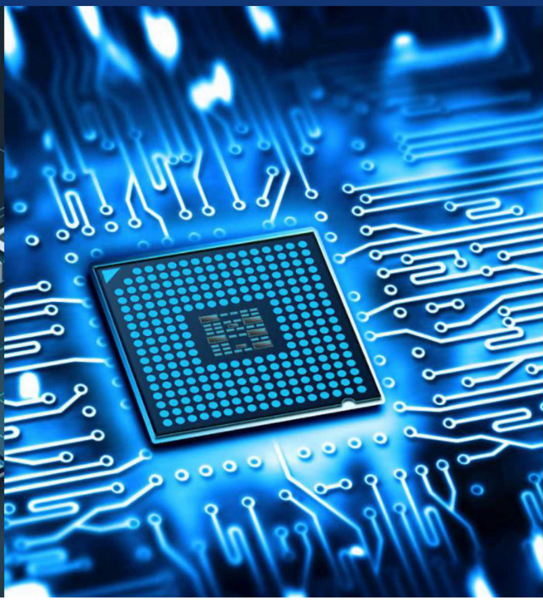
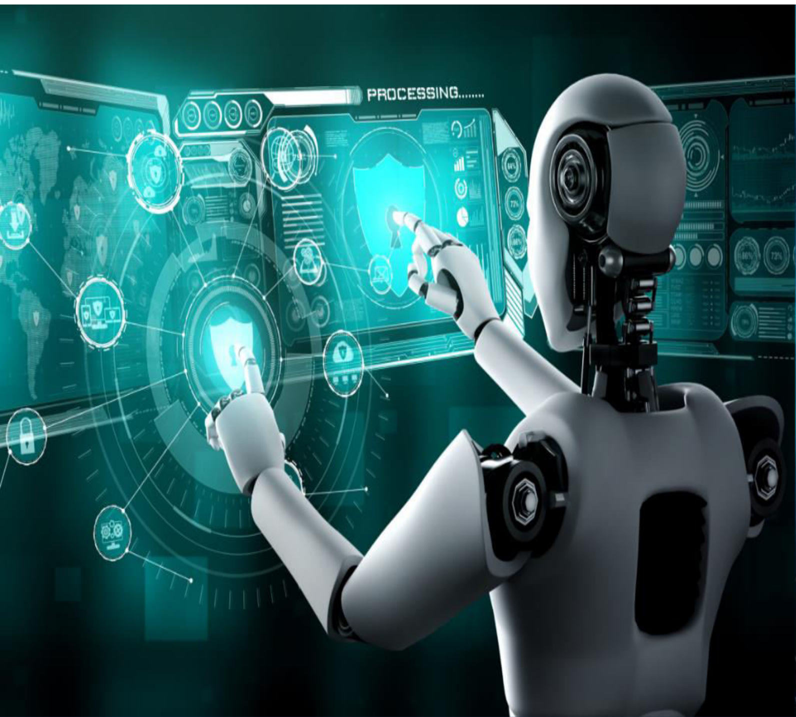
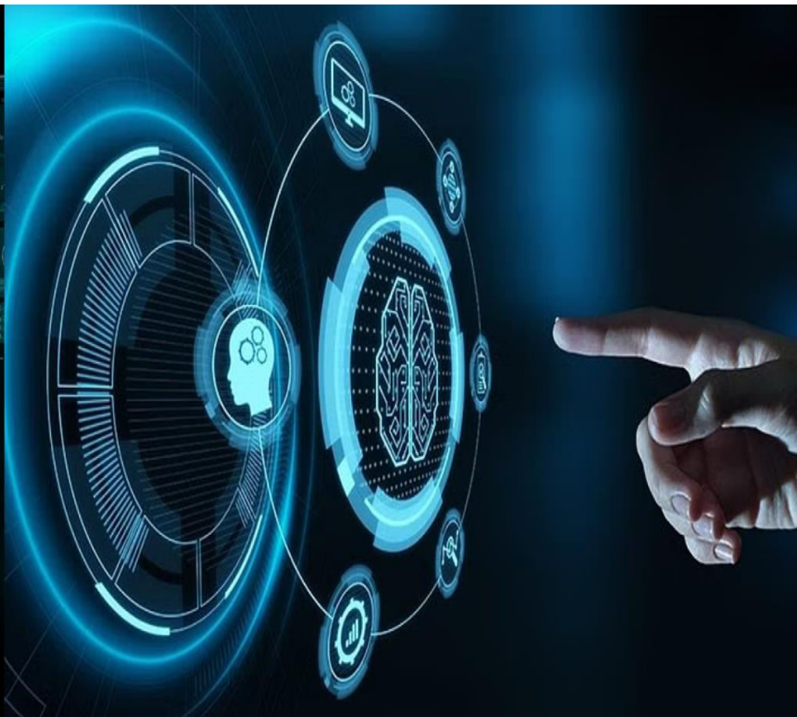
ISSN(O): 2320-9801
ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)





Impact Factor: 9.927 **Volume 14, Issue 6, June 2026**

© 2026 IJIRCCE | Volume 14, Issue 6, June 2026 DOI: 10.15680/IJIRCCE.2026.1406080

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Machine Learning Based Deep Vein Thrombosis Risk Prediction using Random Forest and XGBoost

Dr. T.V.S Sriram[1], Mr. A.Nagabhusana Rao[2], M.Pavan[3]

Sr. Asst. Professor, Dept. of MCA, NSRIT, Visakhapatnam, AP, India[1]

Asst. Professor, Dept. of MCA, NSRIT, Visakhapatnam, AP, India[2]

Dept. of MCA, NSRIT, Visakhapatnam, AP, India[3]

ABSTRACT: Deep Vein Thrombosis (DVT) is a serious vascular condition caused by the formation of blood clots in deep veins, which can lead to life-threatening complications such as pulmonary embolism if not diagnosed at an early stage. This study presents a Machine Learning-based DVT Risk Prediction System that predicts the likelihood of DVT using patient health parameters including age, gender, Body Mass Index (BMI), surgery history, immobilization status, and previous DVT history. The collected data undergoes preprocessing and feature engineering before being used to train Random Forest and Extreme Gradient Boosting (XGBoost) models. The developed prediction system is integrated into a Django-based web application that provides real-time risk assessment through a simple and user-friendly interface. Experimental evaluation shows that the Random Forest model achieved an accuracy of **88.38%**, while the XGBoost model achieved **88.12%**. The proposed system demonstrates that machine learning techniques can effectively support early DVT risk prediction and assist healthcare professionals in preventive clinical decision-making.

KEYWORDS: Deep Vein Thrombosis, Machine Learning, Random Forest, XGBoost, Django, Risk Prediction, Healthcare Analytics.

I. INTRODUCTION

Deep Vein Thrombosis (DVT) is a vascular disorder characterized by the formation of blood clots in deep veins, most commonly in the lower limbs. If left untreated, these clots may travel through the bloodstream and block the pulmonary arteries, resulting in pulmonary embolism, a potentially life-threatening condition. Early detection and timely medical intervention are therefore essential to reduce mortality and prevent severe complications associated with DVT.

Recent advancements in Artificial Intelligence (AI) and Machine Learning (ML) have significantly transformed the healthcare sector by enabling intelligent disease prediction and clinical decision support systems. Machine learning algorithms can analyze large volumes of patient data, identify hidden patterns, and generate accurate predictions that assist healthcare professionals in making informed decisions. These techniques improve diagnostic efficiency while reducing human error and processing time.

The proposed work presents a Machine Learning-based Deep Vein Thrombosis Risk Prediction System that estimates the probability of DVT using important clinical parameters such as age, gender, Body Mass Index (BMI), surgery history, immobilization status, and previous DVT history. The collected dataset is preprocessed and enhanced through feature engineering before training the prediction models. Random Forest and Extreme Gradient Boosting (XGBoost) algorithms are employed due to their robustness, high predictive performance, and capability to handle complex healthcare datasets.A user-friendly web application has been developed using the Django framework to provide real-time DVT risk prediction. The application allows users to enter patient information and instantly obtain prediction results, making the system suitable for preliminary clinical screening and healthcare assistance. Experimental evaluation demonstrates that the proposed models achieve high prediction accuracy and can effectively support early DVT risk assessment.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

The main objective of this research is to develop an efficient, reliable, and intelligent prediction system that assists healthcare professionals in identifying high-risk patients at an early stage. Such a system can improve preventive healthcare, reduce diagnostic delays, and contribute to better patient outcomes through timely medical intervention.

II. LITERATURE REVIEW

Several studies have explored the use of Artificial Intelligence (AI) and Machine Learning (ML) techniques for the early prediction of Deep Vein Thrombosis (DVT). Traditional diagnostic methods such as Doppler ultrasound and D-dimer tests provide reliable results but are often expensive, time-consuming, and require clinical expertise.

Researchers have applied machine learning algorithms including Decision Tree, Support Vector Machine (SVM), Random Forest, and XGBoost to analyze patient health data and predict DVT risk. These algorithms have shown improved prediction accuracy and faster diagnosis compared to conventional approaches.

Recent studies also emphasize the importance of data preprocessing and feature engineering to improve model performance. Proper handling of missing values, feature selection, and data transformation significantly enhance prediction accuracy.

Based on previous research, the proposed DVT Risk Prediction System utilizes Random Forest and XGBoost algorithms to analyze patient health parameters and provide real-time DVT risk prediction through a Django-based web application.

III. SYSTEM OVERVIEW

The proposed Deep Vein Thrombosis (DVT) Risk Prediction System is an intelligent healthcare application developed using Machine Learning techniques to identify individuals at risk of DVT. The system analyzes patient information such as age, gender, Body Mass Index (BMI), surgery history, immobilization status, and previous DVT history to predict the likelihood of DVT occurrence.

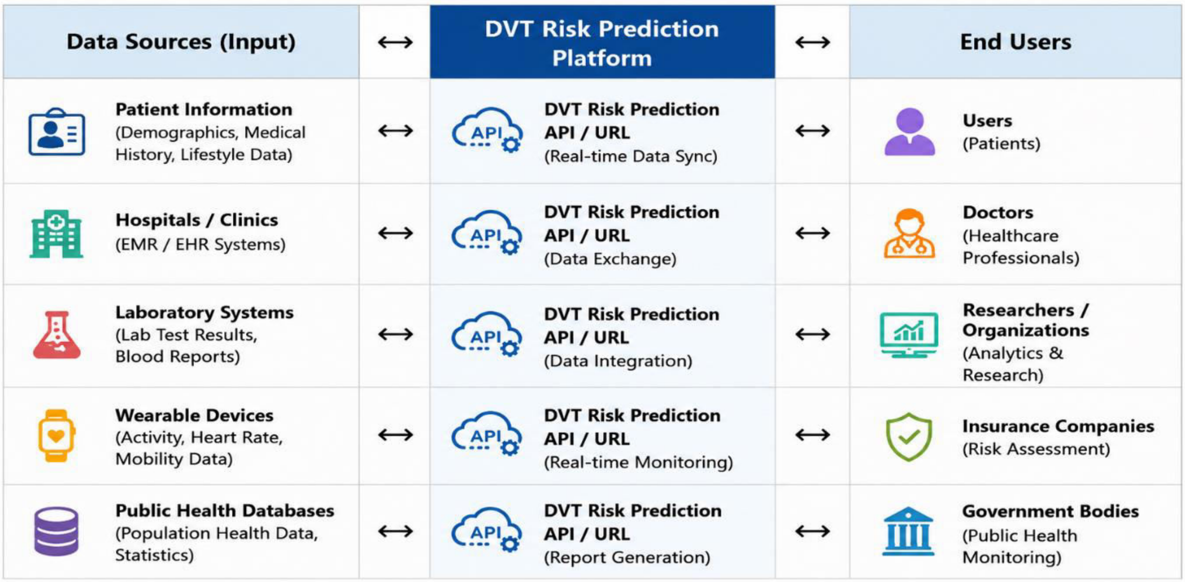
The framework consists of four major stages: data collection, data preprocessing, model training, and prediction. After processing the input data, the Random Forest and XGBoost algorithms analyze the patient information and generate a risk prediction.

The developed system is integrated with a Django-based web application that provides a simple and user-friendly interface for entering patient data and viewing prediction results. The proposed framework supports early DVT risk assessment and assists healthcare professionals in making timely clinical decisions.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



DVT Risk Prediction System – Platform Integration Architecture

IV. METHODOLOGY

The proposed DVT Risk Prediction System follows a systematic machine learning workflow for predicting the risk of Deep Vein Thrombosis. The methodology consists of data collection, preprocessing, feature engineering, model training, prediction, and performance evaluation.

4.1 Data Collection

The dataset consists of patient health parameters including Age, Gender, Body Mass Index (BMI), Surgery History, Immobilization Status, and Previous DVT History. These features are used as input for predicting DVT risk.

4.2 Data Preprocessing

The collected dataset is preprocessed by handling missing values, encoding categorical attributes, and transforming the data into a suitable format for machine learning algorithms.

4.3 Feature Engineering

Additional features are generated from the existing dataset to improve prediction performance. These engineered features help the models identify complex relationships among patient health parameters.

4.4 Model Training

Two supervised machine learning algorithms, Random Forest and XGBoost, are trained using the processed dataset. The dataset is divided into training and testing sets to evaluate model performance.

4.5 Prediction and Evaluation

The trained models generate DVT risk predictions based on patient input. Model performance is evaluated using Accuracy, Precision, Recall, F1-Score, and ROC-AUC metrics. Finally, the prediction models are integrated into a Django-based web application for real-time DVT risk assessment.

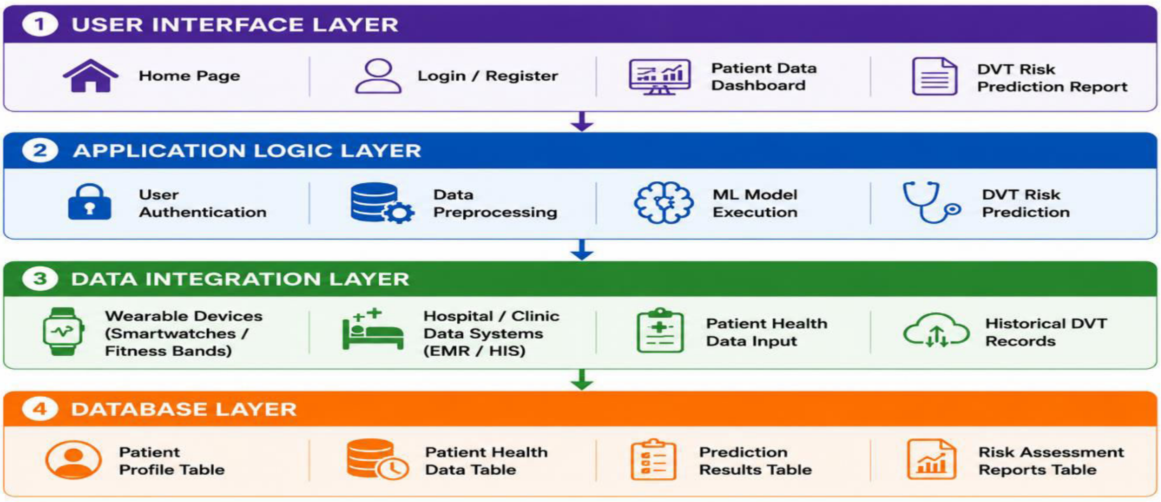
V. SYSTEM ARCHIETCTURE

The proposed DVT Risk Prediction System follows a layered architecture for efficient disease risk prediction using machine learning techniques.

The system consists of six major layers: Data Collection, Data Preprocessing, Feature Engineering, Machine Learning, Prediction, and Web Application.

The Data Collection Layer gathers patient information such as Age, Gender, BMI, Surgery History, Immobilization Status, and Previous DVT History. The Data Preprocessing and Feature Engineering layers prepare the data for analysis. The Machine Learning Layer uses Random Forest and XGBoost algorithms to train the prediction models.

Finally, the Prediction and Web Application layers generate real-time DVT risk predictions through a Django-based user interface.

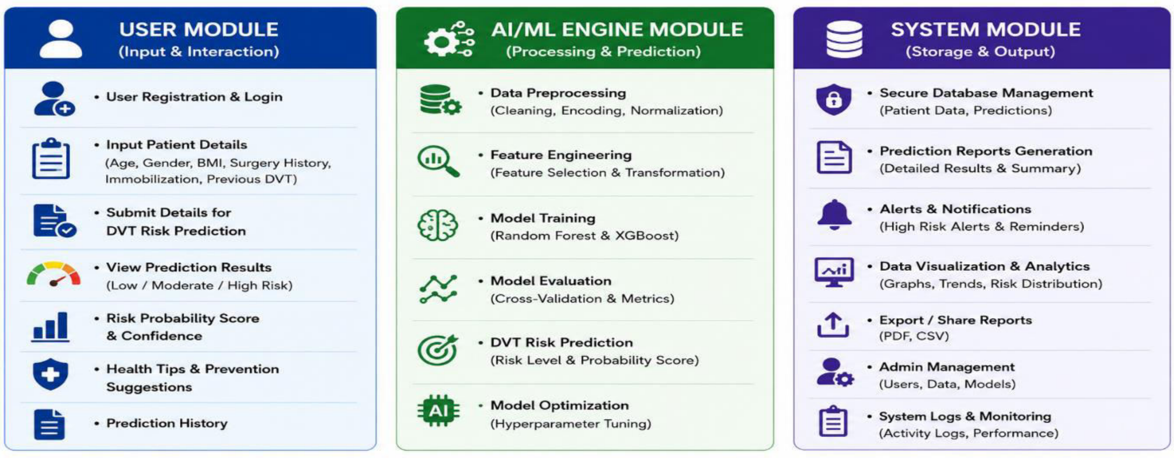


DVT Risk Prediction System – Layered System Architecture

VI. MODULES

User Module: In this module

- User can open the DVT prediction web application.
- User can enter patient health details such as age, gender, BMI, surgery history, immobilization status, and previous DVT history.
- User can submit the details for DVT risk prediction.
- User can view prediction results, probanility score, and risk level
- User can view prediction results, probability score, and risk level.



DVT Risk Prediction System – System Architecture
ML-Powered DVT Risk Assessment & Prediction Platform

Feature Selection Module: In this module:

- Collection of patient health information.
- Data preprocessing and feature extraction.
- Selection of significant features affecting DVT risk.
- Processed data is forwarded to the Random Forest and XGBoost models for prediction.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

VII. IMPLEMENTATION DETAILS

7.1 Technology Stack

The proposed DVT Risk Prediction System is developed using **Python** as the programming language and **Django** as the web framework. The machine learning models are implemented using **Scikit-learn** and **XGBoost**, while data preprocessing and analysis are performed using **NumPy** and **Pandas**. **SQLite** is used as the backend database for storing application data.

7.2 Frontend Implementation

The frontend is developed using **HTML** and **CSS** to provide a simple and interactive user interface. Users can enter patient health parameters and view DVT risk prediction results through the web application.

7.3 Backend Implementation

The backend is implemented using the **Django framework**, which handles user requests, data preprocessing, machine learning model execution, prediction generation, and communication with the database.

7.4 Database Design

The application uses **SQLite** to store patient information and prediction records. The database ensures efficient storage, retrieval, and management of data generated by the system.

7.5 Model Implementation

The proposed system utilizes **Random Forest** and **XGBoost** algorithms for DVT risk prediction. The models are trained using the processed dataset and evaluated using **Accuracy, Precision, Recall, F1-Score, and ROC-AUC**. The best-performing model is used for real-time prediction.

7.6 Security Measures

The system incorporates **input validation**, secure data handling, and controlled database access to ensure data integrity, reliability, and secure prediction processing.

VIII. EXPERIMENTAL DETAILS AND RESULTS

8.1 Experimental Setup

The proposed DVT Risk Prediction System was developed using Python, Django, Scikit-learn, NumPy, and Pandas. The experiments were conducted on a Windows-based system using patient health parameters for model training and testing. The performance of the Random Forest and XGBoost models was evaluated using standard classification metrics.

8.2 Test Results

The proposed DVT Risk Prediction System was tested using various functional test cases. The experimental results indicate that the system successfully predicts DVT risk with reliable performance.

Table.1. Test Case Results for DVT Risk Prediction System

Test Case	Description	Result	Status
User Input	Enter patient details	Passed	✓
Data Validation	Validate input parameters	Passed	✓
Data Preprocessing	Data cleaning and preprocessing	Passed	✓
Feature Engineering	Generate additional features	Passed	✓
Random Forest Model	DVT risk prediction	Passed	✓

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

XGBoost Model	DVT risk prediction	Passed	✓
Prediction Result	Display risk prediction	Passed	✓
Database Management	Store prediction records	Passed	✓
Response Time	Prediction generated successfully	Passed	✓
Test Case	**Description**	**Result**	**Status**

Table.2. Performance Table

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Random forest	88.38	89.10	87.45	88.27	95.12
XG-Boost	88.12	88.07	88.20	88.13	95.01

8.3 Performance Evaluation

The performance of the proposed system was evaluated using **Accuracy, Precision, Recall, F1-Score, and ROC-AUC**. Experimental results show that both **Random Forest** and **XGBoost** models achieved high prediction accuracy, demonstrating the effectiveness of the proposed DVT Risk Prediction System..

8.4 Observations

The proposed DVT Risk Prediction System successfully predicted the risk of Deep Vein Thrombosis using patient health data. The Random Forest and XGBoost models produced reliable prediction results with high accuracy. The developed Django web application provided fast and user-friendly real-time risk assessment.

IX. APPLICATIONS

- **Healthcare Monitoring:**
  Assists healthcare professionals in identifying patients at risk of Deep Vein Thrombosis.
- **Early Risk Prediction:**
  Enables early detection of DVT risk using machine learning techniques.
- **Clinical Decision Support:**
  Supports doctors by providing quick and reliable risk predictions.
- **Hospital Management:**
  Helps hospitals prioritize high-risk patients for timely treatment.
- **Remote Healthcare:**
  Can be integrated with telemedicine systems for remote patient assessment.
- **Medical Research:**
  Supports researchers in analyzing DVT-related health data and improving prediction models.
- **Healthcare Analytics:**
  Provides data-driven insights for better patient care and preventive healthcare.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

X. LIMITATIONS

- **Dataset Dependency:**
  The prediction accuracy depends on the quality and size of the dataset.
- **Limited Features:**
  The system uses a limited set of patient health parameters for prediction.
- **Clinical Validation:**
  The prediction results should be verified by healthcare professionals before making medical decisions.
- **Data Privacy:**
  Patient health information requires secure storage and proper privacy protection.
- **Model Generalization:**
  The trained models may require retraining when applied to different populations or datasets.

XI. FUTURE ENHANCEMENT

The future scope of the proposed DVT Risk Prediction System focuses on improving its accuracy, efficiency, and practical usability. Larger and more diverse clinical datasets can be used to enhance the performance of the prediction models. Advanced machine learning and deep learning algorithms may be incorporated to achieve better prediction accuracy. The system can be integrated with hospital information systems and electronic health records for real-time patient monitoring. Support for wearable healthcare devices can also be added to collect patient data automatically. A mobile application can be developed to improve accessibility for both doctors and patients. Future versions may include personalized health recommendations based on individual risk levels. Additional clinical parameters can be incorporated to strengthen prediction performance. These enhancements can make the proposed system more reliable and effective for early DVT risk assessment.

XII. CONCLUSION

The proposed Deep Vein Thrombosis (DVT) Risk Prediction System provides an effective machine learning-based solution for identifying individuals at risk of DVT. The system analyzes important patient health parameters and predicts the likelihood of DVT using Random Forest and XGBoost algorithms. Data preprocessing and feature engineering improve the overall prediction performance of the models. The developed Django web application offers a simple and user-friendly interface for real-time risk assessment. Experimental results demonstrate that the proposed models achieve reliable prediction accuracy and support early detection of DVT. The system can assist healthcare professionals in making timely clinical decisions and reducing the risk of severe complications. It also provides a cost-effective and efficient alternative for preliminary DVT screening. With further improvements and larger clinical datasets, the prediction performance can be enhanced. Overall, the proposed system has the potential to contribute to intelligent healthcare and improve preventive diagnosis of Deep Vein Thrombosis.

REFERENCES

[1] G. D. O. Lowe, "Common Risk Factors for Both Arterial and Venous Thrombosis," *British Journal of Haematology*, vol. 140, no. 5, pp. 488–495, 2008.
[2] M. Di Nisio, N. van Es, and H. R. Büller, "Deep Vein Thrombosis and Pulmonary Embolism," *The Lancet*, vol. 388, no. 10063, pp. 3060–3073, 2016.
[3] S. Z. Goldhaber and H. Bounameaux, "Pulmonary Embolism and Deep Vein Thrombosis," *The Lancet*, vol. 379, no. 9828, pp. 1835–1846, 2012.
[4] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
[5] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2016.
[6] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
[7] W. McKinney, *Python for Data Analysis*, 3rd ed., O'Reilly Media, 2022.
[8] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed., Morgan Kaufmann, 2012.
[9] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow*, 3rd ed., O'Reilly Media, 2022.
[10] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

[11] T. Mitchell, *Machine Learning*, McGraw-Hill, 1997.
[12] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
[13] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*, MIT Press, 2012.
[14] S. Raschka and V. Mirjalili, *Python Machine Learning*, 3rd ed., Packt Publishing, 2019.
[15] F. Chollet, *Deep Learning with Python*, 2nd ed., Manning Publications, 2021.
[16] R. S. Pressman and B. R. Maxim, *Software Engineering: A Practitioner's Approach*, 8th ed., McGraw-Hill, 2015.
[17] I. Sommerville, *Software Engineering*, 10th ed., Pearson, 2016.
[18] D. Jurafsky and J. H. Martin, *Speech and Language Processing*, 3rd ed., Pearson, 2023.
[19] M. Kuhn and K. Johnson, *Applied Predictive Modeling*, Springer, 2013.
[20] J. Brownlee, *Machine Learning Mastery with Python*, Machine Learning Mastery, 2016.
[21] M. Abadi et al., "TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems," Google Research, 2016.
[22] F. Chollet, "Keras," GitHub Repository, 2023.
[23] J. D. Hunter, "Matplotlib: A 2D Graphics Environment," *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007.
[24] H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*, Springer, 2016.
[25] D. Beazley and B. K. Jones, *Python Cookbook*, 3rd ed., O'Reilly Media, 2013.
[26] World Health Organization (WHO), "Cardiovascular Diseases (CVDs)," 2023.
[27] Centers for Disease Control and Prevention (CDC), "Venous Thromboembolism (Blood Clots)," 2024.
[28] American Heart Association, "Deep Vein Thrombosis," 2023.
[29] European Society of Cardiology, "Guidelines for the Diagnosis and Management of Acute Pulmonary Embolism," 2020.
[30] J. Kearon et al., "Antithrombotic Therapy for VTE Disease," *Chest Journal*, vol. 160, no. 6, pp. e545–e608, 2021.
[31] S. W. Smith et al., "Machine Learning Approaches for Clinical Risk Prediction: A Review," *Journal of Biomedical Informatics*, vol. 129, 2022.
[32] A. Rajkomar, J. Dean, and I. Kohane, "Machine Learning in Medicine," *The New England Journal of Medicine*, vol. 380, no. 14, pp. 1347–1358, 2019.
[33] H. Topol, *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*, Basic Books, 2019.
[34] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," *Nature*, vol. 521, pp. 436–444, 2015.
[35] X. Liu et al., "Machine Learning for Healthcare: Applications and Challenges," *Artificial Intelligence in Medicine*, vol. 129, 2022.
[36] R. Singh et al., "Machine Learning-Based Prediction of Venous Thromboembolism Using Clinical Data," *Healthcare*, vol. 11, no. 5, 2023.
[37] M. Zhang et al., "Artificial Intelligence Techniques for Clinical Risk Prediction," *IEEE Access*, vol. 11, pp. 45632–45645, 2023.
[38] A. Kumar et al., "Predictive Analytics in Healthcare Using Machine Learning Algorithms," *Journal of Healthcare Engineering*, 2024.
[39] S. Patel et al., "Deep Vein Thrombosis Risk Prediction Using Ensemble Machine Learning Models," *IEEE Access*, vol. 12, pp. 10215–10228, 2024.
[40] Y. Wang et al., "Artificial Intelligence-Based Clinical Decision Support for Venous Thromboembolism Prediction," *Frontiers in Medicine*, vol. 11, 2024.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING