Harvesting Automation: Cobots and Generative AI in Real-Time Crop Monitoring and Yield Optimization

Vignesh Vane
Dept. of ECS,
Pillai HOC College of
Engineering and Technology,
Rasayani Maharashtra, India
Email: vigneshvane400@gmail.com

Nayan Patil
Dept. of ECS,
Pillai HOC College of
Engineering and Technology,
Rasayani Maharashtra, India
Email: nayanpatil813@gmail.com

Mansi Subhedar
Dept. of ECS,
Pillai HOC College of
Engineering and Technology,
Rasayani Maharashtra, India
Email: mansissubhedar@gmail.com

***Abstract*—Agriculture is changing significantly with the adoption of artificial intelligence and automation to address growing food needs and mitigate the lack of trained personnel. However, conventional hardware-based automation systems are expensive and not affordable for small and medium-sized holdings, restricting their mass adoption. This article introduces SimAgro, a software-based platform intended to solve this problem by integrating cobot simulation with generative AI for real-time crop monitoring and yield optimization. The system utilizes Random Forest models that have been trained on multimodal agricultural data such as NDVI, soil moisture, rainfall, pest population, and fertilizer application, and a virtual cobot advisor gives actionable advice. A cloud-hosted dashboard developed using Streamlit enables interactive monitoring and decision support. Experimental testing proved 98.84% accuracy in crop health classification and an R^2 value of 0.9719 for yield prediction. The originality of this research is in presenting a software-based, affordable, and scalable alternative to hardware-based approaches, promoting authentic precision farming and facilitating future integration with real farm IoT platforms.**

***Index Terms*—Cobots Simulation, Generative AI, Real-Time Crop Monitoring, Yield Prediction, Agricultural Data Analysis, Smart Farming Software, AI in Precision Agriculture, Automation in Agriculture**

1. INTRODUCTION

Artificial intelligence (AI) and automation in agriculture sector are transforming farming techniques. As the global population is increasing and expected to reach over 9.7 billion by 2050, agriculture must produce more food, improve quality, and do so sustainably. Smart farming technologies help farmers make better decision, efficiently increase the crop yields, and use resources effectively. Conventional farming practices are heavily depended on manual labor and experience-based decision making, often insufficient in meeting these demands. The integration of Collaborative Robots (Cobots) and Generative AI in farming processes presents an innovative approach for managing these challenges by enabling efficient, scalable and data-driven agricultural management [1], [2], [3].

In recent years, Cobots are becoming very popular in farming because they can work safely and efficiently with people. They help with tasks like picking fruits, trimming plants and sorting produce, making work easier and faster [4]. Unlike traditional industrial robots, cobots are designed for flexibility, ease of deployment, and easy adaptability to different crop types and farming environments [5]. Computer vision and Machine learning have made cobots smarter. Cobots can see, recognize the crops better, and handle crops more accurately, helping farmers with tasks like sorting and harvesting [2], [6]. However, deploying cobots in real-world agricultural fields often requires significant investment in hardware, maintenance and field calibration, which can limit affordability for small and medium scale farms. To address this challenge, software-based simulation platforms have emerged as a low-cost, scalable alternative for testing and optimizing AI-driven cobot systems before physical deployment [7]. Normalized Difference Vegetation Index (NDVI) and Leaf Area Index (LAI) are used to train predictive models capable of assessing crop under varying environmental conditions [8], [9].

Generative AI can create realistic crop growth examples and add more data to help AI models learn better and make more accurate predictions [6], [10]. When combined with predictive algorithms such as Random Forest classifiers [11], [12] and implemented through interactive platforms like Streamlit [13], these systems can deliver actionable observations directly to farmers via consumer devices such as smartphones, tablets and personal computers. This shift toward software-based farming tools fits well with the rise of consumer technology in agriculture, like mobile apps, cloud dashboards and smart advisory systems [14], [15].

This manuscript introduces a real-time cobot simulation and generative AI-based framework for crop monitoring and yield optimization. Unlike previous research that mainly focuses on hardware-based solutions [1], [4], [7], our approach is an entire virtual system, enabling early testing, system scaling, and iterative AI model improvement without costly field operations.

2. LITERATURE SURVEY

Automation in agriculture is a major transformation in recent years, driven by the integration of robotics, artificial intelligence (AI), and data-driven decision making [1]. The transition from manual labor to intelligent and automated systems is fueled by advancements in collaborative robots (cobots), machine vision, and real-time analytics, enabling precision agriculture to meet the growing demand for food production.

2.1. Traditional and Automated Harvesting Methods

Traditionally, crop harvesting has relied strongly on manual labor, which is time-consuming and subject to seasonal shortages of labors [1]. Machines combining agricultural robots and automated threshers have changed large-scale farming. They work faster and reduce the need for manual labor, making farming more efficient [2]. However, such machinery often lacks adjustability to delicate crops, leading to product damage and inefficiencies in mixed-crop environments.

Recent developments in machine vision-assisted harvesters improving selectively and precision in harvesting operations [3], [6]. For instance, tomato-picking robots that use color to check ripeness have improved accuracy, but they still struggle with blocked views and changing light conditions. These limitations emphasize the need for more advanced AI-driven solutions.

2.2. Cobot Integration in Agriculture

Collaborative robots (cobots) are a flexible solution for tasks requiring adjustability and human-machine cooperation [4], [5]. Unlike traditional industrial robots, cobots are designed to work safely, efficiently and with easy adaptability to different crop types alongside humans, with force-limiting joints and intelligent control systems. In agriculture, cobots are used for fruit picking, pruning, and greenhouse management which enables more precise handling and reducing crop damage [5].

Research in [7] demonstrated that cobots equipped with technology could accurately identify and harvest strawberries with minimum waste. Another study [16] implemented real-time cobot path planning to navigate between rows of crops independently, significantly reducing operational disruptions.

2.3. AI and Generative AI for Crop Monitoring

The integration of AI in agriculture has allowed AI-based forecasting, real-time monitoring, and intelligent decision making [10]. Disease detection, yield estimation and crop maturity assessment are successfully performed by deep learning models using multispectral and RGB imaging.

Precision agriculture has started to be transformed by generative AI, which is capable of creating synthetic data and generating predictive models. Research in [10] shows that synthetic crop images could be created by GANs for training AI models, improving performance in low-data scenarios. LLMs integrated with IoT data streams were applied by the study from [11] to provide intelligent agricultural recommendations, showing improvements in resource allocation and pest management strategies.

2.4. IoT and Sensor-Based Yield Optimization Systems

Smart farming IoT has been integral with IoT, which offers continuous environmental monitoring through temperature, humidity, soil moisture, and chlorophyll content sensors [12]. These sensor networks enable real-time monitoring for decision-making systems, reducing water and fertilizer wastage [8].

Integration of IoT with distributed computing has enhanced responsiveness and intelligent control in agricultural devices [13]. For instance, in [15], a network of soil nutrient sensors connected to a distributed AI system allows dynamic adaptability of irrigation and fertilization schedules, resulting in 23% increase in yield efficiency.

Combining IoT with cobots and generative AI provides a collaborative platform where sensor-driven data informs AI models, and cobots execute optimized harvesting operations in real time. Foundation of the system proposed in this paper is formed by such integration.

3. Proposed Methodology

The proposed methodology focuses on the advancement of a software and integration of collaborative robots (cobots) and Generative Artificial Intelligence (AI) within the software simulation to optimize real-time crop monitoring and yield prediction.The workflow of the software is divided into several functional stages.Figure 1 represents an overview of the methodology

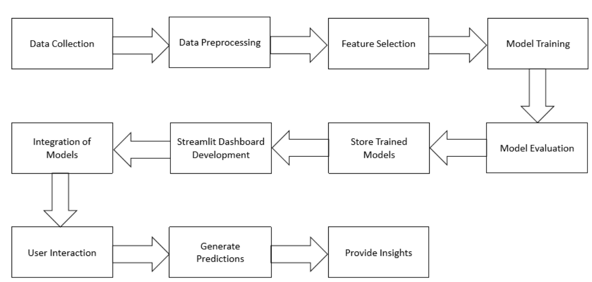

Figure 1. System Overview

3.1. System Architecture Design

The architecture consists of three layers:
1. Simulation Environment Layer:- This layer models the agricultural field ,patterns of plant growth and environmental entities such as temperature, humidity and soil moisture through physical and data-driven simulation tools [1], [2].
2. Perception Layer:- Multi-spectral imaging, thermal cameras and feeding the raw data into the AI pipeline is simulated by virtual sensors
3. Decision and Control Layer:- The system runs the AI models for plant health classification ,task scheduling for the cobots and yield prediction within a simulated farming setup

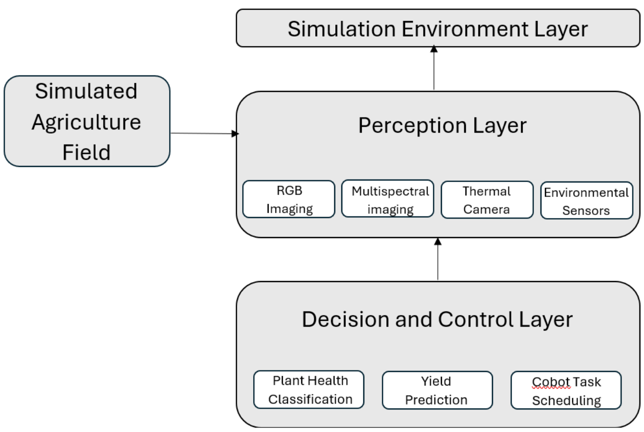

Figure 2. System Architecture

3.2. Data Simulation and Preprocessing

3.2.1. Data Simulation.

- Environmental parameters are generated using statistical models and synthetic data generators which includes temperature, humidity, soil moisture and sunlight intensity.
- Crop image datasets are collected from publicly available source PlantVillage and Kaggle agricultural datasets [16].
- Virtual sensor models are created in the simulation that reflects the real cobot sensors.
- These data streams represents the inputs that would be collected in a real farm

3.2.2. Data Preprocessing. To make the simulated data usable for the AI-driven monitoring models following techniques are used:

- Noise Filtering: Random variations in data are smoothened using various filters.
- Normalization and scaling: Environmental parameters are scaled between 0-1 for uniform input to the AI models [12].
- Feature Extraction:
  - From images- plant leaf texture, color and shape features for identifying the crop health.
  - From environmental data- Temperature/humidity correlation with growing stages is modelled.
- Data Augmentation(for images):The datasets are expanded to improve the robustness of the AI models by the techniques such as rotation, scaling, flipping and color enhancement [8].

3.3. Virtual Cobot Advisor Framework

The Virtual Cobot Advisor is an smart assistant in real-time crop monitoring that suggests when and what agricultural operations(like irrigation, fertilization and harvesting) should be performed in order to maintain and improve the crop health.

3.3.1. Advisory Simulation Environment.

- The framework is developed in the simulation environment which combines ML models, crop growth datasets and weather condition simulators.
- The system outputs recommendations such as irrigation required within 24 hours, Harvesting optimal next week, Lack of Fertilizers.

3.3.2. Data Inputs for Cobot Advisor. The advisor relies on following data sources:

- Synthetic Crop Growth data: Simulated plant images and metrics.
- Environmental Data: Temperature, soil moisture, weather values from the datasets
- Historical Analysis: Previous yields and crop cycle records used to increase prediction accuracy.

3.3.3. Advisory Functions.

- Pest/Disease Alert: Provides warnings if the simulation detects patterns similar to that of crop stress or pest outbreaks.
- Irrigation scheduling: Suggests optimal irrigation time based on the simulated input data of soil and weather.
- Harvest Prediction: Determines when the crops are ready to harvest for optimizing yield.
- Fertilizer recommendation: Handles the nutrient management depending on the plant growth stages.

3.3.4. Human-System Interaction.

- The advisor output is displayed on the User Interface so the farmers or researchers can accept, reject or reschedule the advised operation
- This simulates how cobots could assist human decision making

3.4. Generative AI For Real-Time Crop Monitoring

3.4.1. Role of Generative AI.

- Data Augmentation: Generative AI is useful in improving the robustness of the classifier usually by creating variations of existing crop images resulting in expansion of datasets.
- Predictive Simulation: AI models generates future crop states such as drought stress conditions ,helping farmers take early and responsive actions.

3.4.2. Technical Approach.

- Generative Model Selection: Realistic plant states conditioned on environmental variables are generated using conditional GAN(Generative Adversarial Networks).
- Integration with CNN: The accuracy in detecting crop health issues is increased by implementing CNN-based classification models that uses generated images along with the real datasets.

3.4.3. Generative AI outcome.

- Monitoring Dashboard: Displays both real and generated images for crop condition tracking.
- Risk Prediction: Shows the areas of a field likely to face stress.

3.4.4. Features of Generative AI.

- Generative AI eliminates dependency on large-scale datasets, allows safe experiments through synthetic simulations.
- It helps improve the accuracy of crop health classification due to larger and varied training datasets.
- The simulation is highly useful that helps predict and prevent crop health degradation before it occurs.

3.4.5. Simulation Environment. The simulation environment and User Interface are the backbone of the proposed system. These components allow researchers or farmers to visualize real-time crop conditions and interact with the virtual advisor to take necessary precautions.

- A replica of a farm is created using datasets and generative AI. It contains crop rows, soil conditions, environmental parameters and potential stress.
- The algorithm simulates different growth stages (such as seeding, vegetative, flowering, harvesting) which enables the user to monitor how environmental changes affect the crops.
- Various virtual conditions such as pest infestation, over-fertilization and disease outbreaks are modelled to evaluate and test AI decision-making under diverse conditions.

4. Result and Discussion

This section represents the experimental results of the proposed system considering the performance of the models for yield prediction and crop health classification, along with their effectiveness. The model performance is evaluated using metrics such as accuracy, RMSE, R^2 score along with the feature analysis.

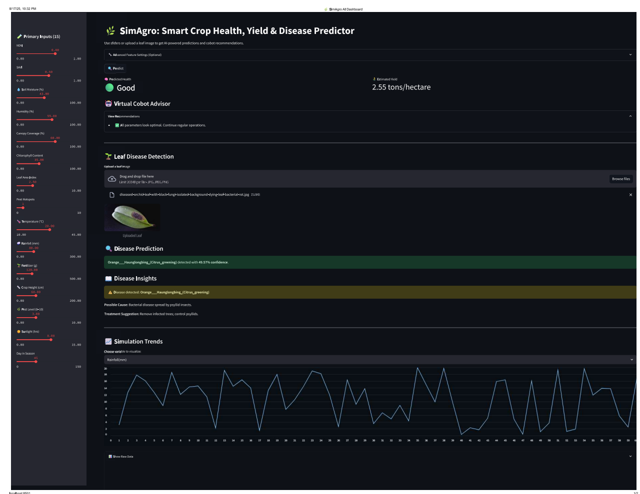

Figure 3. Visual Dashboard

The figure 5 represents an interactive dashboard for analyzing the crop health and yield classification. The user can modify the input parameters and check the models output instantly.

4.1. Yield Prediction

The Random Forest Regressor obtained an R^2 score of 0.9719 which indicates that the model is able to capture more than 97% of the variations in the crop yield. This illustrates that the predictions fits very well with the actual yield values, with very small proportion of undescribed variance due to the noise or the external factors. The model has also obtained a Root Mean Square Error (RMSE) of 0.2844 tons/hectare, which represents the average deviation between the observed and predicted yields. These results indicate that the model provides strong predictive accuracy and data explanation ability across different crops and environmental conditions.

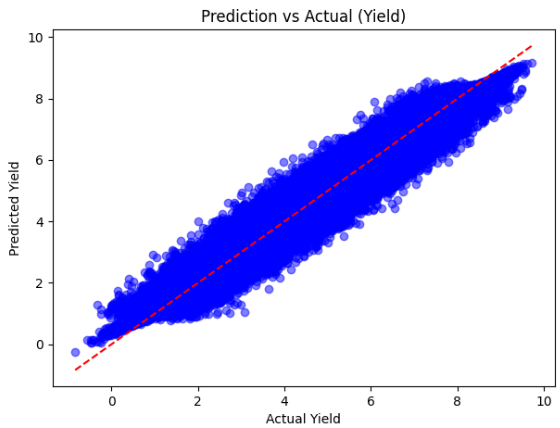

Figure 4. Predicted vs Actual Yield

The Scatter Plot shows the high accuracy of the regression model, which displays a strong alignment of the actual and predicted yield values along the diagonal line.

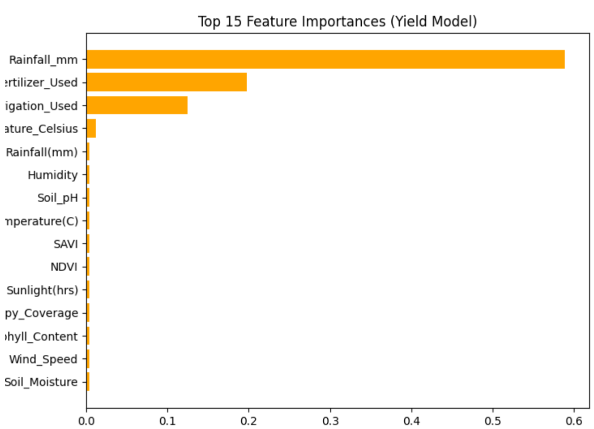

Figure 5. Yield Feature Importance

The most important factors influencing yield are irrigation, fertilizer application and rainfall, according to feature importance analysis. This improves the model's transparency by being consistent with agronomic knowledge.

4.2. Feature Corelation and System Insights

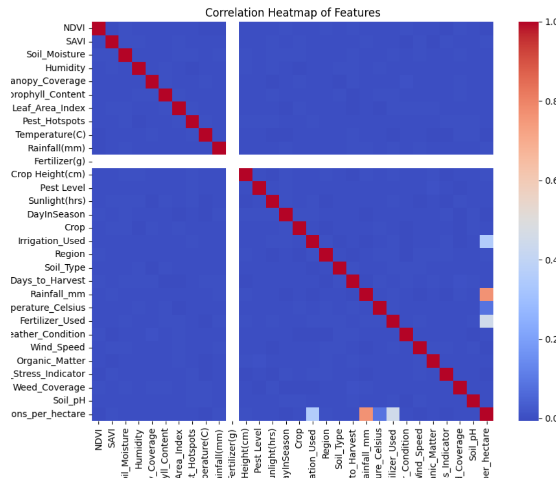

Figure 6. Correlation Heatmap

The heatmap shows that there are strong links between rainfall, soil moisture and NDVI. These three features are very important for predicting both health and yield. This shows that the system is good at finding important agronomic drivers.

4.3. Comparison with Prior Work

Table 1 provides a comparative overview of prior work in farming automation, cobot technology, and machine learning–based agricultural solutions. Earlier studies such as Patel and Desai [1] and Liu et al. [2] were primarily surveys and reviews, offering useful background insights but lacking experimentally validated results. Other contributions, such as Singh et al. [4] and Escobar and Sanchez [7], proposed cobot-assisted harvesting prototypes that demonstrated feasibility, but their scalability and adaptability to varied farm conditions remained limited. IoT-based solutions, such as Patel et al. [15], improved monitoring and resource management efficiency but did not integrate advanced predictive modeling or virtual cobot simulation.

The work proposed here builds upon these task-oriented approaches by integrating multiple capabilities into a unified framework. Unlike prior studies that addressed individual components in isolation, our system combines Random Forest models for crop health classification and yield prediction, Generative methods for synthetic data augmentation, and a Virtual Cobot Advisor with an interactive Streamlit dashboard. This synergy delivers high crop health classification accuracy (98.84%) and robust yield prediction performance (R^2=0.9719), while also providing a scalable, software-based alternative to costly hardware-centric implementations.

TABLE 1. COMPARISON WITH PRIOR WORK IN AGRICULTURE

Ref.	Dataset(s)	Focus / Technique	Performance	Limitation
[16]	PlantVillage, custom leaf RGB	Transfer Learning + GAN for disease recognition	∼98% accuracy	Focused only on disease detection, no yield prediction or advisory
[4]	Leaf datasets (texture, color)	AI-based disease detection	∼95% accuracy	Limited to specific diseases, not scalable to multi-feature yield estimation
[15]	IoT sensor datasets (soil nutrients, weather, moisture)	Smart dashboard + ML prediction	Yield efficiency ↑ 23%	IoT-based, lacked advanced predictive models and cobot integration
Proposed Work	PlantVillage, Kaggle + synthetic data	Random Forest (health classification, yield regression), GAN augmentation, Virtual Cobot Advisor, Streamlit dashboard	Health: **98.84%**, Yield: $R^2 = 0.9719$	Software-only simulation (future work: real IoT farm integration)

4.4. Task Performance Comparison

Results from other tasks were added like using a ResNet18 CNN on the PlantVillage dataset to find diseases.

TABLE 2. MODEL PERFORMANCE ACROSS TASKS

Task	Crop Health	Yield Prediction	Disease Detection
Algorithm	Random Forest Classifier	Random Forest Regressor	ResNet18 CNN
Dataset	Agronomic dataset	Agronomic dataset	PlantVillage (Kaggle)
Accuracy/R^2	98.84%	R^2=0.9719, RMSE=0.28	98%

This table shows that all parts of the worked very well.

4.5. Crop Health Classification Performance

Random Forest Classifier had a total accuracy of 98.84% in classifying crop health. Class-wise performance indicated that the Average category attained precision of 1.00, recall of 0.80, and F1-score of 0.89, while the Good category attained precision of 0.99, recall of 1.00, and F1-score of 0.99. The findings testify to the high capability of the model in distinguishing crop health classes with high reliability.

5. CONCLUSION

Overall, this research showed the efficacy of SimAgro, a software platform incorporating cobot simulation and generative AI for real-time crop health determination and yield forecasting. The system attained 98.84% accuracy in crop health classification and an R^2 value of 0.9719 in yield forecasting, illustrating its promise as a dependable agricultural decision-support system. In contrast to hardware-demanding approaches, SimAgro is a scalable, low-cost solution deployable on consumer hardware, which brings precision farming within reach. Subsequent research will emphasize incorporating real farm IoT data for field evaluation, improving the virtual cobot advisor with adaptive planning and scheduling, and using advanced generative models for pest and disease prediction. In addition, full-scale simulations over various crop types will be investigated to enhance robustness and generalizability, bringing SimAgro nearer to practical application in precision agriculture.

References

[1] J. A. Patel and M. K. Desai, "Automation in agriculture using robotics and artificial intelligence: A review," *IEEE Access*, vol. 9, pp. 103 195–103 212, 2021.

[2] Y. Liu, S. Huang, and R. H. Taylor, "A review of computer vision-based agricultural systems using deep learning," *Computers and Electronics in Agriculture*, vol. 187, 2021.

[3] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Computers and Electronics in Agriculture*, vol. 147, pp. 70–90, 2018.

[4] G. Pantazi, D. Moshou, and A. Tamouridou, "Automated leaf disease detection in crops using ai," *Biosystems Engineering*, vol. 146, pp. 102–111, 2016.

[5] M. J. Escobar and L. M. Sanchez, "Ai-enabled cobot architectures for smart harvesting," *Sensors*, vol. 21, no. 3, p. 873, 2021.

[6] B. Singh, T. Yadav, and K. S. Rathore, "Autonomous cobot system for precision agriculture," in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*, Paris, France, 2020, pp. 985–991.

[7] T. Bacco, F. Delmastro, and A. Zanella, "Smart farming: Opportunities, challenges and use cases," *Computer Networks*, vol. 187, p. 107946, 2021.

[8] "Leaf area index (lai) – nasa earthdata," https://earthdata.nasa.gov/, 2021.

[9] "Normalized difference vegetation index (ndvi) dataset," https://modis.gsfc.nasa.gov/data/dataprod/mod13.php, 2021.

[10] T. Singh and H. Kaur, "Generative ai applications in smart farming," in *Proc. Int. Conf. Artificial Intelligence & Data Analytics (ICAIDA)*, 2023.

[11] "Scikit-learn: Machine learning in python," https://scikit-learn.org/stable/, 2021.

[12] "Random forest algorithm - explained with examples," https://www.analyticsvidhya.com/blog/2021/06/random-forest-classifier/, 2021.

[13] "Streamlit: Turn data scripts into apps," https://streamlit.io/, 2021.

[14] "Precision agriculture using iot and ai," https://www.agritech.org/reports/precision-ai-report.pdf, 2022.

[15] S. D. Patel, M. R. Joshi, and A. Jain, "Smart dashboard for crop prediction using machine learning and sensor data," *International Journal of Computer Applications*, vol. 183, no. 9, pp. 1–7, 2021.

[16] S. Lin, R. Zhang, and B. Liu, "Crop disease recognition based on transfer learning and gan-augmented data," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 11, pp. 3123–3131, 2020.