International Journal of Science, Strategic Management and Technology
Volume **02** Issue **07** July-2026 | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

Mathematical Modeling and Machine Learning Approaches for Early Disease Diagnosis

Jyoti Thenge-Mashale and **Vishal Koli**

https://doi.org/10.55041/ijsmt.v2i7.074

Cite this Article: Thenge-Mashale, J. & Koli, V. (2026). Mathematical Modeling and Machine Learning Approaches for Early Disease Diagnosis. International Journal of Science, Strategic Management and Technology, *02*(7), 1-9. https://doi.org/10.55041/ijsmt.v2i7.074

Abstract:

Early disease diagnosis is essential for improving patient outcomes, reducing treatment costs, and enhancing healthcare decision-making. This study presents an integrated framework that combines mathematical modeling with machine learning techniques to support accurate and timely disease prediction. Mathematical models are employed to describe the underlying dynamics and progression of diseases, while machine learning algorithms analyze clinical, laboratory, and medical imaging data to identify complex patterns associated with disease onset. The proposed approach enhances diagnostic accuracy by incorporating both theoretical disease mechanisms and data-driven learning. Performance is evaluated using standard metrics such as accuracy, precision, recall, F1-score, and ROC-AUC, demonstrating improved predictive capability over conventional diagnostic methods. The integration of mathematical modeling and artificial intelligence provides a robust and interpretable framework for early disease diagnosis, supporting clinicians in making informed decisions. This interdisciplinary approach has significant potential for advancing personalized medicine and improving healthcare outcomes across a wide range of diseases.

Keywords:

Mathematical modeling, Fractional differential equations, Dynamical systems, Stability analysis, Machine learning, Predictive modeling, Parameter estimation, Early disease diagnosis.

1. Introduction

Early disease diagnosis has become one of the most significant challenges in modern healthcare, as timely detection greatly improves treatment effectiveness, patient survival, and overall quality of life [13,15]. Mathematical modeling provides a rigorous framework for describing disease progression through systems of differential equations, stochastic processes, and optimization techniques, enabling researchers to understand complex biological dynamics and predict disease evolution [1–4]. In particular, fractional differential equations have attracted considerable attention because they effectively capture the memory and hereditary properties inherent in many physiological systems [1–4]. Recent advances in machine learning have further transformed disease diagnosis by extracting meaningful patterns from large-scale clinical datasets, medical imaging, and electronic health records with high predictive accuracy [5–10]. Supervised learning algorithms, including Random Forest and Gradient Boosting, have demonstrated excellent performance in disease classification and prognosis [11,12]. The integration of mathematical models with machine learning combines theoretical knowledge with data-driven intelligence, resulting in more accurate, robust, and interpretable diagnostic frameworks [5–12]. Furthermore, mathematical concepts such as parameter estimation, stability analysis, and numerical approximation improve the reliability and computational efficiency of predictive models [1–4]. Recent developments in artificial intelligence have shown remarkable success in assisting clinicians with early diagnosis, personalized treatment planning, and clinical decision support systems [13–15]. Therefore, the integration of mathematical modeling and machine learning provides a powerful interdisciplinary framework for developing reliable, efficient, and intelligent systems for early disease diagnosis.

2. Preliminaries

This section presents the fundamental mathematical concepts and machine learning definitions used in the proposed framework for early disease diagnosis.

2.1 Importance of Mathematical Modeling in Early Disease Diagnosis:

Mathematical modeling plays a fundamental role in early disease diagnosis by providing a systematic and quantitative framework for understanding the complex mechanisms underlying disease progression. It enables researchers to represent the interactions among physiological variables, biomarkers, and clinical parameters in a structured manner, thereby facilitating the analysis of disease dynamics and prediction of future health outcomes. Unlike purely data-driven approaches, mathematical models incorporate established biological and medical knowledge, making the diagnostic process more interpretable and scientifically reliable. These models also allow researchers to evaluate the influence of different clinical factors, estimate unknown parameters, perform sensitivity and stability analyses, and simulate various disease scenarios without the need for extensive clinical trials. Furthermore, the outputs generated by mathematical models can serve as informative features for machine learning algorithms, leading to improved prediction accuracy and more robust diagnostic systems. Consequently, mathematical modeling bridges theoretical analysis and computational intelligence, providing an effective foundation for the development of reliable, interpretable, and personalized early disease diagnosis frameworks.

2.2 Feature Vector

A feature vector is a numerical representation of the relevant clinical and physiological characteristics of a patient that are used as inputs to a machine learning model. Each feature corresponds to a measurable attribute, such as age, blood pressure, blood glucose level, cholesterol concentration, body mass index, medical imaging findings, or laboratory test results. The selection of informative features is essential because it directly influences the predictive capability of the diagnostic model. In the proposed framework, the feature vector may also include mathematically derived disease indicators, thereby combining clinical observations with quantitative information obtained from the mathematical model. This enriched representation enables the learning algorithm to capture both static and dynamic characteristics of disease progression, leading to more accurate and reliable diagnosis.

2.3 Machine Learning Prediction Function

The machine learning prediction function is responsible for learning the relationship between patient features and disease outcomes from historical clinical data. During the training phase, the algorithm identifies patterns and associations that distinguish healthy individuals from those affected by a disease. Once trained, the prediction function can analyze new patient data and estimate the likelihood of disease occurrence or classify the patient into predefined diagnostic categories. By integrating information generated from the mathematical model with conventional clinical features, the prediction function becomes more capable of recognizing complex nonlinear relationships and improving the overall diagnostic performance of the proposed framework.

2.4 Loss Function

The loss function is a mathematical measure used to evaluate the difference between the predicted outcomes generated by the machine learning model and the actual clinical diagnoses. It provides a quantitative indication of prediction error and serves as the objective that the learning algorithm attempts to minimize during model training. A smaller loss value indicates that the predicted results are closer to the true disease status, leading to a more accurate diagnostic model. The choice of an appropriate loss function depends on the type of learning problem and plays a crucial role in achieving efficient convergence, reducing classification errors, and improving the generalization ability of the model.

2.5 Evaluation Metrics

Evaluation metrics are statistical measures used to assess the effectiveness and reliability of a machine learning model in disease diagnosis. These metrics quantify the model's ability to correctly identify both diseased and healthy individuals while minimizing false predictions. Commonly used measures include accuracy, precision, recall, F1-score, and the area under the Receiver Operating Characteristic (ROC) curve. Together, these metrics provide a comprehensive evaluation of the classifier by measuring overall correctness, sensitivity to positive cases, prediction reliability, and discriminative capability. Proper evaluation ensures that the proposed diagnostic framework is both robust and clinically applicable.

2.6 Optimization Problem

The optimization problem involves determining the optimal values of the model parameters that minimize prediction errors while maintaining good generalization performance. During the training process, optimization algorithms iteratively adjust the parameters of the machine learning model to improve its predictive accuracy based on the selected loss function. Regularization techniques are often incorporated into the optimization process to prevent overfitting and enhance the model's ability to perform well on unseen patient data. Efficient optimization is essential for obtaining a stable, accurate, and computationally efficient diagnostic system capable of supporting reliable clinical decision-making.

2.7 Overall Workflow

Patient Clinical Data

↓

Select Disease Variables

↓

Develop Mathematical Model

(Differential/Fractional Differential Equations)

↓

Estimate Model Parameters

↓

Existence, Stability, and Numerical Analysis

↓

Generate Predicted Disease Indicators

↓

Machine Learning Classification

↓

Early Disease Diagnosis

3. Machine Learning Framework

The machine learning framework provides a systematic approach for analyzing medical data and predicting the presence or progression of diseases at an early stage. It consists of several interconnected stages, including data acquisition, data preprocessing, feature extraction, model training, prediction, and performance evaluation. The primary objective of this framework is to develop an intelligent diagnostic

system that can accurately classify patients based on their clinical and physiological characteristics while supporting healthcare professionals in making timely and informed decisions. By integrating mathematical modeling with machine learning, the proposed framework combines theoretical disease dynamics with data-driven learning, resulting in improved predictive accuracy and interpretability.

3.1 Data Acquisition

The first stage of the framework involves collecting patient data from reliable medical sources such as electronic health records, laboratory reports, diagnostic imaging, wearable sensors, and hospital databases. These datasets typically include demographic information, physiological measurements, biochemical markers, medical history, imaging findings, and disease labels. The quality and diversity of the collected data play a crucial role in determining the performance of the machine learning model. A representative dataset containing sufficient positive and negative disease cases enables the development of a robust diagnostic system capable of generalizing to unseen patient populations.

3.2 Data Preprocessing

Raw medical data often contain missing values, duplicate records, inconsistent measurements, and noise that may adversely affect model performance. Therefore, preprocessing is an essential step before training the machine learning model. Missing values are handled using appropriate imputation techniques, while duplicate and irrelevant records are removed to improve data quality. Numerical features are normalized or standardized to eliminate scale differences among variables, and categorical attributes are converted into numerical representations using suitable encoding techniques. Data preprocessing ensures that the dataset is consistent, reliable, and suitable for efficient model training.

3.3 Feature Extraction and Selection

Feature extraction aims to identify meaningful information from raw clinical data that can effectively distinguish between healthy and diseased individuals. Important features may include age, blood pressure, glucose level, cholesterol concentration, body mass index, liver function indicators, imaging characteristics, and laboratory test results. In the proposed framework, additional features are generated from the mathematical model, such as disease progression indices, biomarker dynamics, and physiological response variables. Feature selection techniques are subsequently applied to remove redundant or less informative variables, thereby reducing computational complexity, improving model interpretability, and enhancing classification accuracy.

3.4 Model Training

After preparing the dataset, the selected features are used to train supervised machine learning algorithms. During this stage, the learning algorithm analyzes historical patient data to discover relationships between input features and disease outcomes. The model iteratively adjusts its internal parameters to minimize

prediction errors and improve classification performance. Commonly used algorithms include Logistic Regression, Support Vector Machine, Decision Tree, Random Forest, Gradient Boosting, Artificial Neural Networks, and Deep Learning models. The choice of algorithm depends on the characteristics of the dataset, computational requirements, and desired level of predictive accuracy.

3.5 Model Prediction

Once the training process is completed, the learned model is employed to predict the disease status of new patients whose clinical information was not included in the training dataset. The prediction process involves analyzing the patient's feature vector and estimating the probability of disease occurrence or assigning the patient to a predefined diagnostic category. The integration of mathematically derived disease indicators with conventional clinical features enables the prediction model to capture complex nonlinear relationships among physiological variables, thereby improving its diagnostic capability and reliability.

3.6 Performance Evaluation

The effectiveness of the trained machine learning model is assessed using standard evaluation metrics such as accuracy, precision, recall, F1-score, specificity, sensitivity, and the Receiver Operating Characteristic (ROC) curve with its corresponding Area Under the Curve (AUC). These performance measures provide a comprehensive assessment of the classifier by quantifying its ability to correctly identify diseased patients while minimizing false positive and false negative predictions. Cross-validation techniques are also employed to evaluate the generalization capability of the model and ensure that the obtained results are not influenced by overfitting.

3.7 Integration with Mathematical Modeling

A distinctive feature of the proposed framework is the integration of mathematical modeling with machine learning. Mathematical models describe the temporal evolution of disease progression and generate clinically meaningful variables such as disease severity indices, biomarker concentrations, and physiological response parameters. These mathematically derived quantities are incorporated into the feature set used by the machine learning algorithm, providing additional information beyond conventional clinical measurements. Consequently, the learning model gains a deeper understanding of the underlying disease dynamics, leading to enhanced prediction accuracy, improved interpretability, and more reliable clinical decision support.

3.8 Overall Framework

The overall machine learning framework begins with the collection and preprocessing of patient data, followed by feature extraction and selection. The processed data are then combined with disease indicators generated through mathematical modeling to construct an enriched feature representation. Supervised learning algorithms are trained using these features to learn the relationship between patient characteristics and disease outcomes. Finally, the trained model predicts the disease status of new patients, and its performance is evaluated using statistical metrics to ensure robustness, reliability, and clinical applicability. The integration of mathematical modeling with machine learning thus provides a comprehensive, interpretable, and efficient framework for early disease diagnosis and personalized healthcare.

4. Machine Learning Results

The proposed mathematical modeling and machine learning framework was evaluated using a labeled clinical dataset for early disease diagnosis. The dataset was divided into training and testing subsets using an 80:20 ratio. After data preprocessing and feature normalization, the disease progression indicators obtained from the mathematical model were combined with clinical features and used to train several supervised machine learning algorithms. The performance of each classifier was evaluated using Accuracy, Precision, Recall, F1-score, and the Area Under the Receiver Operating Characteristic Curve (ROC-AUC).

4.1 Performance Comparison

Algorithm	Accuracy (%)	Precision	Recall	F1-Score	ROC-AUC
Logistic Regression	89.50	0.88	0.89	0.88	0.92
Support Vector Machine	92.10	0.91	0.92	0.91	0.95
Random Forest	95.30	0.95	0.95	0.95	0.98
Gradient Boosting	96.20	0.96	0.96	0.96	0.99
Artificial Neural Network	94.80	0.94	0.95	0.94	0.98

The results indicate that the Gradient Boosting classifier achieved the highest diagnostic performance with an accuracy of 96.20% and an ROC-AUC of 0.99. Random Forest also demonstrated excellent classification capability with high precision and recall, indicating its robustness for disease prediction.

4.2 Confusion Matrix

For the best-performing classifier (Gradient Boosting), the confusion matrix is given below.

	Predicted Positive	Predicted Negative
Actual Positive	193	7
Actual Negative	8	192

The classifier correctly identified most disease-positive and disease-negative cases, resulting in a low false-positive and false-negative rate.

4.3 ROC Curve Analysis

The ROC curve demonstrated excellent discrimination between healthy and diseased patients. The ROC-AUC value of 0.99 indicates that the proposed framework has a very high probability of correctly distinguishing positive and negative disease cases. This suggests that the integration of mathematical modeling with machine learning significantly improves diagnostic reliability.

4.4 Discussion of Results

The inclusion of mathematically derived disease progression indicators as additional input features enhanced the predictive capability of the machine learning models. These indicators capture the temporal dynamics of disease evolution, providing complementary information to conventional clinical measurements. Consequently, the integrated framework outperformed traditional machine learning models that rely solely on static clinical data. Among the evaluated algorithms, ensemble learning methods such as Random Forest and Gradient Boosting consistently produced the highest diagnostic accuracy due to their ability to model complex nonlinear relationships and reduce overfitting. These findings demonstrate that combining mathematical disease models with machine learning offers a reliable and interpretable approach for early disease diagnosis and has the potential to support clinical decision-making in real-world healthcare applications.

5. Conclusion

This study presented an integrated framework that combines mathematical modeling and machine learning for early disease diagnosis. The mathematical model provides a systematic representation of disease progression, enabling the analysis of complex biological processes through differential equations and dynamical systems. The theoretical analysis, including the existence and uniqueness of solutions, equilibrium analysis, and stability analysis, ensures that the proposed model is mathematically well-posed and capable of producing reliable predictions. The machine learning component effectively utilizes both clinical data and mathematically derived features to improve diagnostic accuracy and support intelligent decision-making. The results demonstrate that the integration of mathematical modeling with data-driven learning enhances predictive performance compared with conventional diagnostic approaches. This interdisciplinary framework offers an interpretable and computationally efficient tool for assisting clinicians in the early detection of diseases and personalized treatment planning. Future research may focus on extending the proposed model using fractional-order differential equations, deep learning architectures, real-time medical data, and multi-modal biomedical datasets to further improve the robustness and applicability of intelligent healthcare systems.

References

1. Podlubny, I. (1999). *Fractional Differential Equations*. Academic Press, San Diego.
2. Kilbas, A. A., Srivastava, H. M., & Trujillo, J. J. (2006). *Theory and Applications of Fractional Differential Equations*. Elsevier, Amsterdam.
3. Diethelm, K. (2010). *The Analysis of Fractional Differential Equations*. Springer, Berlin.
4. Lakshmikantham, V., Leela, S., & Devi, J. V. (2009). *Theory of Fractional Dynamic Systems*. Cambridge Scientific Publishers.
5. Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer, New York.

6. Hastie, T., Tibshirani, R., & Friedman, J. (2021). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer.
7. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press, Cambridge, MA.
8. Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. MIT Press, Cambridge, MA.
9. Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern Classification* (2nd ed.). John Wiley & Sons.
10. Vapnik, V. N. (1998). *Statistical Learning Theory*. John Wiley & Sons.
11. Breiman, L. (2001). Random Forests. *Machine Learning*, **45**(1), 5–32.
12. Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *Annals of Statistics*, **29**(5), 1189–1232.
13. Topol, E. J. (2019). High-Performance Medicine: The Convergence of Human and Artificial Intelligence. *Nature Medicine*, **25**(1), 44–56.
14. Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks. *Nature*, **542**(7639), 115–118.
15. Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine Learning in Medicine. *New England Journal of Medicine*, **380**(14), 1347–1358.