IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 5, May 2026

DOI: 10.17148/IARJSET.2026.135102

# Patient Risk Identification Using Machine Learning

**Dr. T. Amalraj Victoire[1], K. Sanjai[2]**

[1]Associate Professor, Department of MCA, Sri Manakula Vinayagar Engineering College, Madagadipet, Puducherry – 605 107, India.

[2]Student, Department of MCA, Sri Manakula Vinayagar Engineering College, Madagadipet, Puducherry – 605 107, India.

**Abstract:** These days, hospitals and clinics keep a lot of patient information on computers. Trying to manage all these records by hand gets really hard as more and more patients come in every day. Doctors often have to look at medical reports, lab results, and a patient's past quickly. Because of this, healthcare places are slowly moving towards smart systems that can help them analyze things and make predictions faster.

This project is about finding out patient health risks using computer programs called machine learning algorithms. The system figures out if a patient is at low or high risk by looking at health details like blood pressure, sugar level, cholesterol, heart rate, and age. While we were building it, we tried and tested different algorithms, including Logistic Regression, Decision Tree, Support Vector Machine, and Random Forest.

We built the application using Python and gave it a simple screen so people could easily use it. After trying out the different models, the Random Forest algorithm gave us better results than the other methods we used. This system can help medical staff with their first look at a patient's case and potentially help them spot health risks sooner.

**Keywords**: Patient Risk Prediction, Machine Learning, Healthcare Analysis, Disease Prediction, Random Forest

## 1. INTRODUCTION

In recent years, healthcare companies have started using technology that relies on data to make medical services and patient care better. Every hospital creates a huge amount of patient information, like lab reports, prescriptions, scan results, and how they were treated before. Handling all this by hand takes a lot of time and constant effort from doctors and nurses. When you analyze things by hand, there's a chance that diagnosis might get delayed if there are too many patient records. Sometimes, important patterns hidden in healthcare data are hard to see using old-fashioned methods. That's why machine learning techniques are now being used in many healthcare apps.

Machine learning helps computer systems learn from old healthcare records and then make predictions based on what they've learned from that data. These prediction systems can give doctors quick analysis and help them make decisions during treatment.

The main goal of this project is to build a system that uses machine learning to find patient risks. The system takes health information as input and then guesses what kind of risk category the patient might be in. Our work mainly aims to make predictions more effective and cut down on how much manual analysis is needed.

## 2. LITERATURE SURVEY

Many research projects have looked into predicting health issues using machine learning. Older healthcare systems mostly relied on statistical math and doctors manually figuring out patient conditions. While these ways were helpful, they weren't very good at dealing with huge amounts of healthcare data.

Logistic Regression algorithms have been widely used for sorting healthcare problems because they are simple to set up and give easy-to-understand results. Decision Tree algorithms are also common in medical prediction systems because of their straightforward structure and ability to classify things.

Support Vector Machine algorithms did a better job at handling complicated healthcare datasets with many different details. Random Forest algorithms became popular because they combine several decision trees and tend to be more accurate, while also helping avoid problems where the model learns too much from the training data.

© IARJSET This work is licensed under a Creative Commons Attribution 4.0 International License

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 5, May 2026

DOI: 10.17148/IARJSET.2026.135102

Lots of recent studies suggest that machine learning can make disease prediction and patient monitoring systems better. However, researchers have also pointed out issues with the quality of the data, missing values, imbalanced data, and choosing the right features. These things can affect how well healthcare prediction models actually perform.

## 3. THEORETICAL FRAMEWORK

Our proposed system is built using ideas from predictive analytics and machine learning. This setup looks at healthcare data and tries to guess patient risk levels based on patterns it learned from past data.

We use medical details like blood pressure, sugar level, cholesterol, age, and heart rate as the information we feed into the system. Before we train the models, we clean up the dataset using special preprocessing steps. This means we get rid of duplicate records, deal with any missing values, and standardize the healthcare parameters.

After cleaning, the machine learning algorithms are trained using healthcare datasets. The trained models then put patients into either a low-risk or high-risk group. The whole point of this system is to make predictions faster and help healthcare professionals when they're looking at patient information.

## 4. METHODOLOGY

Our work involved several steps during its creation.

First, we gathered healthcare data from public healthcare sources and some example medical records. This collected data included different health details needed for making predictions.

Since raw healthcare data can often have gaps and messy parts, we cleaned it up before training our models. We carefully handled missing values, took out any duplicate entries, and made sure numbers were consistent where needed.

We split the dataset into two parts: one for training and one for testing. Then, we trained different machine learning algorithms, including Logistic Regression, Decision Tree, Support Vector Machine, and Random Forest, using the cleaned data.

Once trained, the system predicts patient risk levels based on the health information a user types in. The prediction result shows up on a graphical screen we made for the project.

| Feature | Existing System \| | Proposed System |
|---|---|---|
| Manual Analysis | Time-consuming | Automated ML Prediction |
| Accuracy | Lower Accuracy | Improved Accuracy |
| Decision Support | Limited | Intelligent Support |
| Risk Prediction | Manual Diagnosis | Automated Prediction |
| Data Processing | Difficult | Efficient Handling |

Table:4.1 Comparison of Existing System vs Proposed System

## 5. EXISTING SYSTEM

Older healthcare systems mostly relied on doctors and other medical staff to figure out a patient's health. They would manually go through patient reports, symptoms, and lab results before making a diagnosis.

A big problem with this way is how long it takes to go through tons of patient data. In busy hospitals or clinics, analyzing things by hand can be tough and sometimes causes delays in catching serious conditions.

These older systems also don't have smart prediction features. They mainly focus on what's happening with the patient right now, instead of trying to predict future risks based on health patterns. Because of these downsides, automated healthcare prediction systems are becoming much more important.

© IARJSET This work is licensed under a Creative Commons Attribution 4.0 International License

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 5, May 2026

DOI: 10.17148/IARJSET.2026.135102

## 6. PROPOSED SYSTEM

Our system uses a machine learning approach to find patient risks. The program looks at health details and tries to predict if a patient falls into a low-risk or high-risk group.

We built the system using the Python programming language and included a user-friendly graphical interface. We tried out and compared different machine learning algorithms during our testing.

This model can help medical professionals by cutting down on the manual work needed for analysis and giving quick prediction results during initial checks.

6.1 Modules of the Proposed System

A. Data Collection Module

This part gathers health-related information from datasets and medical records.

B. Data Preprocessing Module

This handles missing information, removes repeated entries, and tidies up the data.

C. Machine Learning Module

This is where the machine learning algorithms are put into action for predicting and sorting.

D. Prediction Module

This module predicts how risky a patient is using the machine learning models we trained.

E. Visualization Module

This shows the prediction results and the outcomes of the health analysis.

F. User Interface Module

This lets users type in health details and see the prediction results.

## 7. SYSTEM ARCHITECTURE

The system's setup includes several parts that work together to predict health outcomes.

First, patient health details are entered through the input screen. The preprocessing part cleans and changes the data before sending it to the prediction part.

The machine learning models then look at the health information and create prediction results. These results are then shown on the graphical screen for health analysis.

Having a system built with different parts like this makes it more flexible and easier to keep up with and improve later on.

© IARJSET This work is licensed under a Creative Commons Attribution 4.0 International License

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 5, May 2026

DOI: 10.17148/IARJSET.2026.135102

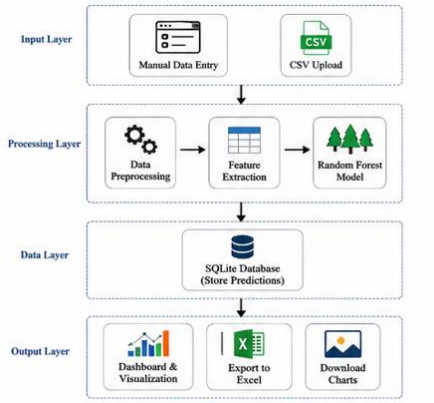


Fig:7.1 System Architecture of Supply Chain of Risk Prediction System

## 8. IMPLEMENTATION

We built this project using the Python programming language. We used libraries like NumPy, Pandas, Scikit-learn, and Matplotlib while developing and testing it. Tkinter was our choice for making the graphical user interface.

We cleaned up the healthcare datasets before training the machine learning models. We separately implemented and tested Logistic Regression, Decision Tree, Support Vector Machine, and Random Forest algorithms using the healthcare data.

Out of all the models we tested, Random Forest showed better prediction accuracy when we checked its performance. The trained model predicts patient risk levels and shows the result on the screen.

## 9. EVALUATION METRICS

We checked how well the system worked using measures like accuracy, precision, recall, and F1-score.

Accuracy tells us how often the prediction results were correct overall. Precision checks how precisely positive cases were identified, while recall measures how well the system found all the actual positive cases.

The F1-score gives a balanced picture by combining precision and recall. These measurements help us figure out how good and reliable our prediction system is.

## 10. RESULTS AND ANALYSIS

We tested the machine learning algorithms we built using healthcare datasets to compare how well they predicted things.

Here are the accuracy results we got during testing:

- Logistic Regression – 84%
- Decision Tree – 87%
- Support Vector Machine – 89%
- Random Forest – 93%

Looking at these results, Random Forest did better than the other algorithms we used in this project. The model handled the data more effectively and gave consistent prediction outputs during testing.

This project showed that machine learning methods can indeed help healthcare prediction systems and cut down on the amount of manual analysis needed.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 5, May 2026

DOI: 10.17148/IARJSET.2026.135102

Confusion Matrix

| Predicted \ Actual | Low | Medium | High |
|---|---|---|---|
| Low | 78 | 24 | 9 |
| Medium | 19 | 60 | 17 |
| High | 10 | 23 | 96 |

| Performance Metrics | |
|---|---|
| Accuracy | 61.9% |
| Balanced Accuracy | 67.0% |
| Precision (Low) | 0.71 |
| Precision (Medium) | 0.55 |
| Precision (High) | 0.78 |
| Recall (Low) | 0.71 |
| Recall (Medium) | 0.57 |
| Recall (High) | 0.73 |
| F1-Score (Weighted Avg) | 0.62 |

Fig:10.1 Confusion Matrix & Performance Metrics

## 11. APPLICATIONS

Our system could be used in places like:

- ❖ Hospitals
- ❖ Clinics
- ❖ Healthcare Research Centers
- ❖ Medical Diagnostic Systems
- ❖ Health Monitoring Applications

It can help medical professionals when they're keeping an eye on patients and doing their first checks.

## 12. ADVANTAGES AND LIMITATIONS

Advantages

- Helps spot patient risks early
- Cuts down on manual analysis work
- Gives faster health predictions
- Helps in making healthcare decisions
- Makes predictions more efficient

Limitations

- How accurate the prediction is depends on the quality of the data
- Requires the healthcare data to be properly prepared
- Its performance might not be the same with different dataset.

## 13. ETHICAL AND PRACTICAL CONSIDERATIONS

Patient health information needs to be handled very carefully because medical data is extremely private. We must prevent anyone unauthorized from getting to healthcare records to protect patient privacy.

Bias in healthcare datasets can mess up prediction accuracy and lead to unfair outcomes. So, it's really important to use balanced datasets and clean them up properly when building the models.

Also, we need to regularly check and update the system to keep it working well and reliably.

© IARJSET This work is licensed under a Creative Commons Attribution 4.0 International License

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 5, May 2026

DOI: 10.17148/IARJSET.2026.135102

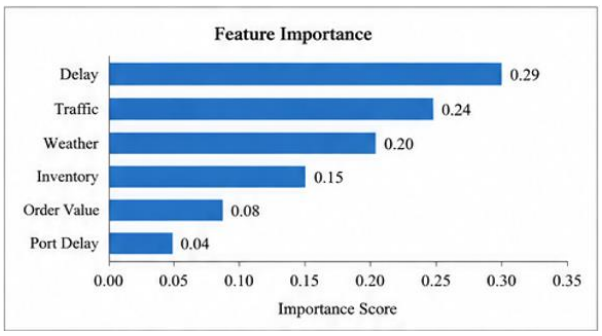


Fig 13.1: Future Importance Chart

## 14. FUTURE ENHANCEMENTS

Our system could be made even better in the future by using bigger healthcare datasets and more advanced deep learning techniques.

Some possible improvements for the future include:

- Watching patients in real-time
- Healthcare systems that run on the cloud
- Mobile apps for healthcare
- Health monitoring using IoT devices
- More advanced prediction tools
- Connecting with hospital management systems

## 15. CONCLUSION

This project introduced a system that uses machine learning to figure out patient risks, predicting potential health issues based on various health details. We tried out and tested several different machine learning algorithms while working on it.

Among the models we used, Random Forest was the most accurate and gave steady results during testing. This system can help healthcare professionals by reducing their manual analysis tasks and supporting earlier predictions about health.

One key thing we learned from this project was how important the quality of the data is for prediction performance. Missing values, inconsistent records, and imbalanced datasets directly impact how accurate the model is. Because of this, cleaning up the data properly became a crucial part of how we built the system.

Our system isn't meant to replace doctors or medical experts. Instead, it's designed to be a helpful tool for analyzing health data and keeping an eye on patients. In the future, we can expand this project by using bigger datasets, advanced deep learning methods, and systems that monitor health in real-time.

## REFERENCES

[1] Leo Breiman, “Random Forests,” Machine Learning Journal, 2001.
[2] Tom M. Mitchell, Machine Learning, McGraw-Hill Education, 1997.
[3] Christopher M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.
[4] Ian H. Witten, Eibe Frank, Mark A. Hall, Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann, 2016.
[5] Scikit-learn Documentation, “Machine Learning in Python,” 2024.
[6] Sebastian Raschka and Vahid Mirjalili, Python Machine Learning, Packt Publishing, 2019.
[7] Jiawei Han, Micheline Kamber, Jian Pei, Data Mining: Concepts and Techniques, Elsevier, 2011.
[8] Trevor Hastie, Robert Tibshirani, Jerome Friedman, The Elements of Statistical Learning, Springer, 2009.
[9] Ian Goodfellow, Yoshua Bengio, Aaron Courville, Deep Learning, MIT Press, 2016.
[10] World Health Organization, “Artificial Intelligence in Healthcare: Opportunities and Challenges,” 2023.

© IARJSET This work is licensed under a Creative Commons Attribution 4.0 International License