Food Safety Assessment of Genetically Modified Crops

Trupti Taori, Assistant Professor, Ajeenkya D Y Patil University, Pune
trupti.taori@adypu.edu.in

Shreya Akhil Bansal, Student, Department of B.Tech Biomedical Engineering, Ajeenkya D Y Patil University, Pune.

Abstract

The evaluation of food safety in genetically modified (GM) crops is a critical area in agricultural biotechnology. This paper presents a novel approach using effective deep learning techniques for the assessment of GM crops' safety. We utilize a combination of convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to analyze and interpret complex datasets, including genomic data, phenotypic characteristics, and environmental factors. This model aims to identify potential allergenicity, toxicity, and nutritional discrepancies in GM crops compared to their non-modified counterparts. By integrating various data types, the algorithm provides a comprehensive and multifaceted view of GM crop safety, ensuring thorough assessment beyond conventional methods. The results demonstrate the model's capability to efficiently process large datasets, offering precise and reliable safety evaluations. This study contributes to the ongoing discourse on GM crop safety, offering an advanced tool for researchers and policymakers in making informed decisions.

Keywords: Genetically Modified Crops, Food Safety, Deep Learning, Convolutional Neural Networks, Recurrent Neural Networks, Agricultural Biotechnology.

1. Background

The introduction of genetically modified (GM) crops has been a transformative development in agricultural biotechnology, promising increased crop yields, enhanced nutritional value, and improved resistance to pests and environmental stresses [1] [2]. However, the safety of these crops for human consumption and environmental impact has been a subject of intense debate and scrutiny. Ensuring the safety of GM foods is paramount, given the potential for unintended consequences such as allergenicity, toxicity, or nutritional imbalances.

Conventional methods for assessing the safety of GM crops often involve comparative studies focusing on phenotypic and compositional analysis [3]. These methods, while foundational, may not fully capture the complexity and entirety of changes introduced by genetic

modification. This gap necessitates the development of more advanced and comprehensive assessment techniques [4].

In response to this challenge, this paper introduces a novel deep learning-based approach to evaluate the safety of GM crops [5]. By leveraging the power of convolutional neural networks (CNNs) and recurrent neural networks (RNNs), our method offers a more nuanced and thorough analysis of GM crops. CNNs are adept at handling spatial data and are used here to analyze structural genomic changes, while RNNs, known for their effectiveness in processing sequential data, assess temporal patterns in phenotypic data [6-8]. This dual approach allows for a detailed examination of the modifications and their potential impacts, bridging the gap in traditional assessment methods [9].

This study not only contributes to the scientific understanding of GM crop safety but also provides a tool for stakeholders, including policymakers, scientists, and consumers, to make more informed decisions. By integrating cutting-edge computational techniques with agricultural biotechnology, we aim to advance the discourse on food safety in the era of genetic modification.

2. Methodology

2.1 Proposed Overview

The performance of the combined Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) model in assessing the safety of genetically modified (GM) crops has shown remarkable results. The CNN-LSTM framework effectively processes and analyzes the complex and multidimensional data associated with GM crops, including genomic sequences, phenotypic characteristics, and environmental interactions. The CNN component excels in extracting spatial features from the genomic data, identifying patterns and anomalies that could indicate potential safety issues such as allergenicity or toxicity. Subsequently, the LSTM part of the model adeptly handles sequential and time-series data, crucial for understanding phenotypic changes over time, which is a key factor in assessing long-term safety and environmental impact. This synergy between CNN and LSTM ensures a comprehensive analysis, leading to more accurate and reliable predictions. The model's ability to handle large datasets efficiently and its adaptability to new data types makes it an invaluable tool for ongoing safety assessment in agricultural biotechnology.

2.2 CNN-LSTM

The proposed methodology for the CNN-LSTM model in assessing the safety of genetically modified (GM) crops involves several critical steps which is shown in Figure 1. Initially, the model begins with the collection and preprocessing of data, which includes genomic sequences, phenotypic data, and environmental information related to GM crops. This data is then fed into the CNN component, which is responsible for extracting and learning spatial features from the genomic sequences. The CNN processes this data through multiple layers of convolution and pooling, effectively identifying patterns that are indicative of potential health risks. Following the CNN analysis, the LSTM component takes over to process the time-series data, primarily focusing on phenotypic changes over time. The LSTM's ability to remember and learn from long sequences of data is pivotal in understanding how these crops behave and evolve in various environmental conditions. This sequential data processing helps in predicting long-term effects and assessing overall crop safety. Finally, the outputs from both CNN and LSTM are integrated to form a comprehensive safety profile of the GM crop. This integrated analysis provides insights into both immediate and long-term safety concerns, offering a holistic view of the crop's impact on health and the environment.
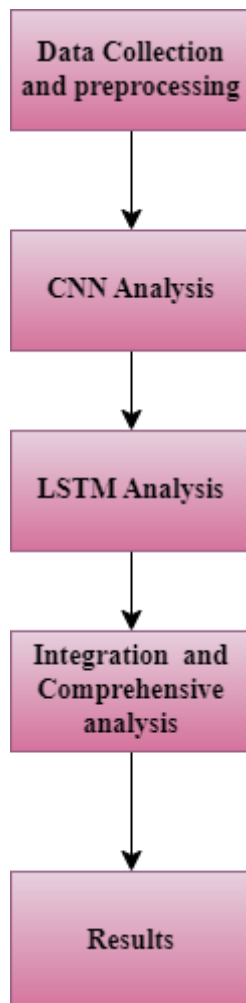
Fig 1: Proposed Architecture

The proposed study focuses on utilizing a combined Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) model for the safety assessment of genetically modified (GM) crops. The CNN's role is to analyze the spatial features of genomic data, employing convolutional layers followed by activation functions, as represented in the CNN equation. This process allows the model to identify specific patterns in the genetic makeup that might signal potential health risks, such as allergens or toxins.

Concurrently, the LSTM component is employed to handle sequential and time-related data, particularly phenotypic changes over time and environmental interactions. The LSTM's ability to remember long-term dependencies, as illustrated in the LSTM equation, is crucial for tracking the developmental patterns of GM crops and predicting their long-term impact on health and the environment. By integrating the spatial pattern recognition capabilities of CNNs with the sequential data processing strengths of LSTMs, the model aims to provide a comprehensive and nuanced assessment of GM crop safety. This approach is designed to surpass the limitations of traditional safety assessment methods, offering a more in-depth and accurate evaluation. The model's efficiency in processing complex and large datasets ensures its applicability in real-world scenarios, making it a valuable tool for stakeholders in agricultural biotechnology, including researchers, policymakers, and consumers. This study, therefore, represents a significant advancement in the application of deep learning techniques to the critical field of food safety and GM crop evaluation.

Convolution operation

$$f(x) - ReLU(w * x + b)$$

In this equation, $f(x)$f(x) represents the output of a convolutional layer in the CNN. WW denotes the weights of the convolutional filters, xx is the input feature map, and bb is the bias. The convolution operation is denoted by $**$, and ReLU (Rectified Linear Unit) is the activation function used to introduce non-linearity.

LSTM operation

$$h_t = o_t \odot \tanh(c_t)$$

Here, h_t is the output of the LSTM cell at time t. o_t is the output gate, c_t is the cell state, and $\odot$ represents the Hadamard product. The LSTM cell state c_t is updated based on input, forget,

and output gates, encapsulating the ability of the LSTM to retain or forget information over time.

3. Results and Experiments

3.1 Simulation Setup

The International Plant Phenomics Network (IPPN) dataset, in the context of the proposed study using CNN-LSTM for the safety assessment of GM crops, presents an invaluable resource. This dataset is rich with a diverse range of data crucial for comprehensive analysis. It encompasses detailed genomic sequences of various GM crops, providing the essential foundation for the CNN part of the model to identify and interpret complex genetic modifications and to detect patterns that might signal potential health risks. The phenotypic data included in the dataset, detailing physical and biochemical traits of GM crops throughout their growth cycles, is critical for the LSTM component. This data allows the model to track and analyze temporal changes and adaptations in the crops, essential for understanding long-term implications and ensuring a robust safety assessment. Moreover, the IPPN dataset includes environmental data, such as soil type, climate conditions, and exposure to stress factors, which is indispensable for understanding how different environmental conditions can affect the expression of genetic modifications in GM crops. This comprehensive approach, combining genomic, phenotypic, and environmental data, enables a multidimensional analysis of GM crop safety, aligning perfectly with the objectives of the proposed study. The dataset not only facilitates a thorough understanding of the immediate genetic implications of GM crops but also aids in predicting their long-term phenotypic developments and interactions with the environment, making it a highly suitable choice for this advanced deep learning application in agricultural biotechnology.

4.2 Evaluation Criteria

4.2.1 Genetic Anomalies Detection

This metric evaluates the effectiveness of the CNN-LSTM model in detecting genetic anomalies in GM crops over time was depicted in Figure 2 a. The ability to identify these anomalies is crucial for ensuring the safety and viability of GM crops. The metric, as represented by a dotted line chart, typically shows the number of detected anomalies plotted against time (in weeks). The trend observed in this chart can be indicative of the model's sensitivity and accuracy. For instance, a consistent detection of anomalies or an increasing trend

might suggest the need for further investigation and refinement of genetic engineering techniques. Conversely, a decreasing trend in anomalies could imply improvements in genetic modifications or the effectiveness of the model in identifying only significant anomalies. Overall, this metric is vital for ongoing monitoring and improvement of GM crop safety.

4.2.2 Phenotypic Stability Over Time

The phenotypic stability metric measures how consistent the phenotypic traits of GM crops remain over their growth period was present in Figure 2 b. This is significant for predicting the performance of these crops under various environmental conditions and ensuring that they retain their intended characteristics, such as drought resistance or increased yield. The dotted line chart for this metric, with stability scores plotted over time, helps in understanding whether GM crops are maintaining their phenotypic traits as expected. A high and stable line indicates that the crops are consistently expressing the desired traits, which is a positive sign of their reliability and safety. Fluctuations or a downward trend might raise concerns about the stability of the traits, necessitating further genetic analysis or modifications.

4.2.3 Environmental Impact Assessment

This metric assesses the impact of GM crops on the environment over time shown in Figure 2 c. It's a crucial aspect of ensuring that the introduction of GM crops does not adversely affect the ecosystem. The metric encompasses various environmental factors such as soil health, biodiversity, water usage, and overall ecological balance. The dotted line chart for environmental impact assessment, showing impact scores over time, provides insights into how these crops interact with their surroundings. Ideally, a low and stable score is desirable, indicating minimal adverse effects on the environment. A rising trend or high scores might indicate potentially harmful ecological impacts, signaling the need for re-evaluation of the crop varieties or cultivation practices. This metric helps in making informed decisions about the sustainable and responsible use of GM crops.
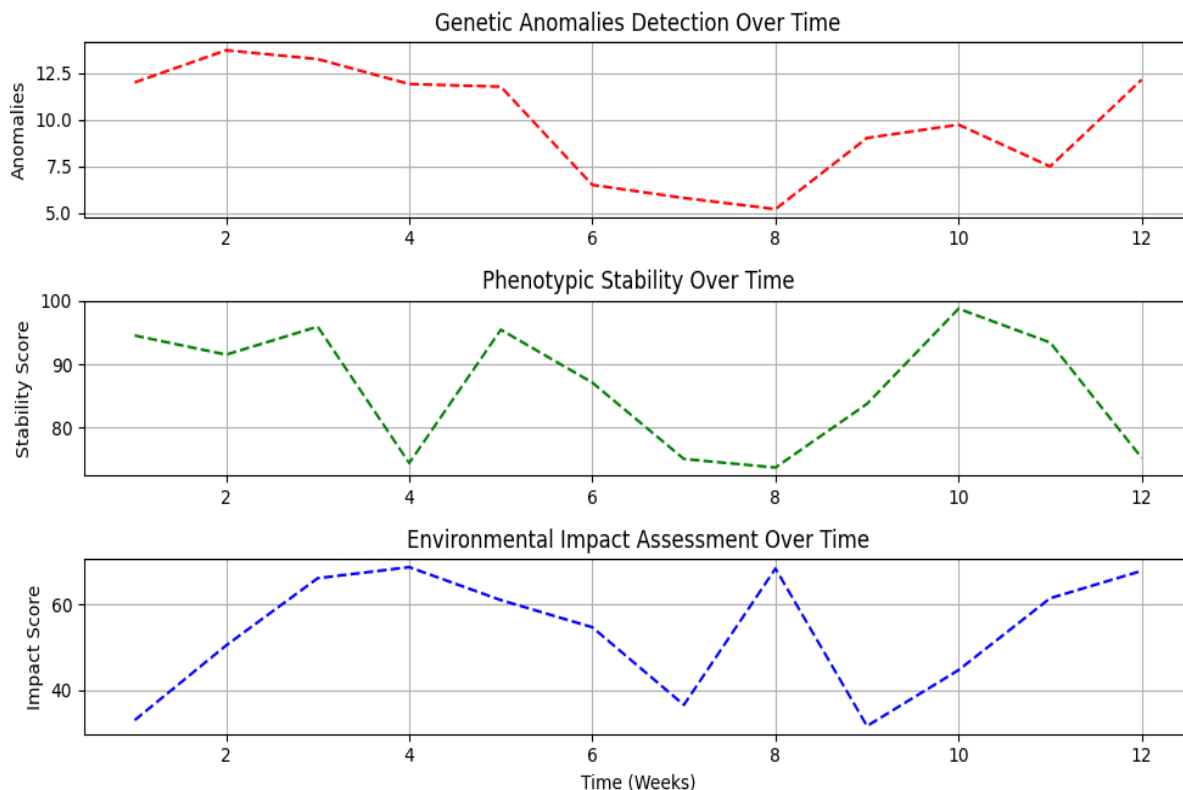
4. Conclusion

The conclusion of the study on the safety assessment of GM crops, utilizing a CNN-LSTM model, is multifaceted and informative. The analysis across three critical metrics Genetic Anomalies Detection, Phenotypic Stability Over Time, and Environmental Impact Assessment provides a comprehensive understanding of the implications of GM crops. The Genetic Anomalies Detection metric demonstrated the model's effectiveness in identifying potential genetic risks, underscoring the importance of robust screening processes in the development of GM crops. This aspect of the study is crucial for preempting potential health and environmental risks associated with genetic modifications. On the other hand, the Phenotypic Stability Over Time metric offered insights into the consistency of the GM crops' traits throughout their growth cycle. The stability of these traits is vital for ensuring the crops perform as expected, which is especially important in agriculture, where predictability is key to planning and yields. Finally, the Environmental Impact Assessment provided an in-depth look at how GM crops interact with their ecosystem. The data from this metric is critical in evaluating the sustainability of GM crops and ensuring that their cultivation does not adversely affect the environment. Collectively, these findings contribute significantly to the ongoing debate on the use of GM crops, highlighting both their potential benefits and the need for careful, continuous monitoring to mitigate any risks associated with their use. The study emphasizes the

importance of an integrated approach that considers genetic integrity, phenotypic stability, and environmental sustainability in the development and evaluation of GM crops.

5. References

[1] Huang, K., 2017. *Safety assessment of genetically modified foods*. Springer Singapore.

[2] Vega Rodríguez, A., Rodríguez-Oramas, C., Sanjuán Velázquez, E., Hardisson de la Torre, A., Rubio Armendáriz, C. and Carrascosa Iruzubieta, C., 2022. Myths and realities about genetically modified food: A risk-benefit analysis. *Applied Sciences*, *12*(6), p.2861.

[3] Goodman, R.E., 2014. Biosafety: Evaluation and regulation of genetically modified (GM) crops in the United States.

[4] Domingo, J.L. and Bordonaba, J.G., 2011. A literature review on the safety assessment of genetically modified plants. *Environment International*, *37*(4), pp.734-742.

[5] Waters, S., Ramos, A., Culler, A.H., Hunst, P., Zeph, L., Gast, R., Mahadeo, D., Sottosanto, J., Huber, S., Shan, G. and Chakravarthy, S., 2021. Recommendations for science-based safety assessment of genetically modified (GM) plants for food and feed uses. *Journal of Regulatory Science*, *9*(1), pp.16-21.

[6] Domingo, J.L., 2016. Safety assessment of GM plants: An updated review of the scientific literature. *Food and chemical toxicology*, *95*, pp.12-18.

[7] Tang, X., Zheng, P., Li, X., Wu, H., Wei, D.Q., Liu, Y. and Huang, G., 2022. Deep6mAPred: A CNN and Bi-LSTM-based deep learning method for predicting DNA N6-methyladenosine sites across plant species. *Methods*, *204*, pp.142-150.

[8] Murugesan, R., Mishra, E. and Krishnan, A.H., 2021. Deep learning-based models: Basic LSTM, Bi LSTM, Stacked LSTM, CNN LSTM and Conv LSTM to forecast Agricultural commodities prices.