



Python-based Risk Classification Tool for Medical Devices

Abstract: This paper presents a Python-based risk classification tool specifically designed for medical devices, aimed at streamlining the assessment process and enhancing decision-making capabilities for healthcare professionals and regulatory bodies. The paper addresses existing regulatory gaps in Software as a Medical Device (SaMD) regulation, highlighting the challenges faced by stakeholders in ensuring compliance and safety and offers this tool as a suggestion for improvement. Leveraging advanced machine learning algorithms, the tool analyzes various risk factors associated with medical devices, including design, usage, and potential failure modes. By providing a scalable solution for risk classification, this work aims to bridge the gap between technological advancements and regulatory frameworks, ultimately fostering a safer healthcare environment.

Amina Zehra ¹, Tanisha Shetty ², Yashvi Tripathi ³, Dr. Narendra Kumar ⁴

Affiliation

^{1,2,3} B.E Student, Department. of Biotechnology,
RV College of Engineering, Bangalore, India

⁴Assistant Professor, Department. of
Biotechnology, RV College of Engineering,
Bangalore, India

Article History:

Received Date : Jun 15, 2025

Revised Date : Jun 30, 2025

Accepted Date : Jul 03, 2025

Published Date : Jul 11, 2025

Introduction

The complexity and diversity of medical devices, including Software as a Medical Device (SaMD), have increased as a result of the quick development of medical technology. With its standalone software that can perform medical functions like diagnosis, treatment planning, and monitoring without the use of traditional hardware, SaMD is transforming the healthcare industry [1]. Driven by developments in digital health and the integration of AI and ML, SaMD provides creative solutions, but also poses particular regulatory difficulties. Its rapid evolution necessitates quick oversight to assure safety, efficacy and reliability. From advanced diagnostic devices to mobile health tools, SaMD covers a broad spectrum of applications [2].

As these devices become more integral to patient care, the need for effective risk classification tools has become paramount. For patient safety, legal compliance, and efficient risk management throughout the lifecycle of medical devices, risk classification is crucial.

Various Regulatory bodies like as the U.S. Food and Drug Administration (FDA) and the European Medicines Agency (EMA), are established for the above purpose. Frameworks to assess the safety and efficacy of medical devices [3], [4]. Devices are divided into different classes according to traditional risk classification systems, with Class I devices representing the least danger and Class III devices representing the maximum risk [2]. However, the emergence of SaMD has introduced new challenges, as these software solutions can vary widely in functionality and risk profiles. The lack of standardized risk classification methodologies for SaMD has resulted in regulatory uncertainty and potential safety concerns. These gaps can lead to inconsistencies in risk assessment and hinder the timely approval of innovative technologies [5].

The motivation behind developing a Python-based risk classification tool stems from the need to address these regulatory gaps and enhance the efficiency of risk assessments for medical devices. Recent approaches majority rely on manual processes that can be time-taking and prone to human error [6]. By leveraging machine learning algorithms, this tool aims to automate the risk classification process, providing a more accurate and consistent approach to evaluating medical devices [7]. Furthermore, the tool seeks to empower healthcare professionals and regulatory bodies with actionable insights that can facilitate informed decision-making and improve patient safety [8].

This paper's main goal is to create a Python-based risk classification tool that analyzes risk indicators related to medical devices using machine learning methods. This



tool is designed to automate the risk assessment process, thereby improving accuracy and efficiency. Further, this paper aims to address the regulatory gaps present in SaMD regulations by providing a scalable solution that not only aligns with current regulatory frameworks but also enhances compliance. By achieving these objectives, we aspire to contribute significantly to the ongoing efforts to improve safety standards in medical devices and support regulatory bodies in adapting to the rapidly evolving technological landscape.

Literature Review

An extensive literature review was conducted by searching through the PUBMED database utilizing keywords such as SaMD, risk classification of medical devices and others in order to identify the various regulatory gray areas that have been identified and documented in current research to provide a clear understanding of the standing of SaMD in the clinical regulatory aspect.

A. Overview of existing tools and methodologies

A crucial step in deciding the regulatory approach for product approval and post-market surveillance in the medical device industry is risk classification. Medical devices are categorized into various classes by regulatory bodies like the European Medicines Agency (EMA) and the U.S. Food and Drug Administration (FDA) according to the risks they represent to patients and their intended usage. Class I (low risk), Class II (moderate risk) and Class III (high risk) are the three main categories into which devices are often categorized [2]. While Class III devices like implantable pacemakers, are subject to the strictest regulatory monitoring and require substantial clinical evidence to prove safety and efficacy, Class I devices, like tongue depressors, are subject to the least amount of regulation [9]. Because incorrect classification can result in serious patient safety hazards and regulatory issues, the classification process is crucial to ensure that devices are properly assessed and tracked throughout their lifecycle. However, the emergence of Software as a Medical Device (SaMD) has complicated this landscape, as these software solutions can vary widely in functionality and risk profiles, necessitating a more nuanced approach to risk classification

that considers both software- specific factors and traditional device characteristics [10].

SaMDs are now essential to clinical practice, and as such, the paperwork that goes along with them must be managed and controlled effectively. Many software applications are categorized at higher risk levels and must go through more complicated certification procedures as a result of the European Regulation MDR (EU) 2017/745's introduction of a specific classification rule (rule 11, Annex VIII) for software, which imposes stricter requirements than earlier regulations [11]. Software as Medical Devices (SaMD) standards are essential to the safe and efficient creation, testing, and implementation of these software programs. Apart from the Medical Device Regulation (MDR), the SaMD industry can benefit from a number of important standards, such as IEC 62304, which describes the primary software lifecycle procedures, such as problem analysis and change management [12]. The IEC 82304 standard is also relevant, addressing post-market activities and requirements [13].

Furthermore, standards such as ISO 14971 provide guidance on risk identification and control [14], while IEC 62366 focuses on design, implementation, and validation of usability requirements [15]. Additionally, maintaining quality in the software development context is essential, and standards such as ISO 13485 provide a framework for achieving this goal [16]. Stakeholders can guarantee that SaMDs are created and implemented in a way that satisfies legal requirements and puts patient safety first by following these guidelines. But while these standards support regulation of software lifecycle, risk classification remains a vague subject.

Several tools and methodologies have been developed to assist in the risk classification of medical devices. Traditional approaches often rely on expert judgment and manual processes, which can be time-consuming and prone to human error. Some existing tools utilize decision trees and rule-based systems to guide users through the classification process based on predefined criteria [17]. For instance, the FDA provides a 510(k) premarket notification process that includes a risk classification tool to help manufacturers determine the appropriate regulatory pathway for their devices [18]. Additionally, various software solutions have



emerged that incorporate machine learning techniques to enhance risk assessment capabilities. These tools analyze historical data and device characteristics to predict risk levels, offering a more data-driven approach to classification. However, while these methodologies have made strides in improving the efficiency of risk assessments, they often lack the flexibility and adaptability required to address the unique challenges posed by SaMD, which can lead to inconsistencies in risk evaluation and regulatory compliance [19].

B. Gaps in current approaches

Despite the advancements in risk classification tools and methodologies, several limitations persist that hinder their effectiveness, particularly in the context of SaMD. One significant challenge is the lack of standardized criteria for evaluating the risk associated with software-based devices, which can lead to variability in classification outcomes [20]. Many existing tools are primarily designed for traditional medical devices and do not adequately account for the unique aspects of software, such as algorithmic complexity, data privacy concerns, and the potential for cybersecurity threats [21]. Furthermore, the reliance on historical data for machine learning models can introduce biases, as these models may not generalize well to new or innovative devices that lack sufficient historical precedents [22]. Additionally, the manual nature of many current methodologies can result in delays in the approval process, ultimately impacting patient access to potentially life-saving technologies [23]. These limitations underscore the need for a more comprehensive and adaptable risk classification framework that can effectively address the complexities of modern medical devices, particularly in the rapidly evolving landscape of digital health and SaMD.

Methodology

A. System Architecture

The proposed Python-based risk classification tool for medical devices is structured around a modular pipeline comprising data ingestion, preprocessing, classification, and interpretability modules. This architecture supports automation and scalability while maintaining transparency—essential in regulatory environments. The input to the system

is structured metadata for various medical devices, including characteristics such as intended use, duration of contact, user environment, and invasiveness. The classification model processes this data to assign a risk category in accordance with recognized frameworks, such as those defined by the International

Medical Device Regulators Forum (IMDRF) [5] or regional authorities like the U.S. FDA or EU MDR [3], [4].

B. Data Collection and Pre-processing

A curated dataset of medical device descriptions and associated risk classes was assembled from publicly available regulatory databases and guidance documents. Each device entry includes attributes such as device type, contact duration, anatomical site of use, and regulatory classification. Preprocessing included handling missing values, standardizing text entries, and encoding categorical variables using LabelEncoder [24]. The data was divided into training and testing sets using a stratified approach to preserve the proportional distribution of risk classes, ensuring a fair and balanced evaluation of the classifier's performance.

C. Risk Classification Algorithm

A Decision Tree Classifier was implemented for categorizing medical devices into risk levels such as Class I, II, or III. Decision trees are particularly suited to this application due to their high interpretability and ability to handle different type of data types [25]. The algorithm employs the Gini impurity index as a splitting criterion to optimize decision boundaries. This choice allows regulators and developers to trace the reasoning behind risk assignments, a critical requirement for compliance and auditability in the medical device industry. Hyperparameters such as maximum tree depth and minimum samples per leaf were tuned using k-fold cross-validation to avoid overfitting while maintaining generalizability [26].

D. Implementation in Python

Python was used to develop the classification tool, with scikit-learn for model building, matplotlib for display, and pandas for

data manipulation. Using scikit-learn's DecisionTreeClassifier, the model was built, and classification attributes were numerically encoded using LabelEncoder [27]. Metrics like accuracy, confusion matrix, and F1-score were used to assess the model's performance. Plot tree, which offers a graphical depiction of decision rules and improves explainability for regulatory assessment, was used to illustrate the trained decision tree [17]. To guarantee usability, reproducibility, and cooperation between software developers, regulatory analysts, and medical specialists, the complete codebase is housed on Google Colab.

Results and Discussion

The proposed risk classification model was evaluated on a structured dataset representing medical device features aligned with regulatory classification guidelines. The Decision Tree Classifier obtained an overall accuracy of 83.33% on the test set following data preparation and model training. Detailed performance metrics including precision, recall, and F1-score for each class are presented in Table I.

The model's performance for each risk class is thoroughly summarized in the classification report using the following important metrics: precision, recall, f1-score, and support. Precision measures how well a model avoids false positives by dividing the number of true positive predictions by the total number of expected positives for a class. The ratio of true positives to all actual instances of a class is called recall (sometimes called sensitivity), and it shows how well the model captures all pertinent instances. The harmonic mean of precision and recall, or f1-score, provides a fair way to assess model performance. Finally, support shows how many real examples of each class there are in the test dataset.

The model's performance is constant across all three risk groups, according to the results. With the greatest f1-score, Class II demonstrates a consistent trade-off between recall and precision. This is to be expected since, in contrast to borderline cases in Class I or Class III, moderate-risk devices frequently exhibit more distinct feature patterns. Class III has a slightly poorer recall, but this is acceptable for early-stage risk instruments because, from a

regulatory standpoint, overestimating risk is better than underestimating it. All classes exhibit consistent performance, according to the classification report (see Table I). Despite variations in class distribution, the model's balanced classification ability was demonstrated by its macro-averaged F1-score of 0.83.

The balanced performance metrics support the model's robustness in multi-class classification and demonstrate its potential utility for initial risk classification in compliance with IMDRF and EU MDR guidelines.

Table I Classification Report [Test Data]

Class	Parameters accounted		
	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
0 [Low Risk]	0.75	0.75	0.75
1 [Moderate Risk]	0.83	0.91	0.87
2 [High Risk]	0.89	0.80	0.84

A confusion matrix is used to assess the decision tree model's classification performance on the test dataset, as shown in Fig. 1. With rows representing true labels and columns representing predicted labels, this matrix contrasts anticipated risk classes with actual known classes. The diagonal shows correct predictions, whereas off-diagonal parts show misclassifications, giving information about the accuracy and error trends of the model [17].

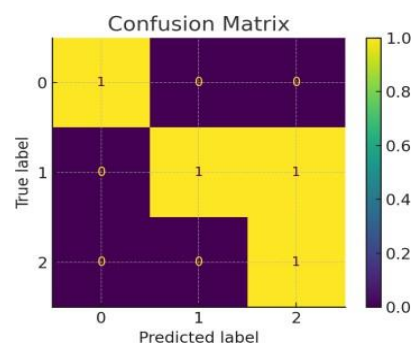


Fig. 1. Confusion Matrix

As seen in the figure, the model demonstrates a strong ability to correctly classify most devices, especially in

Class II (Moderate Risk), where the majority of predictions align with the true labels. Misclassifications primarily occur between adjacent risk classes, such as Class I and Class II or Class II and Class III, which is a reasonable outcome considering the real-world overlap in features like invasiveness and duration of contact. Importantly, no high-risk (Class III) devices were misclassified as low-risk (Class I), which reinforces the model's safety in regulatory contexts. This behavior ensures that more stringent regulatory requirements are not erroneously bypassed due to under-classification.

A decision tree is generated using core device-related features that influence regulatory risk classification as shown in Fig. 2. The model uses features such as *Device_Type*, *Contact_Duration*, *Invasiveness*, and *Use_Environment* to predict the appropriate risk class (Class I, II, or III). Each node in the tree represents a binary decision rule based on these features, and the branches indicate the logical path followed depending on the feature value. The leaf nodes at the bottom of the tree indicate the predicted risk class, with color intensity representing the model's confidence (i.e., class purity) at that node [28].

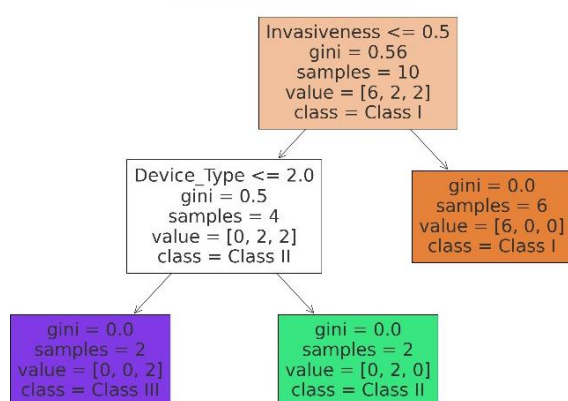


Fig. 2. Decision Tree Visualization based on Risk Magnitude

The structure of the tree aligns well with international medical device classification frameworks, such as the IMDRF principles and the European Medical Device Regulation (EU MDR) Annex VIII. For instance, devices characterized as long-duration, invasive, and

used in clinical environments are correctly routed toward higher-risk classifications (Class III). Conversely, non-invasive, short-duration, and home-use devices are often assigned to Class I. This interpretability is crucial for regulatory transparency, as it enables manufacturers and reviewers to understand the rationale behind each classification decision, thereby supporting traceability and conformity assessment documentation.

The decision tree serves not only as a predictive model but also as an explanatory tool that reflects real-world decision criteria embedded in international regulations. Its logical flow mirrors the classification logic outlined in regulatory guidelines, making it a practical aid in early product development, premarket submission, and internal risk assessment processes [25].

Discussion

One of the key advantages of using a decision tree for this application lies in its interpretability. The decision tree visualization clearly shows how different device features—such as duration of use, invasiveness, and user environment—affect classification outcomes. For example, the model identified long-term internal use and implantable devices as strong indicators for high-risk classification, which aligns with regulatory expectations [29].

This interpretability supports regulatory submissions, internal quality assurance reviews, and risk management documentation under ISO 14971. Additionally, the tool can serve as an initial triaging mechanism for manufacturers preparing premarket submissions, reducing manual burden and increasing classification consistency.

Future Scope and Conclusion

While the model performed well, it has limitations that warrant further investigation. First, the dataset size and scope were limited to publicly available device metadata; expanding this to include more granular features (e.g., device material, target population) could improve model accuracy. Second, the model is currently rule-based and static; integrating ensemble methods like Random Forest or Gradient Boosted Trees could improve generalization and handle class imbalance more effectively. In the future, we plan to integrate

external ontologies such as the Global Medical Device Nomenclature (GMDN) and harmonized classification codes to enrich feature representation and improve classification granularity. Furthermore, real-time risk prediction interfaces can be built atop this tool using lightweight Python web frameworks for deployment in regulatory and clinical settings [30].

This paper presents a Python-based risk classification tool for medical devices that leverages a decision tree algorithm to categorize devices into regulatory risk classes. The system demonstrates strong performance, achieving an 83.33% classification accuracy and maintaining explainability—an essential requirement in the highly regulated domain of medical device development. The model's interpretability, ease of deployment, and compliance with international classification guidelines position it as a valuable asset for regulatory teams and manufacturers. By enabling rapid, transparent, and repeatable risk classification, this tool has the potential to streamline the device approval pipeline and enhance decision-making processes during premarket assessment.

Future enhancements will focus on expanding the feature set, integrating more advanced machine learning techniques, and building user-facing applications to broaden its practical impact.

Acknowledgement

The authors are grateful to the Head of Biotechnology Department, RVCE, Bengaluru, for providing valuable guidance and mentorship during the research process. We would also like to express gratitude to Mr. Senthil and Mr. Sarat Kumar (Manager, Central RA, Baxter) for their continuous support and guidance in the fulfillment of this project.

References

- [1] U.S. Food and Drug Administration, "Proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD)," 2019
- [2] A. K. Gupta, "Navigating the Regulatory Landscape of Software as a Medical Device (SaMD): Compliance Challenges, Best Practices, and Future Trends," ResearchGate, 2020. [Online]. Available:
- [3] U.S. Food and Drug Administration, "Medical Device Overview," 2021. [Online]. Available: <https://www.fda.gov/medical-devices>. [Accessed: Jun. 3, 2025].
- [4] European Medicines Agency, "Guideline on the regulation of medical devices," 2020. [Online]. Available: https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-regulation-medical-devices_en.pdf. [Accessed: Jun. 3, 2025].
- [5] D. Tobey, "IMDRF publishes draft software risk classification document," DLA Piper, Feb. 23, 2024. [Online]. Available: <https://www.dlapiper.com/en/insights/publications/2024/02/diagnostic-landscape-complex-fda-rule>. [Accessed: Jun. 3, 2025].
- [6] M. Haimerl and C. Reich, "Risk-based evaluation of machine learning-based classification methods used for medical devices," BMC Med. Inform. Decis. Mak., vol. 25, no. 126, 2025. doi: 10.1186/s12911-025-02909-9.
- [7] D. Lyell, E. Coiera, J. Chen, P. Shah, and F. Magrabi, "How machine learning is embedded to support clinician decision making: an analysis of FDA-approved medical devices," BMJ Health Care Inform., vol. 28, no. 1, p. e100301, Apr. 2021. doi: 10.1136/bmjhci-2020-100301.
- [8] U.S. Food and Drug Administration, "Artificial Intelligence and Machine Learning in Software as a Medical Device," 2021. [Online]. Available: <https://www.fda.gov/media/145022/download>. [Accessed: Jun. 3, 2025].
- [9] T. Melvin, "The European Medical Device Regulation - What Biomedical Engineers Need to Know," IEEE J. Transl. Eng. Health Med., vol. 10, Art. no. 3194415, pp. 1–10, 2022. doi: 10.1109/JTEHM.2022.3194415.
- [10] A. Lal et al., "Regulatory oversight and ethical concerns surrounding software as medical device (SaMD) and digital twin technology in healthcare," Ann. Transl. Med., vol. 10, no. 18, pp. 950, 2022. doi: 10.21037/atm-22-4203.
- [11] M. Bretthauer, S. Gerke, C. Hassan, O. F. Ahmad, and Y. Mori, "The New European Medical Device Regulation: Balancing Innovation and Patient Safety," Ann. Intern. Med., vol. 176, no. 6, pp. 844–848, 2023. doi: 10.7326/M23-0454.
- [12] International Electrotechnical Commission, IEC 62304:2006 - Medical Device Software - Software Life Cycle Processes, 2006. [Online]. Available: <https://webstore.iec.ch/publication/6170>. [Accessed: Jun. 3, 2025].
- [13] International Electrotechnical Commission, IEC 82304-1:2016 - Health Software - Part 1: General Requirements for Product Safety, 2016. [Online]. Available:



- <https://webstore.iec.ch/publication/6170>.
[Accessed: Jun. 3, 2025].
- [14] International Organization for Standardization, ISO 14971:2019 - Medical Devices - Application of Risk Management to Medical Devices, 2019. [Online]. Available: <https://www.iso.org/standard/72704.html>. [Accessed: Jun. 3, 2025].
- [15] International Electrotechnical Commission, IEC 62366-1:2015 - Medical Devices - Part 1: Application of Usability Engineering to Medical Devices, 2015. [Online]. Available: <https://webstore.iec.ch/publication/6170>. [Accessed: Jun. 3, 2025].
- [16] International Organization for Standardization, ISO 13485:2016 - Medical Devices - Quality Management Systems - Requirements for Regulatory Purposes, 2016. [Online]. Available: <https://www.iso.org/standard/59752.html>. [Accessed: Jun. 3, 2025].
- [17] A. Ceross and J. Bergmann, "A Machine Learning Approach for Medical Device Classification," in Proc. 14th Int. Conf. Theory Pract. Electron. Govern., 2021, pp. 285–291. doi: [10.1145/3460097.3460111](https://doi.org/10.1145/3460097.3460111).
- [18] U.S. Food and Drug Administration, "510(k) Premarket Notification," 2021. [Online]. Available: <https://www.fda.gov/medical-devices/how-study-and-market-your-device/510k-premarket-notification>. [Accessed: Jun. 3, 2025].
- [19] U. J. Muehlematter, P. Daniore, and K. N. Vokinger, "Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015–20): a comparative analysis," Lancet Digit. Health, vol. 3, no. 3, pp. e195–e203, 2021. doi: [10.1016/S2666-7568\(20\)30264-4](https://doi.org/10.1016/S2666-7568(20)30264-4).
- [20] L. Svempe, "Regulation and its impact on innovation in healthcare: SAMD case," Socrates, vol. 22, no. 1, pp. 43–52, 2022. doi: [10.25143/socr.22.2022.1.043-052](https://doi.org/10.25143/socr.22.2022.1.043-052).
- [21] European Commission, "Guidance on Qualification and Classification of Software in Regulation (EU) 2017/745 – MDR and Regulation (EU) 2017/746 – IVDR," 2019. [Online]. Available: <https://ec.europa.eu/docsroom/documents/37581>. [Accessed: Jun. 3, 2025].
- [22] L. Peter, L. Hájek, P. Maresov, M. Augustynek, and M. Penhaker, "Medical Devices: Regulation, Risk Classification, and Open Innovation," J. Open Innov. Technol. Mark. Complex., vol. 6, no. 2, p. 42, 2020. doi: [10.3390/joitmc6020042](https://doi.org/10.3390/joitmc6020042).
- [23] E. Balogun and C. Luetge, "Gap Analysis on Regulatory Frameworks for AI-embedded Medical Devices," in 2025 IEEE Symp. Trustworthy, Explainable and Responsible Comput. Intell. (CITREx Companion), Trondheim, Norway, 2025, pp. 1–5. doi: [10.1109/CITRExCompanion65208.2025.10981493](https://doi.org/10.1109/CITRExCompanion65208.2025.10981493).
- [24] Scikit-learn, "LabelEncoder," 2021. [Online]. Available: <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.LabelEncoder.html>. [Accessed: Jun. 3, 2025].
- [25] D. Straub and I. Welp, "Decision-making under risk: a normative and behavioral perspective," in Risk – A Multidisciplinary Introduction, C. Klüppelberg, D. Straub, and I. M. Welp, Eds. Cham, Heidelberg: Springer, 2014, pp. 63–93.
- [26] S. Yadav and S. Shukla, "Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification," in Proc. 2016 IEEE 6th Int. Conf. Adv. Comput. (IACC), 2016, pp. 78–83.
- [27] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," J. Mach. Learn. Res., vol. 12, pp. 2825–2830, 2011.
- [28] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed. New York, NY: Springer, 2009.
- [29] C. Molnar, Interpretable Machine Learning: A Guide for Making Black Box Models Explainable, 1st ed., 2020. [Online]. Available: <https://christophm.github.io/interpretable-ml-book/>. [Accessed: Jun. 3, 2025].
- [30] Flask, "Flask Documentation," 2021. [Online]. Available: <https://flask.palletsprojects.com/>. [Accessed: Jun. 3, 2025].