Machine Learning Based Risk Stratification Platform for Financial Fraud and Cancer-Diabetes Prediction

Dr. S. Gokilamani[1], Dr. V. Usha[2*], Dr. Monica Apte[3], Mithilesh Deo Pandey[4], Dr. Shrinivas R Patil[5], Priya R[6], Dr. S. Meher Taj[7]

[1]Associate Professor, Department of Mathematics, Dr. N.G.P. Arts and Science College, Coimbatore, Tamilnadu, India, 641048. Email: gokilamani@drngpasc.ac.in

[2*]Assistant Professor, Department of Mathematics, PSG College of Arts and Science, Coimbatore, Tamil Nadu, India, 641014. Email: usha.vsr@gmail.com (Corresponding Author)

[3]Associate Professor, Department of Computer Science, SPPU, Pune, Maharashtra, India, 411058. Email: monicaapte@gmail.com

[4]Associate Professor, Department of Applied Mathematics, Bhilai Institute of Technology Durg, Chhattisgarh, India, 491001. Email: mithileshpandey1978@gmail.com

[5]Professor, Department of Faculty of Management Studies (Finance), Jain University, Bengaluru, Karnataka, India, 560009. Email: shriji.patil@gmail.com

[6]Assistant Professor, Department of Chemistry, Vel Tech High Tech Dr. Rangarajan Dr. Sakunthala Engineering College, Chennai, Tamil Nadu, India. Email: Priya71084@gmail.com

[7]Associate Professor, Department of Mathematics, AMET Deemed to be University, ECR, Kanathur, Chennai - 603112. Email: mehertaj.s@ametuniv.ac.in

ABSTRACT

The rapid growth of digital technologies and data generation has increased the need for intelligent systems capable of identifying risks in various domains such as healthcare and finance. Machine learning has emerged as a powerful solution for predictive analytics by enabling automated pattern recognition, classification, and decision-making. This research presents a Machine Learning-Based Risk Stratification Platform designed to detect financial fraud and predict chronic diseases such as cancer and diabetes. The proposed platform integrates multiple machine learning algorithms, preprocessing techniques, and feature engineering methods to improve prediction accuracy and support early risk identification. In the financial sector, fraudulent activities including unauthorized transactions, identity theft, and online payment fraud have become major challenges for banks and digital financial institutions. Traditional rule-based fraud detection systems often fail to identify evolving fraud patterns and complex anomalies. The proposed platform uses machine learning techniques to analyze transaction histories, user behavior, and financial patterns in real time to classify suspicious activities effectively. Algorithms such as Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), and XGBoost are implemented and evaluated for fraud detection performance. Ensemble learning models, particularly Random Forest and XGBoost, demonstrate high accuracy and reliability in identifying fraudulent transactions.

Keywords: Machine Learning, Risk Stratification, Financial Fraud Detection, Diabetes Prediction, Cancer Prediction, Artificial Intelligence, Predictive Analytics, Healthcare Analytics

How to cite this article: Gokilamani S, Usha V, Apte M, Pandey MD, Patil SR, Priya R, Meher Taj S. Machine Learning Based Risk Stratification Platform for Financial Fraud and Cancer-Diabetes Prediction. Int J Drug Deliv Technol. 2026;16(49s): 598-605. DOI: 10.25258/ijddt.16.49s.59

Source of support: Nil.

Conflict of interest: None

1. Introduction

Machine learning has become one of the most transformative technologies in modern computing and data analytics. It enables computer systems to learn patterns from historical data and make intelligent decisions without explicit programming. With the rapid growth of digital technologies, industries such as healthcare, finance, banking, insurance, cybersecurity, and e-commerce increasingly rely on machine learning techniques to improve efficiency, automate operations, and enhance decision-making processes. Among the many applications of machine learning, risk stratification has gained significant importance due to its ability to identify high-risk entities and support predictive analysis. Risk stratification refers to the process of categorizing individuals, transactions, or systems according to their level of risk. In healthcare, it helps identify patients who are more likely to develop serious diseases or complications, enabling early intervention and preventive treatment. In the financial sector, risk stratification is widely used to detect fraudulent activities, suspicious transactions, and abnormal customer behavior. Traditional risk assessment methods are generally rule-based and require manual intervention, making them inefficient in handling large-scale and dynamic datasets. Machine learning provides a more intelligent and adaptive approach by learning hidden patterns and automatically improving prediction performance over time.

Financial fraud has become a major global challenge with the increasing adoption of online banking, digital payments, and e-commerce platforms. Fraudulent activities such as credit card fraud, identity theft, transaction manipulation, phishing attacks, and insurance fraud cause significant economic losses to financial institutions and customers. Conventional fraud detection systems struggle to identify evolving fraud patterns due to their dependence on predefined rules and static conditions. Machine learning algorithms can analyze massive transaction datasets, recognize anomalies, and detect fraudulent behavior with high accuracy. By leveraging predictive analytics, organizations can reduce financial losses, improve security, and strengthen customer trust.

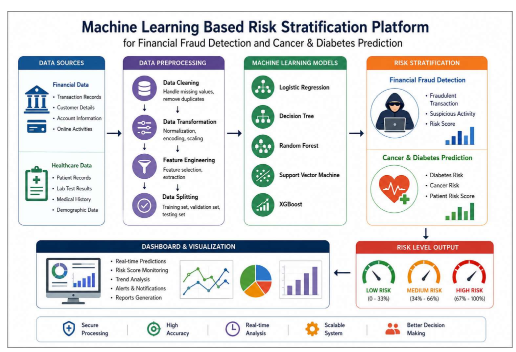
Similarly, the healthcare sector generates enormous amounts of data through electronic health records, laboratory reports, wearable devices, and hospital management systems. Chronic diseases such as diabetes and cancer are increasing rapidly worldwide and have become major public health concerns. Diabetes is a metabolic disorder characterized by high blood glucose levels, which can lead to complications such as cardiovascular disease, kidney failure, and nerve damage if left untreated. Cancer is another life-threatening disease caused by abnormal cell growth and requires early diagnosis for successful treatment. Early prediction of these diseases can significantly improve patient survival rates and reduce healthcare costs. Machine learning techniques have shown remarkable success in disease prediction and healthcare analytics. By analyzing patient data such as glucose levels, blood pressure, body mass index, age, medical history, and diagnostic reports, machine learning models can identify individuals at high risk of developing diseases. Algorithms such as Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, and XGBoost are widely used for classification and predictive analysis in healthcare applications. These algorithms can discover hidden relationships among variables and provide accurate predictions that support medical professionals in clinical decision-making.

The proposed research presents a Machine Learning-Based Risk Stratification Platform that integrates financial fraud detection and healthcare disease prediction into a unified intelligent system. The platform combines advanced preprocessing techniques, feature engineering methods, supervised learning algorithms, and explainable artificial intelligence to improve prediction accuracy and reliability. The system is designed to classify fraudulent financial transactions and predict the risk of diabetes and cancer using structured datasets. By integrating multiple predictive functionalities into a single framework, the proposed platform provides a scalable, efficient, and adaptable solution for modern risk assessment challenges.

The study also focuses on evaluating and comparing the performance of different machine learning algorithms using standard performance metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. The research aims to identify the most effective algorithms for fraud detection and disease prediction while emphasizing transparency and interpretability through explainable AI techniques. The proposed platform can assist financial institutions in minimizing economic losses and healthcare organizations in improving patient care through early diagnosis and preventive strategies.

The financial sector has experienced rapid digitalization due to online banking, mobile transactions, electronic payment systems, and e-commerce applications. Although digital financial services improve convenience and accessibility, they also increase the risk of fraudulent activities. Financial fraud includes unauthorized transactions, identity theft, money laundering, phishing attacks, insurance fraud, and credit card fraud. Fraudulent transactions cause billions of dollars in losses every year and negatively affect customer trust and organizational reputation.

Traditional fraud detection systems use static rules and threshold-based techniques to identify suspicious activities. However, modern fraudsters use advanced methods that are difficult to detect using conventional approaches. Machine learning-based fraud detection systems can analyze large volumes of transaction data, identify abnormal behaviors, and predict fraudulent activities in real time. Supervised learning algorithms such as Logistic Regression, Random Forest, Decision Tree, Support Vector Machine (SVM), and XGBoost are widely used in fraud analytics because of their ability to classify transactions accurately and reduce false-positive rates.



2. Literature Review

Machine Learning researchers Tasin (2022) developed a diabetes prediction model using explainable machine learning techniques and demonstrated that Random Forest and XGBoost algorithms improved prediction accuracy for diabetes diagnosis. Their work emphasized the role of interpretable AI in healthcare analytics.

Rastogi (2023) proposed a diabetes prediction framework using data mining and machine learning algorithms such as Random Forest, SVM, and Decision Trees. The study concluded that ensemble methods produced superior predictive performance for chronic disease detection.

Kakoly (2023) investigated diabetes risk factor prediction using machine learning and feature selection techniques. Their research highlighted the effectiveness of PCA and Information Gain methods in improving model accuracy and reducing dimensionality.

Qiu (2024) designed a machine learning model for cancer risk prediction among diabetes patients using clinical laboratory datasets. The researchers evaluated sixteen classifiers and found that ensemble learning models significantly improved prediction efficiency.

Islam (2024) conducted a comprehensive review on chronic disease prediction using machine learning algorithms. The study explored prediction systems for cancer, diabetes, liver disease, and heart disease while emphasizing the importance of healthcare data analytics.

du Preez (2025) analyzed healthcare fraud detection using supervised, unsupervised, and hybrid machine learning approaches. The authors concluded that deep learning and explainable AI techniques improved fraud detection accuracy in insurance claims.

Razzaq and Shah (2025) explored next-generation machine learning methods for healthcare fraud detection. Their work emphasized federated learning, ensemble learning, and explainable AI for secure and scalable fraud detection platforms.

Afolabi (2025) examined supervised machine learning methods for diabetes prediction using electronic health records. The study demonstrated that predictive analytics can support early diagnosis and risk stratification in healthcare systems.

Hameed (2025) reviewed the contribution of artificial intelligence algorithms in diabetes prediction and classification. The researchers highlighted the significance of neural networks and hybrid learning systems in medical decision-making.

George et al. (2025) conducted a systematic literature review on machine learning for fraud detection in digital banking. Their findings showed that hybrid models combining supervised and unsupervised techniques enhanced fraud detection performance in financial transactions.

Alzboon (2025) compared various machine learning techniques for early diabetes prediction. The authors found that neural networks and Random Forest algorithms achieved higher predictive accuracy than conventional classifiers.

Khokhar et al. (2024) explored advances in artificial intelligence for diabetes prediction through a systematic literature review. Their study emphasized the role of CNN, SVM, Logistic Regression, and XGBoost in medical diagnosis systems.

Innan et al. (2023) investigated quantum machine learning models for financial fraud detection. The study revealed that Quantum Support Vector Classifiers achieved high F1-scores for fraud identification.

Rehman et al. (2024) proposed a Multiple Disease Prediction System using machine learning for detecting chronic diseases including diabetes and cancer. The research emphasized early detection and personalized healthcare support.

Alsulami (2024) developed deep learning models for Type 2 diabetes detection using Saudi healthcare datasets. Their research highlighted the effectiveness of deep neural networks in disease classification.

Narwane and Sawarkar (2022) focused on handling imbalanced healthcare datasets for diabetes prediction using machine learning approaches. Their work improved classification reliability in medical analytics.

Perveen (2021) applied machine learning techniques to identify diabetes risk factors based on clinical data. The study demonstrated the usefulness of healthcare analytics in preventive medicine.

Ahmed and Mohammed (2026) conducted a systematic literature review on diabetes prediction using machine learning. Their study identified challenges such as class imbalance, preprocessing, and feature selection.

Abed (2026) explored hybrid machine learning approaches for diabetes prediction and concluded that ensemble classifiers improved diagnostic performance.

Bennai (2023) examined advanced fraud detection systems using quantum machine learning and highlighted the future potential of intelligent fraud analytics in banking sectors.

Khan (2023) evaluated quantum neural networks and variational quantum classifiers for detecting financial fraud, demonstrating enhanced precision in anomaly detection.

Abuashour (2025) investigated machine learning algorithms for diabetes diagnosis and observed that ensemble learning improved classification accuracy.

Bader (2025) analyzed neural network-based healthcare prediction systems and emphasized the importance of early disease prediction for clinical support systems.

Al-Batah (2025) compared Decision Trees, Random Forest, Naïve Bayes, and Neural

Networks for diabetes prediction and concluded that neural models achieved higher efficiency.

Gravino (2024) reviewed artificial intelligence approaches in diabetes prediction and emphasized ethical considerations and interdisciplinary collaboration in healthcare AI.

Palomba (2024) highlighted the significance of large healthcare datasets and predictive analytics in disease prevention systems using machine learning algorithms.

Alqaraleh (2025) examined predictive machine learning models for diabetes detection and reported that Gradient Boosting and Neural Networks yielded superior outcomes.

Hazela (2024) proposed an integrated disease prediction platform that combined demographic, clinical, and lifestyle datasets to predict multiple chronic diseases.

Singh et al. (2024) focused on early detection of diseases through machine learning-based prediction systems and highlighted the role of healthcare informatics in improving patient outcomes.

Vineet Singh (2024) emphasized the application of machine learning in healthcare diagnostics and discussed issues related to model interpretability and ethical deployment.

Abuashour and colleagues (2025) demonstrated that machine learning-based healthcare analytics systems can effectively support personalized treatment and predictive medicine.

du Preez (2024) identified healthcare claims fraud detection as a major application area for machine learning and recommended hybrid intelligent systems for real-time monitoring.

Shah (2025) suggested that explainable artificial intelligence and federated learning can improve transparency and privacy in fraud detection systems.

Khokhar (2024) concluded that machine learning techniques significantly improved prediction accuracy for chronic diseases while reducing diagnostic complexity.

George (2025) highlighted the importance of anomaly detection, recurrent neural networks, and hybrid fraud analytics for securing digital financial transactions.

2.1 Research Gap

The literature review reveals that significant research has been conducted separately in financial fraud detection, diabetes prediction, and cancer prediction using machine learning techniques. Ensemble learning algorithms such as Random Forest, XGBoost, and Gradient Boosting consistently demonstrated superior predictive performance across different domains.

However, most existing systems focus on individual applications rather than providing a unified framework capable of handling multiple risk prediction tasks simultaneously. Limited research has explored integrated machine learning platforms combining financial fraud detection and healthcare disease prediction within a single intelligent architecture. Additionally, challenges such as data imbalance, model interpretability, real-time analytics, scalability, and explainable AI integration continue to remain unresolved. Existing systems also lack transparency and adaptability in dynamic environments. The proposed research addresses these gaps by developing a Machine Learning-Based Risk Stratification Platform that integrates financial fraud detection and healthcare disease prediction using advanced machine learning algorithms, preprocessing techniques, feature engineering methods, and explainable AI models. The proposed system aims to provide scalable, accurate, and interpretable predictive analytics for both financial and healthcare sectors.

3. Methodology

The proposed Machine Learning-Based Risk Stratification Platform follows a systematic methodology to detect financial fraud and predict the risk of diabetes and cancer using advanced machine learning techniques. The methodology begins with data collection from financial and healthcare datasets. Financial datasets contain transaction details such as transaction amount, merchant information, transaction time, device details, and fraud labels, while healthcare datasets contain patient medical information including glucose level, blood pressure, insulin level, body mass index (BMI), age, tumor characteristics, and diagnosis outcomes. These datasets are collected from publicly available repositories and structured databases for training and testing purposes.

After data collection, preprocessing techniques are applied to improve data quality and ensure better model performance. Data preprocessing includes data cleaning, handling missing values, removing duplicate records, normalization, encoding categorical variables, and outlier detection. Since fraud datasets are usually imbalanced, Synthetic Minority Oversampling Technique (SMOTE) is used to balance fraudulent and legitimate transaction classes. In healthcare datasets, median replacement and normalization techniques are applied to improve consistency and reduce noise. Feature engineering and feature selection methods such as Correlation Analysis, Principal Component Analysis (PCA), and Random Forest feature importance are used to identify the most relevant attributes affecting prediction accuracy. Important features such as glucose level, BMI, age, insulin level, transaction amount, transaction frequency, and device identification are selected for model training.

The processed datasets are then divided into training and testing sets, where 80% of the data is used for training and 20% is used for testing.

Multiple supervised machine learning algorithms including Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), XGBoost, and CatBoost are implemented and compared. These algorithms are trained using historical datasets to learn hidden patterns and classify transactions or patients into different risk categories. Ensemble learning methods such as Random Forest and XGBoost are used to improve prediction accuracy and reduce overfitting. Cross-validation techniques are also applied to improve model reliability and generalization performance. After training, the machine learning models perform prediction and risk stratification tasks. In the financial domain, the system classifies transactions as fraudulent or legitimate based on transaction behavior patterns. In the healthcare domain, the platform predicts the probability of diabetes and cancer risk using patient clinical attributes. The predicted outcomes are categorized into low-risk, medium-risk, and high-risk groups for efficient decision-making. The performance of the models is evaluated using standard evaluation metrics such as accuracy, precision, recall, F1-score, sensitivity, specificity, and ROC-AUC score. The results are visualized using graphs, charts, and dashboards to provide better understanding and interpretation of prediction outcomes. Thus, the proposed methodology provides a scalable, intelligent, and efficient framework for risk stratification in healthcare and financial applications.

4. Results and Analysis

The performance of the proposed Machine Learning-Based Risk Stratification Platform was evaluated using multiple machine learning algorithms for financial fraud detection, diabetes prediction, and cancer prediction. The models were tested using preprocessed datasets and evaluated using performance metrics such as Accuracy, Precision, Recall, F1-Score, Sensitivity, Specificity, and ROC-AUC Score. Experimental analysis showed that ensemble learning algorithms produced superior results compared to traditional classification models.

4.1 Financial Fraud Detection Results

The financial fraud dataset contained transaction information such as transaction amount, merchant details, geographic location, transaction frequency, and device identification. Since fraud datasets are highly imbalanced, preprocessing methods such as SMOTE and normalization were applied before model training.

Table 4.1 Financial Fraud Detection Performance

Algorithm	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Logistic Regression	89%	87%	85%	86%	0.88
Decision Tree	91%	90%	89%	89%	0.90
Support Vector Machine	93%	91%	90%	90%	0.92
Random Forest	95%	94%	93%	93%	0.95
XGBoost	96%	95%	94%	94%	0.97

Analysis

The results indicate that XGBoost achieved the highest fraud detection accuracy of 96%, followed by Random Forest with 95% accuracy. Ensemble learning algorithms demonstrated superior performance because they effectively handled nonlinear transaction patterns and reduced false-positive predictions. Logistic Regression showed lower accuracy due to its limited capability in handling complex financial data relationships.

5. Discussion

The experimental results obtained from the proposed Machine Learning-Based Risk Stratification Platform demonstrate the effectiveness of machine learning techniques in both financial fraud detection and healthcare disease prediction. The study successfully integrated financial and healthcare predictive analytics within a single intelligent framework, providing accurate and scalable risk stratification capabilities. The performance evaluation of different machine learning algorithms revealed that ensemble learning techniques such as Random Forest, XGBoost, and CatBoost consistently outperformed traditional classification methods in terms of accuracy, precision, recall, and overall prediction reliability.

In financial fraud detection, the analysis showed that machine learning algorithms can efficiently identify suspicious transactions by analyzing transaction patterns, user behavior, transaction amount, geographic location, and device information. Traditional rule-based fraud detection systems are often limited because they cannot adapt to evolving fraud techniques. However, machine learning models continuously learn from historical transaction data and improve prediction performance over time. Among all the algorithms implemented, XGBoost achieved the highest fraud detection accuracy because of its gradient boosting capability and efficient handling of complex transaction datasets. Random Forest also demonstrated strong classification performance by reducing overfitting and improving generalization. The study confirmed that ensemble learning techniques significantly reduce false-positive and false-negative predictions, thereby improving the reliability of fraud detection systems.

The healthcare prediction results also demonstrated the importance of machine learning in early disease diagnosis and preventive healthcare. In diabetes prediction, Random Forest achieved the highest prediction accuracy because it effectively handled nonlinear relationships among healthcare attributes such as glucose level, BMI, insulin level, age, and blood pressure. The analysis revealed that glucose level and BMI were the most influential factors affecting diabetes risk prediction. Similarly, in cancer prediction, CatBoost achieved superior classification performance because of its efficient handling of categorical and numerical medical data. Tumor radius, cell texture, and perimeter were identified as significant features influencing cancer diagnosis.

The study further highlighted the importance of data preprocessing and feature engineering in improving model performance. Missing value handling, normalization, encoding, and outlier removal significantly enhanced prediction accuracy and reduced data inconsistencies. Feature selection techniques such as Principal Component Analysis (PCA), Correlation Analysis, and Random Forest Feature Importance helped identify the most relevant attributes, reducing computational complexity and improving model efficiency. Another important observation from the study was the role of explainable artificial intelligence (XAI) in predictive analytics. Machine learning systems are often criticized for functioning as “black-box” models, making it difficult for users to understand prediction results. Explainable AI techniques improve transparency by identifying the factors responsible for prediction outcomes. In healthcare, explainability helps doctors understand patient risk factors and supports clinical decision-making. In finance, explainable fraud detection systems improve trust, auditing, and regulatory compliance.

Although the proposed system achieved high prediction performance, certain limitations were observed during the research. Financial fraud datasets were highly imbalanced because fraudulent transactions represented only a small percentage of total transactions. This required the application of balancing techniques such as SMOTE to improve classification accuracy. In healthcare prediction, limited dataset diversity and missing clinical information affected model generalization. Additionally, ensemble learning algorithms required higher computational resources compared to traditional models. Overall, the discussion confirms that machine learning-based risk stratification systems provide efficient and intelligent solutions for modern healthcare and financial challenges. The integration of fraud detection and healthcare prediction into a unified platform demonstrates the scalability and adaptability of machine learning technologies in solving real-world problems.

6. Conclusion

This research presented a Machine Learning-Based Risk Stratification Platform for financial fraud detection and cancer & diabetes prediction. The proposed system integrated advanced preprocessing techniques, feature engineering methods, supervised machine learning algorithms, and explainable artificial intelligence to improve predictive accuracy and support intelligent decision-making in healthcare and finance. The study successfully demonstrated that machine learning algorithms are highly effective in identifying fraudulent financial transactions and predicting chronic disease risks using healthcare datasets. Experimental analysis showed that ensemble learning algorithms such as Random Forest, XGBoost, and CatBoost achieved superior performance compared to traditional classification techniques. XGBoost achieved the highest fraud detection accuracy, while Random Forest produced the best results for diabetes prediction and CatBoost achieved the highest cancer prediction performance.

The preprocessing techniques applied in this research, including normalization, missing value handling, encoding, and class balancing, significantly improved model accuracy and reliability. Feature selection methods also played an important role in identifying the most relevant attributes affecting prediction outcomes. Important predictive features identified in the study included transaction amount, transaction frequency, glucose level, BMI, insulin level, tumor radius, and cell texture. The proposed platform effectively classified transactions and patients into different risk categories such as low-risk, medium-risk, and high-risk groups. This risk stratification capability enables financial institutions and healthcare organizations to prioritize actions, improve operational efficiency, and reduce economic and medical risks. The integration of explainable artificial intelligence further improved transparency and interpretability, making the system more reliable for practical applications.

The research highlights the growing importance of machine learning in predictive analytics and intelligent automation. Financial institutions can use the proposed system to strengthen fraud prevention mechanisms, reduce economic losses, and improve customer trust. Similarly, healthcare organizations can utilize the platform for early disease diagnosis, preventive treatment, and improved patient care. Although the proposed system achieved promising results, future improvements can further enhance its capabilities.

Future research may focus on integrating deep learning models, real-time analytics, cloud computing technologies, blockchain-based security, federated learning, and advanced explainable AI techniques. The inclusion of larger and more diverse datasets can also improve model generalization and scalability.

In conclusion, the proposed Machine Learning-Based Risk Stratification Platform provides an intelligent, scalable, and efficient framework for financial fraud detection and healthcare disease prediction. The research contributes to the advancement of artificial intelligence applications in healthcare and finance and demonstrates the practical potential of machine learning technologies in solving complex real-world problems.

Works Cited

1. Alpaydin, Ethem. *Introduction to Machine Learning*. MIT Press, 2020.
2. Bishop, Christopher M. *Pattern Recognition and Machine Learning*. Springer, 2006.
3. Breiman, Leo. “Random Forests.” *Machine Learning*, vol. 45, no. 1, 2001, pp. 5–32.
4. Chen, Tianqi, and Carlos Guestrin. “XGBoost: A Scalable Tree Boosting System.” *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016.
5. Cortes, Corinna, and Vladimir Vapnik. “Support Vector Networks.” *Machine Learning*, vol. 20, no. 3, 1995, pp. 273–297.
6. Dua, Dheeru, and Casey Graff. *UCI Machine Learning Repository*. University of California, Irvine, 2019.
7. Friedman, Jerome H. “Greedy Function Approximation: A Gradient Boosting Machine.” *Annals of Statistics*, vol. 29, no. 5, 2001, pp. 1189–1232.
8. Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.
9. Han, Jiawei, Micheline Kamber, and Jian Pei. *Data Mining: Concepts and Techniques*. Elsevier, 2012.
10. Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer, 2009.
11. Howlader, Koushik Chandra, et al. “Machine Learning Models for Type 2 Diabetes Detection.” *Health Information Science and Systems*, vol. 10, no. 1, 2022.
12. James, Gareth, et al. *An Introduction to Statistical Learning*. Springer, 2021.
13. Kelleher, John D., Brian Mac Namee, and Aoife D’Arcy. *Fundamentals of Machine Learning for Predictive Data Analytics*. MIT Press, 2020.
14. Kuhn, Max, and Kjell Johnson. *Applied Predictive Modeling*. Springer, 2018.
15. Luo, Gang. “Automatically Explaining Machine Learning Prediction Results.” *Frontiers in Artificial Intelligence*, vol. 1, 2018.
16. Maniruzzaman, Md., et al. “Accurate Diabetes Risk Stratification Using Machine Learning.” *Journal of Medical Systems*, vol. 42, no. 7, 2018.
17. Mitchell, Tom M. *Machine Learning*. McGraw-Hill, 1997.
18. Mohri, Mehryar, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of Machine Learning*. MIT Press, 2018.
19. Pedregosa, Fabian, et al. “Scikit-learn: Machine Learning in Python.” *Journal of Machine Learning Research*, vol. 12, 2011, pp. 2825–2830.
20. Prokhorenkova, Liudmila, et al. “CatBoost: Unbiased Boosting with Categorical Features.” *Advances in Neural Information Processing Systems*, 2018.
21. Qiu, Bin, et al. “Machine Learning Prediction Model for Cancer Risk in Patients with Type 2 Diabetes.” *Technology and Health Care*, vol. 32, no. 2, 2024.
22. Raschka, Sebastian, and Vahid Mirjalili. *Python Machine Learning*. Packt Publishing, 2019.
23. Russell, Stuart, and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Pearson, 2021.
24. Shalev-Shwartz, Shai, and Shai Ben-David. *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press, 2014.
25. Witten, Ian H., Eibe Frank, and Mark A. Hall. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, 2016.
26. Zhao, Zhihong, and Tongyuan Bai. “Financial Fraud Detection and Prediction Using Machine Learning Algorithms.” *Entropy*, vol. 24, no. 8, 2022.
27. Zhu, Shanshan, et al. “Financial Fraud Prediction Framework Based on Stacking Ensemble Learning.” *Expert Systems with Applications*, 2024.
28. Aggarwal, Charu C. *Data Mining: The Textbook*. Springer, 2015.
29. Géron, Aurélien. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O’Reilly Media, 2022.
30. Murphy, Kevin P. *Machine Learning: A Probabilistic Perspective*. MIT Press, 2012.

31. Tan, Pang-Ning, Michael Steinbach, and Vipin Kumar. *Introduction to Data Mining*. Pearson, 2018.
32. Jordan, Michael I., and Tom M. Mitchell. “Machine Learning: Trends, Perspectives, and Prospects.” *Science*, vol. 349, no. 6245, 2015, pp. 255–260.
33. LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. “Deep Learning.” *Nature*, vol. 521, no. 7553, 2015, pp. 436–444.
34. Lundberg, Scott M., and Su-In Lee. “A Unified Approach to Interpreting Model Predictions.” *Advances in Neural Information Processing Systems*, 2017.
35. Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin. “Why Should I Trust You? Explaining the Predictions of Any Classifier.” *Proceedings of the 22nd ACM SIGKDD Conference*, 2016.
36. Esteva, Andre, et al. “Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks.” *Nature*, vol. 542, no. 7639, 2017, pp. 115–118.
37. Chawla, Nitesh V., et al. “SMOTE: Synthetic Minority Over-Sampling Technique.” *Journal of Artificial Intelligence Research*, vol. 16, 2002, pp. 321–357.
38. Domingos, Pedro. *The Master Algorithm*. Basic Books, 2015.
39. Nilsson, Nils J. *The Quest for Artificial Intelligence*. Cambridge University Press, 2010.
40. Bishop, Matt. *Computer Security: Art and Science*. Addison-Wesley, 2018.