting precipitation, increase and decrease in rainfall and effect on soil in Indian agriculture for collecting keywords. To understand the correlation between the keywords selected, literature related to IPCC, mineral cycles in soil and fulltext articles from PMC dealing with the role of precipitation in agriculture were added as text resource. This will help in collecting information about the less studied or the neglected factors in solving the climate change issues effecting precipitation in India. It includes focusing on general keywords, names associated to crops and all the Indian states.

Methodology

In the present case study, we collected abstracts from PubMed using query phrases complying to our topic of study (Fig. 1 Step 1). Each query phrase generates an individual document file with all the abstracts containing that term. Here, we used seven different query phrases (Fig. 2) associated to the effect of precipitation, different soils and rainfall alteration. Further, text mining was done on these files using the custom-made text mining module scripted in python programming language (https://github.com/nray029/semantic_hackathon_2024) (Fig. 1 Step 2), which removed the commonly used grammar and language words. A list of words with frequency of occurrence was generated from which highly occurring words were selected. Through human intervention, words that appear to be related to our topic of interest were selected as “keywords” (Fig. 1. Step 3). Among these keywords, names of crops were extracted separately.

Extraction of fulltexts from publicly available literature database (steps 4- 5)

In this step, we extracted fulltext version of the articles from PubMed Central (PMC) using the query

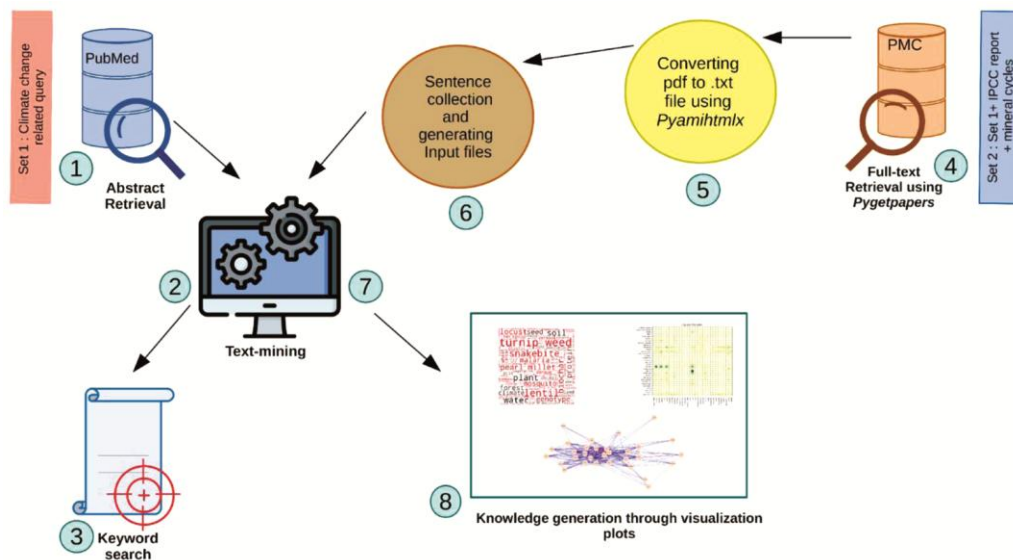


Fig. 1 — Graphical representation of the pipeline of information extraction from the literature corpus. Numbers in the circles represent the consecutive steps of the pipeline. A brief description of each step is provided in Methods. Keywords selection from abstracts (steps 1-3)

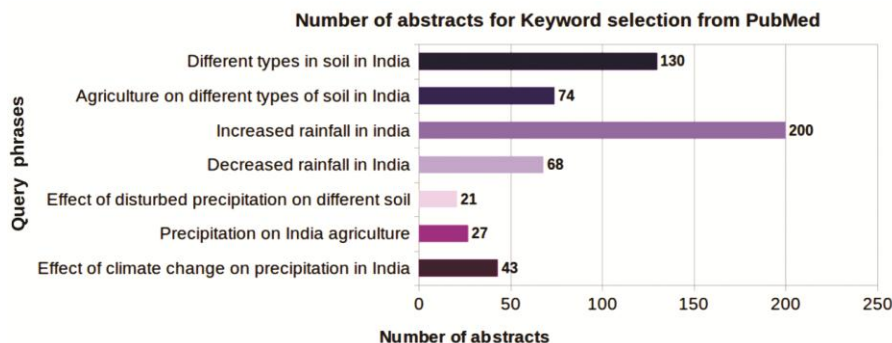


Fig. 2 — Graphical representation of number of abstracts collected for the 07 different query phrases from PubMed.

phrases similar to the query phrases used in step 1 and some additional query phrases related to IPCC, effect of water availability on soil ecosystem and mineral cycle in soil. To overcome the issue of extracting bulk files from PMC, we used *Pygetpapers* which is a python module used to extract all the files related to a query phrases from PMC in .pdf format (Fig.1 Step 4). These .pdf files were converted to .html through *Pyamhtmlx* python module followed by converting all them to .txt format using in-house scripts (Fig.1 Step 5).

Mining Information from Fulltexts (steps 6-8)

Further, these fulltexts were subjected to textmining through our custom-made text mining modules scripted in python programming language to extract sentences containing the keywords selected in step 3³. We have also selected all the Indian states as additional keywords, so as to find their correlation between crops

and climate change keywords selected in step 3. The resultant keywords were distributed in three categories e.g. name of the Indian States, name of the crops and the climate change related words. The sentences were collected and converted back to the file formats required for text mining (Fig.1 Step 6).

The next step involves calculation of “average frequency” of a keyword from occurrence frequency of each keyword along with the number of articles in which these keywords were found (Fig.1 Step 7). The calculation is summarized below:

$$\text{Average frequency of keyword} = \frac{\text{Total number of occurrence of keyword}}{\text{Number of articles where the keyword is found}}$$

Finally, the averaged frequency for each keyword is plotted in three different wordclouds, e.g. first for all the climate change related keywords including

crops, second for names of indian states and third one specifically for name of crops. We also plotted interaction network graphs and heat maps for the three categories using Matplotlib module of python^{3,4}, both of which showcase the number of times these keywords exist together in the articles. Interaction networks depict the relationship among words within each of the three categories, whereas the heatmaps depict the correlation among keywords from different categories. In interaction network for states, if two keywords were found together in an article they are found connected while for rest of the two categories, keywords are found connected if they are found together in a sentence. Similarly, keywords for

different categories, if found together in a sentence, are considered as connected (Fig.1 Step 8).

Results

Keywords selection from abstracts (steps 1-3)

In this study, we selected 7 important query phrases relevant to our topic for retrieving abstracts from PubMed (Fig. 2). Query phrases related to precipitation, soil and rainfall had 85, 268 and 204 abstracts, respectively. After performing the first text mining process (Fig.1 Step 2), we were able to extract keywords relevant to the topic of interest. We found a total of 193 keywords from all the abstracts which are enlisted in Table 1 (Fig.1 Step 3). These keywords

Table 1 — Keywords collected from query phrases

S.No.	Query phrases using in PubMed	Keywords
1.	Effect of climate change on precipitation in India	Heat stress, Emission, Deforestation, Chlorophyll, Planting, Coastal, Climate warming, Flood risk, Salinity, Extreme Salinity, Herbivore Effect on Biodiversity, Aerosol, Forest, Genotype, Mosquito, Spicy, Extreme Precipitation, Pollutant, herbivore, Water, pest, Endemic, Irrigation, Nitrogen, nematode
2.	Increased rainfall in India	Aerobic, rainfall, heatwave, malarial, giardia, soil penetration resistance, co-infection, rain gauge, fruit, forest fire, biotype, lantana, transpiration response, apple, anesthetic, rainfed rice, nematode, watershed management, soybean, corn, saffron, meliodosis, gregarization, snakebite, halophyte, enteric infection, orchard, weed, tribal, chickpea, cholera, genotype, scrub, typhus, wheat, dengue, contaminant, cryptosporidium, coffee, malaria, maize, salt, pollination, drain, mizoram, pollinator, parasite, groundnut, nutrient, irrigation, eutrophication
3.	Agriculture on different types of soil in India	Germination, Ecohydrological, Ecological Flooding, Microbial Abundance, Permafrost, Surface soil water, Shrub, land Cover, Isotopic, Desert, Landscape, Biodiversity, Forest, Rainfall, Vegetation, Tropical, Population, Restoration, Drought, Plateau, Soil, Arthropod, Moisture, Hotspot, Richness, Springtail, soybean, wheat, biodiversity, soil
4.	Role of precipitation on India agriculture	genotype, protease, contaminant, tyrosinase, phosphorus, drought, xenorhabdus, transformation, lentil, protein, indica, allele, plant, death, water availability, histone, polymerization, trait, decontamination, yield loss risk, role of soil, biodiversity
5.	Effect of disturbed precipitation on different soil	Ecological flooding, shrub, windstorm, seed production, ecohydrological, resilience, germination, riparian, isotopic, soil, animal, microbial abundance, rotation, land, desert, microarthropod, biocrust, florida, watershed, vegetarian and edaphic, Human activity, flux, tropical forest, biodiversity, climate, landscape, productivity, arthropod, vegetation, rainfall, ecological forest, area, fire, site, biomass, wetting, plateau, warming, springtail, population, moisture, habitat, hotspot, carbon
6.	Decreased rainfall in India	Locust, Groundnut Yield, Salt Management, Leptospirosis, Giardia, Savanna, Drought, Odisha, Cavitation, Water Quality, Flash Floods, Rain gauge Density, Germination, Dust, Polyethylene, Deforestation, El Nino, Heat Tolerance, Cropping Intensity, desert, Prevalence, Rainfall, ecosystem, Cloud, Diversification, Climate change, Soil, Vegetation, Pollution, Irrigation, Cropping Concentration, Adaptation, Rice yield, Aerosol, Groundwater Level
7.	Different types of soil in India	Phytoremediation Metatranscriptomic, Micorrhizal, Earthworm, Biocide, Intercropping, Azospirillum, Termite, Surface runoff, Phytophthora, Chromium, Detergent, Aspergillus, Biochar, Groundwater, Plasticity, Fusarium, Plastic, Forest Divergence, Pesticide, Acid, Remediation, Cancer, Streptomyce, Fertilizer, Cyanobacteria

also contain 42 words associated to names of the crops (Supplementary Table S1)

Extraction of fulltexts from publicly available literature database (steps 4- 5)

For extracting fulltext articles from PubMed Central (PMC), 18 query phrases were used such that a majority of articles related to the effect of climate change in precipitation affecting Indian agriculture can be included in our study (Fig. 3). The articles also include reports of the Intergovernmental Panel on Climate Change (IPCC). We found a total of 1191 articles directly associated to the precipitation effecting Indian agriculture. Interestingly, increased rainfall in India (286 articles) has more articles than decreased rainfall in India (84 articles), however the summation of both (370 articles) is less than the total articles related to precipitation. Effect of water availability on different aspects related to agriculture such as ecosystem, plant, microbe and mineral cycles brings up 117 articles while IPCC reports sum up to 132 articles.

For better visualization of the occurrence of words in the articles, we chose to depict them as the wordclouds, which displays two parameters at once, the frequency of occurrence of the word in the article (font size) and number of articles in which that word is found (color gradient). Three separate wordclouds were generated. The first included all the keywords selected in step 3 (Fig. 1). The second looked at the mention of individual Indian states while the third was limited to crops as the keywords.

In the wordcloud for all the keywords where the color gradient ranges from red to black indicates increasing number of abstracts in which the word was

found. It was seen that most of the words were red but font size was comparatively large suggesting that the keyword was prominent in small number of articles. For example, the keyword “lentil” (red and large font) was occurring 453 times in 15 articles where the word “rice” (black and small font) was occurring 1959 times in 244 articles where the average frequencies were 30 and 8, respectively, which implied that a large number of articles talk about “rice”, but “rice” may not be the most important word of these articles while a small portion of articles talk about “lentil” and it was the highlight of the articles. This also suggests that “rice” was being talked about in correlation with other important factors (keywords). In Fig. 4A, it was found that font size of some of the crops were larger than other keywords indicating that crops were the major factor considered in association to climate change. Interestingly, words such as pest (mosquito), animal (snakebite) and some diseases (malaria, dengue) were also found to be significantly associated with our topic of study.

The association of states of India is also a major highlight as it offers insights about the regions of India which have been studied more and the neglected states too. It was observed that Punjab, Delhi and Tamil Nadu were the most highlighted states in most of the articles (Fig. 4B). Uttarakhand, Kashmir, Jammu, Sikkim, Assam, Kerala and West Bengal were also found to be prominent among a good number of articles with high average frequency. Some of the states were still little underreported including Andaman, Telangana, Himachal Pradesh, Nicobar, Mizoram, Nagaland, Manipur, Bihar, Madhya pradesh and Ladakh.

A closer look at the wordcloud for crop (Fig. 4C), shows that the keyword “seed” was darkest

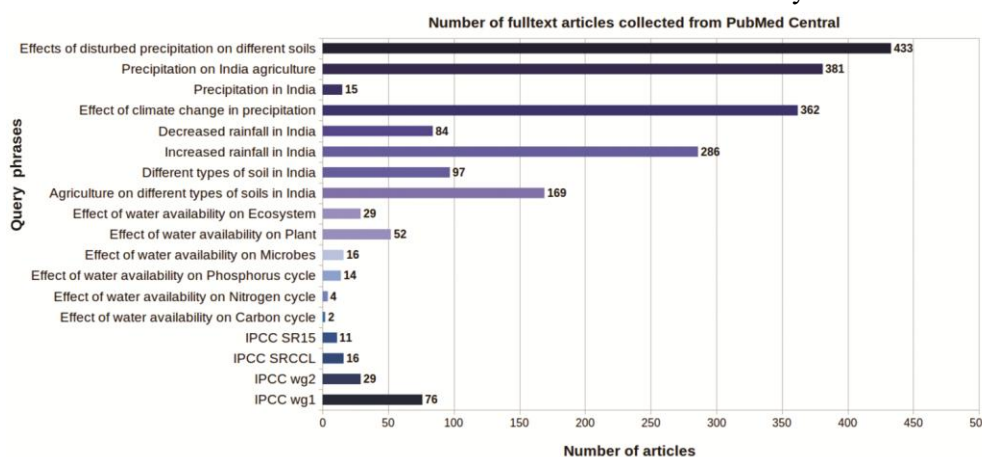


Fig. 3 — Graphical representation of the number of fulltext articles collected from PubMed Central (PMC) against different query phrases.



Fig. 4 — (A-C): Wordclouds generated based upon occurrence of a word indicated with the font size and number of abstracts in which they are found indicated in color gradient from *red* to *black* for low to high number of abstracts. (A). General Keywords (step 3) (B) Indian States (C) Crops.

suggesting that was the most of the associated terms in a large number of articles. The staples such as maize, wheat, corn, sorghum and rice (depicted in brown color) were found prominent in larger number of articles than other crops. Different types of millets, especially pearl millet and bean mostly soybean were also highly associated. Surprisingly, rare crops such saffron, marigold, bamboo, cardamom, moringa, coffee and papaver were also found as keywords.

For understanding the extent to which two keywords were connected, analysis of networks played a major role (Fig. 5). In the network generated for three categories of the keywords, some of the keywords were removed as they were outliers (they made no connections to any other keyword). In the overall network graph, it was found that words like water, soil, land, plant, climate change and climate are connected very closely to each other (Fig. 5A). Similarly, carbon, ecosystem, population, nitrogen, nutrient, biomass are also closely connected (Fig. 5A). The terms such as drought, vegetation, moisture, rainfall, warming, emission and pest were closely connected to the former cluster. The terms such as fertiliser, salt, coastal, plastic, pollution, drain, desert, irrigation, planting, restoration, salinity were not too close to any of the cluster while words such as groundwater, pollutant and plateau were too far to be considered as significant connection with the major keywords according to present studies.

The network generated for states of India indicates which states were closely studied together (Fig. 5B). It was observed that Chhattisgarh, Puducherry, Manipur, Andaman, Mizoram, Sikkim and Arunachal Pradesh were at a larger distance from the other states though connected. However, Arunachal Pradesh was found to be connected to Sikkim, Manipur and Mizoram. It was also found that states such as Gujarat, Rajasthan, West Bengal, Karnataka,

Maharashtra, Punjab, Delhi, Tamil Nadu, Kerela, Uttar Pradesh and Gujarat were very closely connected indicating that they were mostly studied together or compared to each other. Wherein, Telangana, Haryana, Madhya Pradesh, Jharkhand, Andhra Pradesh, Odisha, Himachal Pradesh, Uttarakhand, Jammu, Kashmir, Assam and Chandigarh were found connected to these mostly studied states but to a smaller extent. It was also seen that states like Kashmir, Jammu, Uttarakhand, Assam and Himachal Pradesh were close to each other indicating that these clusters were studied together most of the time. This may be due to their geographic location, as most of these are mountainous states, so they may have similar type of crops.

For crops, the networks show that in how many articles these crops were found to co-occur in same sentences (Fig. 5C). It was found that marigold, pomegranate, moringa, pulse and rainfed rice are at distance from the whole cluster, indicating these crops were not studied in association with other crops. Whereas mungbean, bamboo, strawberry, okra, lentil and pigeon pea were studied more often compared to the former cluster. It was seen that durum wheat, millet, wheat, maize, rice, soybean, corn, pea, sorghum, tomato and chickpea were in very close association with each other indicating they were studied together very commonly. Another cluster including cowpea, tobacco, coffee, banana and apple, was also seen adjacent to the most frequently studied crops.

Heatmaps helped in understanding the correlation of keywords from two different categories based on their co-occurrence (Fig. 6). The heatmap with the name of states of India and the crops, suggest which crops were more strongly associated with particular state (Fig. 6A). It was found that Punjab is highly associated with maize, wheat, seed and rice while

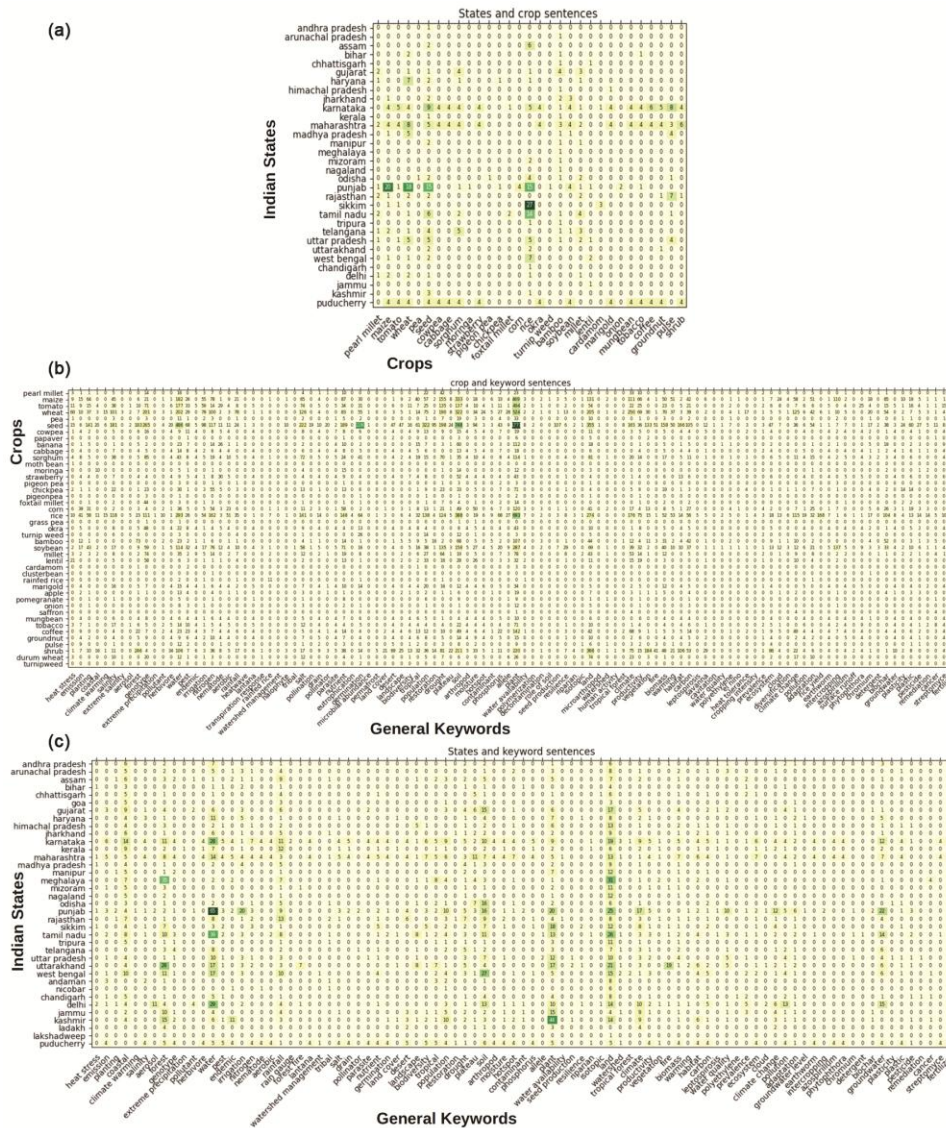


Fig. 6 — (A-C): Heatmaps representing the co-occurrence of words with another word. The color gradient from yellow to green showcase the low to high co-occurrence of two words. (A) Co-occurrence of name of states of India with the name of the crops. (B) Co-occurrence of name of crops with the keywords (C) Co-occurrence of name of the states of India with the keywords.

with the climate change (Fig. 6B). It was found that heat stress is common with the crops like lentil, wheat, maize, tomato and rice. Many studies appear to report “genotype”, for various crop such as wheat, maize, tomato, foxtail millet, chickpea, okra, soybean, millet, lentil and durum wheat. Other than prominent crops, “nitrogen” was found to co-occur with corn, chickpea, banana and groundnut. “Climate”, keyword was seen co-occur also with many crops other than prominent crop which includes pearl millet, corn, bamboo, lentil, coffee and durum wheat. Heat tolerance co-occurs more with wheat but not with other prominent crops. “Carbon” was found to co-

occur with coffee, bamboo and corn where biochar was also prominent with banana.

Similarly, in the heatmap of name of the state of India and the common keywords, it was observed that states such as Karnataka, Maharashtra, Punjab and Puducherry were mostly associated with a large number of keywords. Similarly, some of the terms such as forest, water, rainfall, drought, soil, plant, groundwater and fertilizer were also associated with almost all the states (Fig. 6C). It was observed that keywords such as heat stress, coastal, pollutants, aerosol, climate change, pollution and climate are mostly associated to Delhi and Punjab. Some of the

terms and states were frequently co-occurring such as irrigation with Punjab; forest with Meghalaya and Uttarakhand; water with Karnataka, Punjab, Tamil Nadu, Delhi, Uttarakhand and Uttar Pradesh; rainfall with Karnataka, Kerala, Rajasthan and West Bengal; soil with Gujarat, Odisha, Punjab and West Bengal; fire with Uttarakhand; water quality with Punjab; and groundwater with Karnataka, Punjab, Tamil Nadu and Delhi.

Discussion

Climate change is a major concern all throughout the world. The three working groups e.g, WG I, WG II and WG III of IPCC are playing a crucial role in monitoring the impact of climate change on various sectors. One such sector is food safety which requires the understanding of how agriculture is affected by climate change, globally. India, a country known for its dependence on agriculture for socio-economic needs, is highly impacted through climate change as it alters the agricultural scenario. There was an estimated relative yield loss of 21% of wheat and 6% of rice in India¹⁰. Various topics associated with the effect of climate change on agriculture in India are definitely being addressed by a number of researchers. As their reports pour in, it becomes difficult to compile all the information in a succinct manner. In this study, we have tried a pilot pipeline for parsing relevant information present in these papers though the creation of “keyword” based networks.

Scientific literature associated with the query phrases as associated with “climate change” collect significant number of text identifiers, termed as the “keywords”. Three categories of keywords, such as general words associated to climate change, name of the Indian states and name of the crops were tried and found to be effective in understanding the climate change issues that Indian agriculture is facing. Most of the articles in PMC are related to precipitation, rainfall and soil in Indian agriculture context but a small number of them included the effect of water availability on mineral cycles, microbes and plant which could be a major delimiting agent in understanding the effect of climate change in growth of a plant.

Wordclouds offer interesting insights into the frequency of topics being studied through published research. The keyword occurrence ranged from words with bright red color and large font size and others with dark brown to black color and small font. The

former indicates that these words were highlighted content of some of the articles while the latter indicates that many articles mention these keywords but they may not be the main highlights.

The data presentation for Indian states, suggested not all states have equal representation in the studies. It could potentially be due to the reason that some states are either more advanced in understanding the effect of climate change or are more dependent on agriculture. Similarly, the staples of Indian diet, such as maize, wheat, rice, sorghum and soybean as well as lentil and millets were also the most studied. The clusters of words containing soil, land, plant and climate are closely connected while those associated to soil ecosystem such as minerals and biomass are in close association. Similarly, cluster of states show that the study among these states is similar. The networks in clusters of the staples with chickpea, tomato and pea shows that these are associated with similar issues. Some of the keywords appeared to be outliers or orphans, showing no connection with other words. This indicates that these terms were thought to be important by few researchers but not enough studies have progressed in that direction.

The better picture was presented when we tried to understand the correlation of Indian states with other factors and the crops. Rice is the staple crop for most of the southern and north-eastern regions, this was also evidenced in our study as the association of such regions was high with rice. However, the mentioned neglected states indicate that either these states have lack of resources to perform such studies or there is not much affect from climate change reported yet. So, some efforts need to be taken to understand whether these states really are devoid of climate change effect or it is just lack of resources. If the former is true, these states can be targeted to grow crops which cannot be grown in climate change effected states.

The association of different climate change keywords with the different crops clearly shows that climate change alters various factors of agriculture and not all the factors can be studied all together for all the crops. However, the specific cluster of keywords with specific cluster of crops indicates that these factors are thoroughly monitored in some crops. Similarly, the keywords restricted to the polluted cities like Punjab and Delhi indicate that these cities are more affected and looking for solutions of climate change while some of the specific words with specific state shows that the particular state is associated with

the particular factor. This can be seen in an example as Punjab is affected more with water quality and Uttarakhand is more prone to forest fire. Both are the reason and cause of climate change. Similarly other factors can be monitored for different states.

To fight the climate change affect in agriculture, we need to have better understanding of the geographic location, crops and the type of change occurring in climate at a given place. The adaptation strategy includes increasing cropped area, using drought resistant crops, use of groundwater, using crops with less harvesting duration and using drip irrigation methods for saving water. Different strategies need to be applied to different states. Like, if western region is shown to hit drought too often then drought resistant crop is a better solution in that area, whereas cropped area can be increased where the climate change affects are not too prevalent as of now.

Conclusion

Climate change is a global concern which majorly affects food safety. A good number of researchers are actively dedicated to understanding and documenting this issue resulting in accumulation of a large amount of literature. In handling and extracting useful insights, text mining has been proved to be a very incredible approach. It helped reducing time on knowledge gain from such huge corpus of scientific papers so as to discover potential solutions to such devastating event.

India, an agriculture-based economy has been affected with the climate change which is expected to alter the socio-economic lifestyle of a huge population. Adapting to newer methods and developing a balance between climate change and agriculture is the paramount solution to this issue. An understanding of the common and unique factors affecting different regions as well as specific crops should be taken into account while devising new policies. The preliminary observation with reference to this study done at Climate knowledge hunt hackathon organised by

Semantic Climate on 26-27th February, 2024 indicates that studies similar to ours can help generate that data quickly through keyword-based text mining approaches.

References

- 1 Frijters R, van Vugt M, Smeets R, van Schaik R, de Vlieg J and Alkema W, Literature mining for the discovery of hidden connections between drugs, genes and diseases, *PLoS Computational Biology*, 6(9) (2010) 1–11. <https://doi.org/10.1371/journal.pcbi.1000943>
 - 2 Garg A, Smith-Unna R D and Murray-Rust P, pygetpapers: a Python library for automated retrieval of scientific literature, *Journal of Open Source Software*, 7(75) (2022) 4451. <https://doi.org/10.21105/joss.04451>
 - 3 https://github.com/nray029/semantic_hackathon_2024
 - 4 <https://github.com/petermr/semanticClimate/tree/main/pyami.html>
 - 5 https://semanticclimate.github.io/p/en/past-events/climate_knowledge_hunt_Feb24/
 - 6 Irwandi H, Rosid M S and Mart T, Effects of Climate change on temperature and precipitation in the Lake Toba region, Indonesia, based on ERA5-land data with quantile mapping bias correction, *Scientific Reports*, 13(1) (2023) 1–11. <https://doi.org/10.1038/s41598-023-29592-y>
 - 7 Hunter J D, "Matplotlib: A 2D Graphics Environment", *Computing in Science & Engineering*, 9(3) (2007) 90-95.
 - 8 Pathak H, Impact, adaptation, and mitigation of climate change in Indian agriculture, *Environmental Monitoring and Assessment*, 195(1) (2023). <https://doi.org/10.1007/s10661-022-10537-3>
 - 9 PMC Open Access Subset [Internet]. Bethesda (MD): National Library of Medicine. 2003 - [cited 2024/06/17]. Available from <https://www.ncbi.nlm.nih.gov/pmc/tools/openftlist/>
 - 10 Sharma A, Ojha N, Pozzer A, Beig G and Gunthe S S, Revisiting the crop yield loss in India attributable to ozone, *Atmospheric Environment: X*, 1(December 2018), (2019) 100008. <https://doi.org/10.1016/j.aeaoa.2019.100008>
 - 11 Swami D and Parthasarathy D, Assessing effectiveness of agricultural adaptation strategies in context of crop loss: a case study of the Indian subcontinent, *Regional Environmental Change*, 22(2) (2022). <https://doi.org/10.1007/s10113-022-01921-3>
 - 12 Uyttendaele M, Liu C and Hofstra N, Special issue on the impacts of climate change on food safety, *Food Research International*, 68 (2015) 1–6. <https://doi.org/10.1016/j.foodres.2014.09.001>
- White, Jacob. "PubMed 2.0." *Medical reference services quarterly* vol. 39,4 (2020): 382-387. doi:10.1080/02763869.2020.1826228