# Sentiment Extraction from Naturalistic Audio of Agricultural Expert Opinions

## Zaid Sarfraz[1], Abdul Razzaq[1*], Ayesha Hakim[1],  Zulqrnain Ali[1], Umar Ijaz Ahmad[2], Attiq-ur Rehman[1], Muhammad Aziz-ur Rehman[1], Sharaiz Shahid[1], and Tausif-ur-Rehman Saddique[1]

[1]Department of Computer Science, MNS University of Agriculture, Multan, Pakistan.
[2]Department of Agribusiness and Applied Economics, MNS University of Agriculture, Multan, Pakistan.
[*]Corresponding Author: Abdul Razzaq. Email: abdul.razzaq@mnsuam.edu.pk.

---

**Abstract:** Opinions of Agri Experts play a vital role in planning and managing crop productivity. We use a variety of text mining techniques to search for the strong sentiment keywords in social media, fiction, non-fiction, movies, novels, comics, stock exchanges, etc. Among the most widely used NLP techniques is sentiment extraction. Today, farmers are facing some difficulties maintaining crop productivity. They need to consult with agricultural experts for maintaining productivity. However, it is difficult for farmers to keep in touch with experts in agriculture industry. In the Proposed Study, experts offered different opinions based on their experience. Using decoded ASR (Automatic Speech Recognition) and sentiment model, audio data was processed by a machine learning algorithm. Then it was stored as a context for training. As a result of analyzing the stored text, the input by the user was analyzed and relevant answers were classified as output. Afterwards, it was translated into a national language like Urdu in Pakistan. This was done to make it more understandable to the farmer who lacks formal education, as opposed to textual data, which he may not understand. This solution became easy to understand and resulted in better outcomes before and after the harvesting of crops in different areas.

## 1. Introduction

Over the last decade, many aspects of sentiment analysis have been explored, such as detecting conscious and unconscious sentences, categorizing sentences as positive, negative, or neutral, detecting the person expressing the feelings and the focus of the sentiment classification, and detecting feelings of pleasure, fear, and anger in the text, sentiment analysis has become more sophisticated [10]. conducted surveys that summarized several methodologies. Records are stored electronically in every domain of the modern world, as text data has expanded dramatically in industry, business, technology, and agriculture. IoT [5], Cloud Computing [14], Artificial Intelligence [19], Machine Learning [2], Deep Learning, and Data Science have all helped agriculture become more inventive [16]. It is unclear that fertilization methods affect crop production [23]. Punjab is Pakistan's principal agriculture zone for major and minor crops, accounting for 18.9% of GDP and 42.3 percent of the jobs market [4]. We've had a lot of textual data in the agriculture area, thanks to various platforms like Twitter, Facebook, and LinkedIn groups [12]. The research conducted semantic analysis on agricultural datasets to determine similarities, sentiments, emotions, sensations, and beliefs about crop productivity.

Text sentiment analysis is an enormous field of Natural Language Processing (NLP). Opinion mining analyzes the opinions of people, emotions, appraisals, and evaluations towards entities such as services, individuals, events, products, topics, and their attributes [11]. A farmer cannot maintain optimal crop productivity without consulting with agricultural experts and getting their opinions. Unless they engage in interactive conversations with the experts, they cannot sustain crop productivity at optimal levels. This

study's primary objective is to provide farmers with the information they need in order to preserve crop productivity. A system that can give them expert advice and opinions. It is very interactive with farmers who may not have access to schooling and, as a result, cannot read textual data. However, when they receive expert opinion in audible medium, a better platform for communication between user and system by providing them native language medium. This will play a vital role in agricultural productivity. Using this solution, farmers will get comprehensive knowledge on soil seed, climate change, and synthetic fertilizer. It's our main objective to provide expert advice to farmers. They might use it to discover the effect of crops input and stress on major crops like wheat, maize, cotton, and sugarcane for better productivity. The disastrous effects of climate change, specifically heat, and drought, are affecting crop productivity and performance all over the world. By providing farmers with the opinions of experts, this type of solution is needed in Pakistan to increase agricultural productivity. To the extent of our knowledge, semantics is often used to forecast crop yield, estimate polarity, find support by body motion, and poses [17], but it hasn't yet been applied to farming, so it represents a significant opportunity. The analysis of the expert opinions made major contributions. The main objectives are as follows:

Collection of expert opinions of agricultural experts through interviews. (Audios)

To perform sentiment analysis on naturalistic audio samples.

To perform an analysis on the extracted data of opinions from agricultural experts.

Determine the critical aspects that affect agricultural production that can assist policymakers and farmers.

## 2. Related Work

Semantic analysis defines the process of understanding natural language, that humans communicate based on meaning and context. When analyzing natural language content, the first step is to read every word to capture the true meaning of the text. Semantics is the study of meaning in language. Any part of a text, or even a single word, paragraph, or entire book, can be analyzed.

Text-Mining (TM) is a vital research area. Now a large amount of data is available in the agriculture industry. A technique applied to many types of information, including descriptive documents, sentences, words, books, stories, blogs, agriculture forums, and social media. KDD (Knowledge Discovery in Database) preprocesses documents before they are ready for analysis. Many methods are available for text mining and data clustering. Text mining was applied for compression to extract authentic data for analysis of agriculture productivity, persistence to find semantics for what our agrarian says about different varieties of crops. [16]. In a study similar to the one before, semantic analysis was applied to education to improve student achievement. [21].

The Twitter data set is mainly used for finding polarity and opinion among researchers who use hashtag datasets like Nice, Excellent, Good, Bad, and Omg. This symbol was used to emphasize certain words and sentences. [9]. In light of the ever-increasing amount of information, semantic analysis is performed on both electronic and descriptive documents. Using a suitable clustering method can increase the accuracy of classification and preprocessing. In machine learning (ML), there are various types of classification techniques which can be applied to a dataset [20]. Discussions in the agriculture forum address issues related to agriculture like seed certification, fruits, vegetables, wheat, cotton, chemicals, pests, insects, water, soil, and temperature. Using algorithms, scientists examine quantitative datasets, tweets and Facebook pages to determine polarities such as positive, negative, and neural thoughts. [11]. Many scientists rely on semantic analysis for a variety of purposes, but semantics face challenges due to their accuracy and match with polarity [7]. Deep learning is applied to analyze document term matrices based on Term Frequency (TF) and Inverse Document Frequency (IDF). Using Latent Dirichlet Allocation (LDA), one can discover datasets from agriculture tweets that have never been seen before, and by doing so creates original datasets [6]. Sentiment Analysis is an open-ended research area in text extraction. To extract useful information from text, it requires computational methods. Through social media, reviews and comments, has become a really rich source of information about people's behavior (opinion). There have been numerous techniques aimed at analyzing sentiment within text, but none have been able to cope with the complexity of sentiment. For more accurate predictions, new approaches for deep sentiment analysis are needed. [15].

### 3. Material and Methods

Dataset was collected from interviews of agrarian experts. After data acquisition, the audios were passed through speech to text recognition script where textual data was extracted. After data extraction, text was translated into English from the Google API. After getting the translated text, we stored the text in a pre-trained [22] supervised machine learning model named "ALBERT". The model applied pre-processing to the text and removed relevant context, as shown in Figure 1. After pre-processing, Model trained itself based on the given context. Now, the user can give a query in audible medium, that will be converted to text and then it will be processed through the trained model. Depending on the keyword in question entered by the user, this model will validate it from the context that was stored in the model. Then it will classify the relevant answer based on the user's input. As a result, the system was able to provide an expert opinion in the form of the user's native language. This solution enables farmers who lack formal education to understand expert opinions in their native language without any difficulty. This is because textual data will be of no use to user having lack of schooling especially for the uneducated Pakistani community.
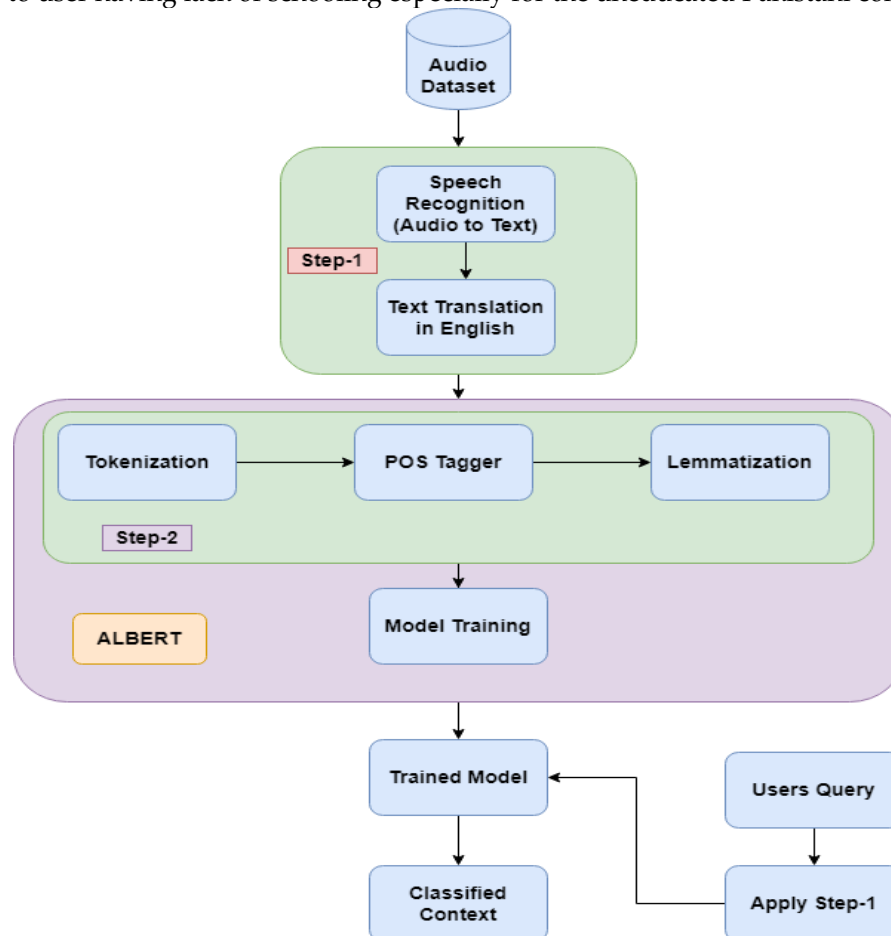


**Figure 1.** Process of Sentiment Extraction from Expert Opinions

3.1 Data Acquistion

Interviews were conducted by different agricultural experts who offered their expert opinions on how farmer can overcome the challenges they face in improving crop productivity. Responses were collected using a questionnaire, as shown in Table 1. Overall 650+ responses were recorded. The audios were converted to text so we could apply sentiments and pre-processing.

**Table 1.** Questionnaire for Data Acquisition

| Sr | English | Urdu |
|----|---------|------|

| 1 | In what ways does soil health influence crop production? | مٹی کی صحت فصل کی پیداواری صلاحیت کو کیسے متاثر کرتی ہے؟ |
| 2 | How does certified seed differ from uncertified seed in terms of crop productivity? | فصل کی پیداوار میں تصدیق شدہ اور غیر تصدیق شدہ بیج کے کیا اثرات ہیں؟ |
| 3 | Is there a way to minimize insect pest problems without harming the environment while increasing agricultural productivity? | ماحولیات کو نقصان پہنچائے بغیر زراعت کی پیداواری صلاحیت کو بڑھانے کے لیے کیڑے مکوڑوں کے مسئلے کو کیسے کم کیا جا سکتا ہے؟ |
| 4 | Is plant disease or insect/pest damage affecting crop production? | پودوں کی بیماری اور کیڑوں/کیڑوں کا فصل کی پیداواری صلاحیت پر کیا اثر ہوتا ہے؟ |
| 5 | Is there a way to overcome the challenges and issues associated with water shortages in agriculture? | ہم زراعت میں پانی کی کمی کے مسائل اور چیلنجز کو کیسے حل کر سکتے ہیں؟ |
| 6 | Farming and agriculture are affected by climate change in what ways? | موسمیاتی تبدیلی زراعت اور کسانوں کو کیسے متاثر کرتی ہے؟ |
| 7 | Is precision agriculture fruitful in improving crop Production? | درست زراعت زراعت اور فصل کی پیداواری صلاحیت کو بہتر بنانے میں کس طرح مددگار ہو سکتی ہے؟ |
| 8 | Can synthetic fertilizer be used in agriculture? What are its pros and cons? | زراعت کے شعبے میں مصنوعی کھاد کے فوائد اور نقصانات کیا ہیں؟ |
| 9 | Are there any ways to reduce the post-harvest losses of crops, fruits, and vegetables? | ہم فصلوں/پھلوں/سبزیوں کے بعد کے نقصانات کو کیسے بہتر کر سکتے ہیں؟ |
| 10 | In what ways does government policy affect the sustainability of agriculture? | حکومت کی پالیسیاں پائیدار زراعت کی نمو پر کیسے اثر انداز ہوتی ہیں؟ |

### 3.2 Features Extraction:

This phase involves determining the accuracy of the machine learning model based on the features and algorithms. Computer scientists have developed a number of techniques to select features, such as a Continuous Bag of Words containing skip Grams (CBOW). Term frequency, N-Grams etc. It is possible to apply or choose among many different methods depending on your dataset and requirements.

We have used ALBERT, a machine learning model that stores the data as context and gets input. As per users' input, the model compares it with the context given and gives the relevant answer. This model has only 12M parameters with an accuracy of 89.4% as per its previous versions like BERT, XLNET, Roberta etc. Table 2 shows that the ALBERT xxlarge configuration has 12 layers with 235M parameters and a hidden size of 4096. It used a much shorter embedding length of 128. However, the biggest parameter savings comes from the column parameter sharing in Table 2.

**Table 2**: Configuration of ALBERT (Pre-Trained ML Model)

| Model | Parameters | Layers | Hidden | Embedding | Parameter-Sharing |
|-------|-----------|--------|--------|-----------|-------------------|
| **Base** | 12M | 12 | 768 | 128 | True |
| **Large** | 18M | 24 | 1024 | 128 | True |

| | | | | | |
|---|---|---|---|---|---|
| **X Large** | 60M | 24 | 2048 | 128 | True |
| **XX Large** | 235M | 12 | 4096 | 128 | True |

In this study, the model is used to generate masked inputs for the MLM targets using n-gram masking [13], with the length of each n-gram mask selected at random. In order to calculate the probability of the length n, by following formula:

$$p(n) = \frac{1/n}{\sum_{k=1}^{N} 1/k}$$

ALBERT has 12 encoder layers and parameters by each encoder are identical, using the same set of weightages as shown in Figure 2. ALBERT Consists of 1 encoder that is applied 12 times to the embeddings. Similar in every layer but 12 times with parameter sharing across all 12 layers. The parameters discussed in Table 1 are set of unique parameters in the model. This has xx-large version, one encoder and the embedding layer. A version of BERT that would be faster to fine tune and scale-up xx-large configuration.
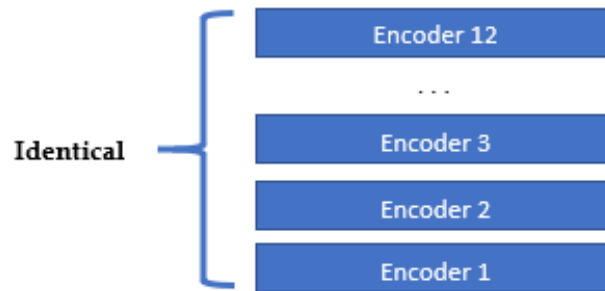


**Figure 2**. Layers of ALBERT (A pre-trained ML Model)

As a result, this ML model is proved to be the key in getting the correct context of user's question that helps in getting the most relevant answer as per the context given.

Beside the output layer, same architecture is utilized in each pre-training and fine-tuning. The identical pretrained model constraints are used for various down streaming tasks. During Fine Tuning phase, all parameters are fine-tuned, separator token is involved for text Question/Answer Separation.

### 4. Results

In Table 3, results are mentioned that were classified from model against the Questions that were passed through the model. Taking farmers as users of the system, they will give there input in audible medium and the audio will be processed by step-1 of the proposed methodology as mentioned in Figure 1. Ultimately, it'll be passed in the model and relevant context will be extracted as mentioned in Table 3.

**Table 3.** Results from the System

| Question Passed in ALBERT | Answer Classified |
|---|---|
| **Effect of Soil in crop productivity?**<br>**فصل کی پیداوار میں مٹی کا اثر؟** | soil is medium for healthy crop growth and crop stand establishment.<br>فصل کی صحت مند نشوونما اور فصل کے اسٹیبلشمنٹ کے لیے مٹی<br>ہے۔ ذریعہ |

**Is there a difference between certified and un-certified seed in terms of crop productivity?**

فصل کی پیداوار میں تصدیق شدہ اور غیر تصدیق شدہ بیج کے اثرات؟

Certified seed results in good crop productivity

تصدیق شدہ بیج کے نتیجے میں فصل کی اچھی پیداوار ہوتی ہے۔

**how to increase Agriculture Productivity?**

زرعی پیداوار کو کیسے بڑھایا جائے۔

herbicides, insecticides, fungicides, nematicides, fertilizers and soil alterations are now being utilized in higher magnitudes than in the past.

جڑی بوٹیوں کی دوائیں، کیڑے مار دوائیں، فنگسائڈز نیماٹائڈس، کھاد اور مٹی میں ترمیم ماضی کے مقابلے اب زیادہ مقدار میں استعمال ہو رہی ہے

**What is impact of plant disease?**

پودوں کی بیماری کا کیا اثر ہے؟

absorbance and translocation of water and nutrients, photosynthesis, flower and fruit development, plant growth and development and cell division and enlargement.

پانی اور غذائی اجزاء کا جذب اور نقل مکانی، فوٹو سنتھیزس پھولوں اور پھلوں کی نشوونما، پودوں کی نشوونما اور خلیوں کی تقسیم اور توسیع

**How we can resolve challenges in Agriculture?**

ہم زراعت میں درپیش چیلنجز کو کیسے حل کر سکتے ہیں؟

by adopting appropriate input and output related polices

متعلق پالیسیوں کو اپنا کر مناسب

**How can precision agriculture affect in improving Agriculture and crop Production?**

درست زراعت زراعت اور فصل کی پیداواری صلاحیت کو بہتر بنانے میں کس طرح مددگار ہو سکتی ہے؟

due to efficient utilization of resources

وسائل کے موثر استعمال کی وجہ سے

**How can resolve water shortage issue?**

پانی کی قلت کا مسئلہ کیسے حل ہو سکتا ہے؟

by introducing such variety of crop that has low water requirement.

پانی کی کم ضرورت والی فصل کی ایسی اقسام متعارف کروا کر

## 5. Discussion

In recent years, scientists have focused on crop inputs such as soil, soil types, soil humidity [3], application of fertilizers [18], crop production, seed, surface temperature, climate change, sustainable development, and chemical spry to evaluate their role in crop productivity. The present research contributed to the analysis of agricultural expert opinions through existing technology. In the absence of prior research, the results consisted of English textual data that could be understood only by an educated person. However, this research can also be used as a multilingual system using English as the standard language and Google Translation API as a translator.

By shifting the text into audio, the person who lacks schooling or cannot read can hear their native language. Farmers are expected to gain expert knowledge from this study, which can be utilized to take safety precautions at early stages and is expected to be beneficial for crop production.
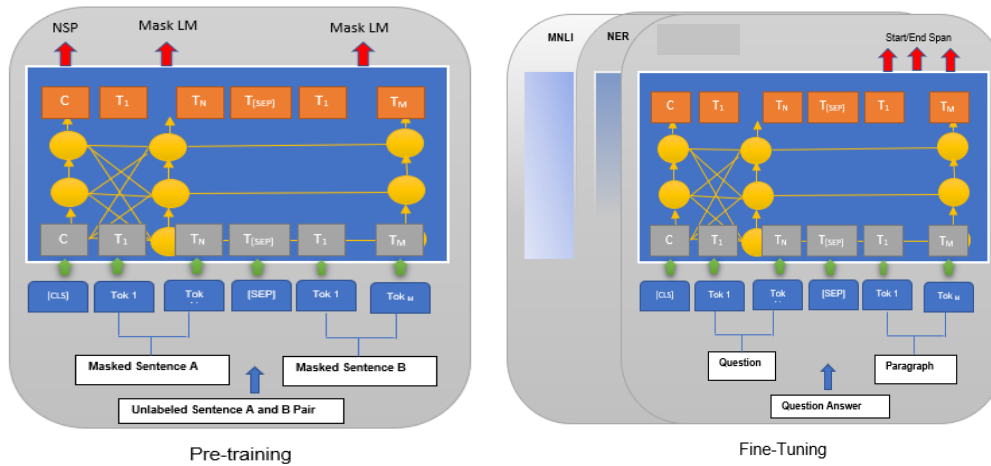
**Figure 3.** Overall Pre-Training and Fine-Tuning Procedure of ALBERT

Pre-training and fine-tuning use the same architecture, besides the output layer. Down streaming tasks are performed with the same pretrained model parameters. During Fine Tuning phase, all parameters are fine-tuned, separator token is involved for text Question/Answer Separation.

## 6. Conclusion

In this study, we have presented a novel approach for collecting and analyzing the expert opinions of agricultural experts. Results show that Certified Seed, Climate change, precision agriculture and Post-Harvest Loss are the most significant factors affecting crop productivity. The use of these types of studies to convey expert opinions through thoughts and emotions to less-educated farmers may prove useful in the future. Agrarian views can be recorded by their body language to obtain effective results. The methodology can be applied to different audio and video recordings from agriculture that are concerned with increasing crop yields. This kind of method can be used for crop yield, polarity estimation, support finding by body motion and poses that helps farmer to gain safety precautions for attaining better crop productivity. However, as far as we know, agriculture experts have reached a consensus on how the technology can be used in naturalistic audio. Therefore, it represents a significant opportunity for industry.

**Conflicts of Interest:** The authors declare no conflict of interest.

**References**

1.  Angiani, G., Ferrari, L., Fontanini, T., Fornacciari, P., Iotti, E., Magliani, F., & Manicardi, S. (2016, September). A Comparison between Preprocessing Techniques for Sentiment Analysis in Twitter. In KDWeb.

2.  Benos, L., Tagarakis, A. C., Dolias, G., Berruto, R., Kateris, D., & Bochtis, D. (2021). Machine learning in agriculture: A comprehensive updated review. Sensors, 21(11), 3758.

3.  Dongare, M. 2020. Smart E-agriculture system using IoT and machine learning. Studies in Indian Place Names 40: 2486–93

4.  Elahi, E., Khalid, Z., Weijun, C., & Zhang, H. (2020). The public policy of agricultural land allotment to agrarians and its impact on crop productivity in Punjab province of Pakistan. Land Use Policy, 90, 104324.

5.  Farooq, M. S., Riaz, S., Abid, A., Umer, T., & Zikria, Y. B. (2020). Role of IoT technology in agriculture: A systematic literature review. Electronics, 9(2), 319.

6.  Fu, R., Qin, B., & Liu, T. (2015). Open-categorical text classification based on multi-LDA models. soft computing, 19, 29-38.

7.  Hussein, D. M. E. D. M. (2018). A survey on sentiment analysis challenges. Journal of King Saud University-Engineering Sciences, 30(4), 330-338.

8.  Karthikeyan, P., Velswamy, K., Harshavardhanan, P., Rajagopal, R., JeyaKrishnan, V., & Velliangiri, S. (2021). Machine learning techniques application: Social media, agriculture, and scheduling in distributed systems. In Research Anthology on Architectures, Frameworks, and Integration Strategies for Distributed and Cloud Computing (pp. 1396-1417). IGI Global.

9.  Kouloumpis, E., Wilson, T., & Moore, J. (2011). Twitter sentiment analysis: The good the bad and the omg!. In Proceedings of the international AAAI conference on web and social media (Vol. 5, No. 1, pp. 538-541).

10. Liu, B., & Zhang, L. (2012). A survey of opinion mining and sentiment analysis. In Mining text data (pp. 415-463). Springer, Boston, MA.

11. Bing, L. (2012). Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, 5(1), 1-167.

12. Martini, D., Mietzsch, E., Schmitz, M., & Kunisch, M. (2011). The agriXchange platform as a means for coordination and support on data exchange in agriculture. E. Gelb & K. Charv t (Eds.), EFITA/WCCA, 11.

13. Joshi, M., Chen, D., Liu, Y., Weld, D. S., Zettlemoyer, L., & Levy, O. (2020). Spanbert: Improving pre-training by representing and predicting spans. Transactions of the Association for Computational Linguistics, 8, 64-77.

14. Mekala, M. S., & Viswanathan, P. (2017, August). A Survey: Smart agriculture IoT with cloud computing. In 2017 international conference on microelectronic devices, circuits and systems (ICMDCS) (pp. 1-7). IEEE.

15. Razzaq, A., Asim, M., Ali, Z., Qadri, S., Mumtaz, I., Khan, D. M., & Niaz, Q. (2019). Text sentiment analysis using frequency-based vigorous features. China Communications, 16(12), 145-153.

16. Rehman, M., Razzaq, A., Baig, I. A., Jabeen, J., Tahir, M. H. N., Ahmed, U. I., ... & Abbas, T. (2022). Semantics analysis of agricultural experts' opinions for crop productivity through machine learning. Applied Artificial Intelligence, 36(1), 2012055.

17. Ryoo, M. S., & Aggarwal, J. K. (2009). Semantic representation and recognition of continued and recursive human activities. International journal of computer vision, 82, 1-24.

18. Saikai, Y., Patel, V., & Mitchell, P. D. (2020). Machine learning for optimizing complex site-specific management. Computers and Electronics in Agriculture, 174, 105381.

19. Smith, M. J. (2018). Getting value from artificial intelligence in agriculture. Animal Production Science, 60(1), 46-54.

20. Vaghela, V. B., Jadav, B. M., & Scholar, M. E. (2016). Analysis of various sentiment classification techniques. International journal of Computer applications, 140(3), 975-8887.

21. Vo, H. T., Lam, H. C., Nguyen, D. D., & Tuong, N. H. (2016). Topic classification and sentiment analysis for Vietnamese education survey system. Asian Journal of Computer Science and Information Technology, 6(3), 27-34.

22. Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., & Soricut, R. (2019). Albert: A lite bert for self-supervised learning of language representations. arXiv preprint arXiv:1909.11942.

23. Tian, S., Zhu, B., Yin, R., Wang, M., Jiang, Y., Zhang, C., ... & Liu, M. (2022). Organic fertilization promotes crop productivity through changes in soil aggregation. Soil Biology and Biochemistry, 165, 108533.