Erciyes Üniversitesi
Fen Bilimleri Enstitüsü Dergisi
Cilt 40, Sayı 3, 2024

Erciyes University
Journal of Institue Of Science and Technology
Volume 40, Issue 3, 2024

Classification of System and Software Related FDA Medical Device Recalls with a Three-Level Taxonomy Approach: Defibrillator Case

Nazire Merve ÖNDER*[1,2], Özden ÖZCAN TOP[1]

*[1] Middle East Technical University, Information Systems Dept., ANKARA, TÜRKİYE
[2] ASELSAN A.Ş., ANKARA, TÜRKİYE

(Alınış / Received: 24.11.2024, Kabul / Accepted: 19.12.2024, Online Yayınlanma / Published Online: 30.12.2024)

Keywords
Defect classification,
Defect taxonomy,
Medical device recall,
Defibrillator,
FDA,
SW91.

Abstract: In the healthcare domain, where safety is paramount, medical device recalls are highly critical events that can pose significant risks to health. The U.S. Food and Drug Administration (FDA) maintains a comprehensive database of recall data, including fields such as product description, product type, recall reason, and termination status. However, the recall reason field lacks a predefined taxonomy for root cause classification, which limits the effectiveness of analyzing and preventing recurring issues. In this study, we aim to classify recalls based on their reasons by proposing a three-level taxonomy to improve the classification process. The first level classification approaches to the problem from a broad perspective, including labels such as operational, environmental, electrical, hardware, software, and mechanical. The second level utilizes the FDA root cause options to provide a more detailed classification. The third level uses the SW91 Classification of Defects in Health Software taxonomy specifically for software, which offers a standardized framework for classifying defects in health software. This taxonomy approach would not only aid in more precise recall classification but also lays the groundwork for subsequent research focused on developing predictive models to prevent recurring defects in healthcare software. By improving classification accuracy, this study aims to increase ways for product safety and enhance regulatory oversight in the medical device domain. In this study, we applied this approach to 271 different defibrillator recall cases reported by the FDA and categorized them based on their recall reasons. This enabled us to better understand recurring issues and allow for a more precise classification of recall reasons. This approach not only aims to analyze and address existing problems more effectively but also seeks to provide a foundation for future research focused on predicting and preventing potential defects.

Sistem ve Yazılım Kaynaklı FDA Tıbbi Cihaz Geri Çağırmalarının Üç Seviyeli Taksonomi Yaklaşımı ile Sınıflandırılması: Defibrilatör Vakası

Anahtar Kelimeler
Hata sınıflandırma,
Hata taksonomisi,
Tıbbi Cihaz Geri Çağırma,
Defibrilatör,
FDA,
SW91.

Öz: Emniyetin her şeyden önemli olduğu sağlık alanında, tıbbi cihaz geri çağırmaları sağlık için önemli riskler oluşturabilecek son derece kritik olayları tanımlar. ABD Gıda ve İlaç Dairesi (FDA) ürün tanımı, ürün türü, geri çağırma nedeni ve sonlandırma durumu gibi alanları içeren kapsamlı bir geri çağırma veritabanı tutmaktadır. Ancak, geri çağırma nedeni alanı, kök neden sınıflandırması için önceden tanımlanmış bir taksonomiden yoksundur ve bu da tekrar eden sorunların analiz edilmesinin ve önlenmesinin etkinliğini sınırlamaktadır. Bu çalışmada, sınıflandırma sürecini iyileştirmek için üç seviyeli bir taksonomi önererek geri çağırmaları nedenlerine göre sınıflandırmayı amaçladık. Birinci seviye sınıflandırma, soruna operasyonel, çevresel, elektriksel, donanımsal, yazılımsal ve mekanik gibi etiketleri içeren geniş bir perspektiften yaklaşmaktadır. İkinci seviye, daha ayrıntılı bir sınıflandırma sağlamak için FDA kök neden seçeneklerini kullanır. Üçüncü seviye, ise sağlık yazılımlarındaki kusurların sınıflandırılması için standart

bir çerçeve sunan SW91 Sağlık Yazılımları için Hata Sınıflandırma taksonomisini kullanmaktadır. Bu çalışmada, bu yaklaşımı FDA tarafından geri çağırılan 271 farklı defibrilatör vakası için uyguladık ve vakaları geri çağırılma nedenlerine göre kategorize ettik. Bu sayede tekrar eden sorunları ve geri çağırma nedenlerinin daha hassas bir şekilde sınıflandırılmasına olanak tanıdık. Bu yaklaşım, sadece mevcut sorunları daha etkili bir şekilde analiz etmeyi ve çözmeyi değil, aynı zamanda ileride oluşabilecek hataları öngörmeye ve önlemeye yönelik araştırmalar için de bir temel oluşturmayı hedeflemektedir.

1. Introduction

Medical devices are safety critical in nature and their production and sale is governed by strict safety rules. Malfunctioning medical devices can lead to significant injuries and even death. When a company discovers an issue with one of its medical devices, it must propose either a correction or a removal based on the situation to regulatory bodies depending on where the medical device is marketed. Correction involves addressing the issue with the medical device at the location where it is used or sold. Removal addresses taking the medical device out of circulation from where it is used or sold. The Food and Drug Administration (FDA) uses the term "recall" when a manufacturer makes a correction or remove the product to address a problem with a medical device that violates FDA laws. Recalls occur when a medical device manufacturer notifies the FDA of a defect that could be a risk to health, or when it is both defective and a risk to health [1].

The FDA classifies recalls into three categories based on health risks posed by device failure. Safety Class I recalls denote situations in which exposure to a "violative product will cause serious adverse health consequences or death" [2]. While Class II exposure may result in temporary or reversible adverse health conditions, Class III recalls only reflect regulatory volitions with minimal or no health risks [3].

The FDA keeps all recall data related to medical devices since 2002, which is publicly accessible on its website [4][10]. Recall reasons can be categorized into several types, including product defects, which involve manufacturing errors or contamination; labeling errors, where there is misleading or incorrect information on the packaging; performance issues, where devices fail to meet safety or efficacy standards; and compliance issues, which involve violations of regulatory requirements. Users can readily access statistics on recalls that are categorized by recall safety class, product type, year, and other relevant criteria. Statistics also shows that medical device recalls are increasing each year [1]. Therefore, recall data can be used as a valuable information for identifying the underlying causes of recalls and implementing preventive measures from the development stage discriminate the potential causes of recalls and take preventive actions during the development process. However, the effectiveness of this approach depends significantly on the level of detail in the defect taxonomy used as it directly affects the accuracy of classification and analysis.

Another challenge arises from the manually entered fields in the recall database. On the FDA website [4], recalls are published based on a predefined template that include structured fields: Date, Brand Names, Product Description, Product Types, Recall Reason Description, Company Name, and Terminated Recall. However, only the Recall Reason Description field is manually written. This reliance on human input introduces variability and inconsistencies, making it difficult to systematically analyze recall reasons. However, examining these human-written descriptions, along with adverse event reports, is essential, as it can uncover underlying safety issues in these devices. Such analysis not only highlights current safety concerns but also offers important insights into potential challenges in the design and development of safety-critical medical devices [5].

In this research, to address the aforementioned issues in defect classification, we introduce a three-level taxonomy designed to provide more precise and systematic recall classification, enabling better identification of safety issues with respect to recall reasons.

We selected defibrillators as the focus of this study because they represent a critical category of medical devices with a high frequency of recalls. Their importance has also been particularly underlined by their extensive use during the COVID-19 pandemic.

Particularly, our aim is to present recall data obtained from the industry, categorized across different levels of defect classification. By examining this data, we present the crucial role of accurate defect classification in ensuring the safety and reliability of medical device development. Additionally, we offer recommendations for improving how recall root causes are described, which can contribute to enhancing safety protocols and supporting the design of safer medical devices in the future.

To achieve these objectives, we first retrieved recall data from the FDA Database, filtering it by device type, specifically defibrillators to create a focused dataset. We then reviewed existing defect classification methods in the literature, selecting those detailed in Section II as the most appropriate for our study. Using these methods, we applied our classification approach to 271 defibrillator recalls, systematically categorizing them to reveal patterns and trends. Finally, we presented the results of this analysis, offering insights that can inform future improvements in defect management and recall prevention strategies.

The rest of the paper is structured as follows: Section 2 describes related works in the field. Section 3 explains the research approach used in this study. Section 4 addresses the results and analysis of defect classification at different levels. Section 5 addresses the discussions. Finally, Section 6 gives the conclusions.

2. Background and Related Work

Bu The FDA is a highly critical regulatory agency in the United States, responsible for ensuring the safety and effectiveness of food products, pharmaceuticals, and medical devices. One of its important functions is to regulate the recall of medical devices, which occurs when a manufacturer takes a correction or removal action to address a problem due to safety concerns or defects that may present health risks to consumers. The FDA keeps all the records of these recalls to facilitate public awareness, enhance consumer safety, and support ongoing surveillance of medical devices, contributing to improved regulatory practices and promoting accountability among manufacturers [4]. FDA keeps the records of all medical device recalls since 2002, data of previous years are partially retained. The database is accesible from FDA Recall Website [4].

There are also other databases from different countries which keeps records of medical devices sold in their countries. European Database on Medical Devices (EUDAMED) in Europe [6], Therapeutic Goods Administration (TGA) [7] from Australia, Medical Device Incident Reporting System (MDIRS) [8] again from Australia are a few examples of these databases. However, the most comprehensive database is maintained by the FDA.

Classification is a systematic methodology for organizing entities into categories based on shared characteristics or attributes. This process is pivotal across various disciplines, including taxonomy in biology, data science, and library science, as it aids in the comprehension and retrieval of information.

2.1. Defect Classification based on Industry Guidence Documents

Industry guidance documents in the context of medical devices are comprehensive resources developed by regulatory authorities to assist manufacturers, stakeholders, and researchers in understanding the regulatory requirements and best practices for the design, development, and commercialization of medical devices.

For the classification process, industry guidance documents are examined to find a predefined root cause taxonomy as an initial step in this study. A guidance which propose a high level taxonomy for sources of hazardous situations is found in *Infusion Pumps Total Product Life Cycle Guidance for Industry and FDA Staff* [9]. These predefined defect sources are given in Table 1.

Table 1. Defect sources based on infusion pump guidance

No	Defect Source
1.	Operational
2.	Environmental
3.	Electrical
4.	Hardware
5.	Software
6.	Mechanical
7.	Biological and Chemical
8.	Use

These labels are further explained with examples of hazards and potential causes in the guidence to help classification [9]. The unintended operation of pump motors, loose connection between delivery parts, broken valves, among other factors, are classified as “Operational” defect sources. Battery leakage, electromagnetic inteference related defects, contamination with toxins, tempereture/humidity/air pressure level thresholds are taken as “Environmental” sources. Loose connections between devices, cooling defects, overcharged/undercharged battery, exceeding supply limits are considered as “Electrical” sources. Discommunuciations, sensor failures, noncalibrated devices and sensors, synchronization errors are classified ad “Hardware” sources. Memory and buffer problems, runtime errors, errors arise from updates and version control, data storage or retrieval problems, library related problems are regarded as “Software” sources. Broken or damaged devices, power cords, motors are classified as “Mechanical” sources. Sterilization problems, local reactions, material damages, contamination problems are considered as “ Biological and Chemical”. User friendliness of softwares, insufficient training,

accidents caused by users, confusinh instructions for use are deemed “Use” sources. These labels are assessed as usable for the initial level of classification.

2.2. Defect Classification based on the FDA Defect Root Causes

We also explored the FDA Recall Search Database [10] to specify any existing predefined root cause options. There are predefined root cause lables which are used as dropdown list on th FDA Recall Database [4]. These predefined possible root causes are given in Table 2.

Table 2. FDA root causes retrieved from [10]

No	Recall Root Cause Types
1.	Component Change Control
2.	Component Design/Selection
3.	Counterfeit
4.	Device Design
5.	Employee error
6.	Environmental Control
7.	Equipment Maintenance
8.	Error in labeling
9.	Finished Device Change Control
10.	Incorrect or No Expiration Date
11.	Labeling Change Control
12.	Labeling False and Misleading
13.	Labeling Design
14.	Labeling mix-up
15.	Manufacturing material removal
16.	Material/Component Contamination
17.	Mixed up of materials/components
18.	No marketing application
19.	Nonconforming material/component
20.	Vendor Change Control
21.	PMA
22.	Packaging
23.	Package Design/Selection
24.	Packaging Change Control
25.	Packaging Process Control
26.	Pending
27.	Process Change Control
28.	Process control
29.	Process Design
30.	Radiation Control for Health and Safety Act
31.	Release of Component/Material prior to receiving test results
32.	Reprocessing Controls
33.	Software Design Change
34.	Software manufacturing/software deployment
35.	Software change control
36.	Software Design
37.	Software Design (Manufacturing Process)
38.	Software in the use environment
39.	Storage
40.	Under Investigation by Firm
41.	Unknown/undetermined by firm
42.	Use error
43.	Other

The Root Cause field in the FDA Database is not mandatory when publishing recalls; consequently, there is limited recall data available for this selection. The handwritten recall reason field provides significantly more information regarding root causes; however, it cannot be utilized for automatic classification and/or filtration. Therefore, the possible root causes identified in FDA Recall Search Database [10] are used as labels for the recalls in this study.

According to researches, four out of every ten medical devices incorporating software have failed due to a problem in the software itself, while compared to the total FDA MD recalls this reaches 18.3% of software failures during this period [11]. The current analysis of recalls has revealed a significant increase in software failures over the past decade. Consequently, a taxonomy to classify software-related causes of recalls was examined, leading to the identification of studies that propose a taxonomy known as “SW91- Classfication of Defects in Health Software”[12].

2.3. Defect Classification based on Software Defect Classification in Health Software

ANSI/AMMI SW91:2018 A Framework for Taxonomy Based Testing Using Classfication of Defects in Health Software [12] is a recognized standard that provides a structured approach for categorizing and managing defects in health software systems. It includes multi-level defect categories such as parent level and child level from planning to maintenance phase of software development process. There are 194 defect category in total, each defect category has a unique defect code which is part of the hierarchical structure [13].

3. Methodology

In this study, we employed a qualitative research methodology that involved a series of steps to analyze and classify FDA recall data for defibrillators. The process began with the collection and filtering of recall data from the FDA database, focusing specifically on defibrillators. The filtered data was then categorized using a three-level classification framework designed to provide a comprehensive understanding of the recall reasons.

The first level of classification was based on the criteria outlined in [9], which provided a broad categorization of defects. The second level refined this classification using more specific guidelines from [10], allowing for a deeper examination of the issues. Finally, the third level applied a detailed classification based on ANSI/AMMI SW91:2018 [12], which aimed to pinpoint the underlying causes of the recalls with greater precision in software components of medical devices.

After completing the classification process, the results were thoroughly evaluated and verified by two academic experts, ensuring accuracy and reliability. Each of these methodological steps is further detailed in the following paragraphs, providing a clear and structured explanation of our approach.

3.1. Case Data Collection Process

To implement this three-level classification method, it was essential to select a specific medical device that aligns with the objectives of the study. This device class should possess the required features and capabilities to permit accurate data collection and analysis applicable to the classification framework. For this assessment, effects of COVID-19 disease were also addressed. The COVID-19 pandemic had harmful effects on prehospital emergency care worldwide, such as decreasing bystander cardiopulmonary resuscitation (CPR), and increasing delays in emergency medical service (EMS) response time [14]. Defibrillators are critical devices detect sudden, dangerous heart rhythms or a cardiac arrest with many types allowing deep analysis of recalls, so defibrillators were chosen to implement this classification method.

To gather the required data, we initially retrieved 495 recall records for defibrillator-type medical devices from the FDA database [15], covering the period from January 2002 to January 2021. Figure 1 below shows the yearly distribution of the recalled 495 defibrillators. The records were exported into an MS Excel sheet. Each record included columns for Date, Brand Names, Product Description, Product Types, Recall Reason Description, Company Name, and Terminated Recall status. Out of these 495 records, 213 records lacked sufficient information and were excluded from the dataset. This resulted in 282 recall records. During a second review of the data, 11 additional recalls were found to have insufficient information for classification and were removed. Ultimately, 271 recall records remained available for applying the three-level classification method.

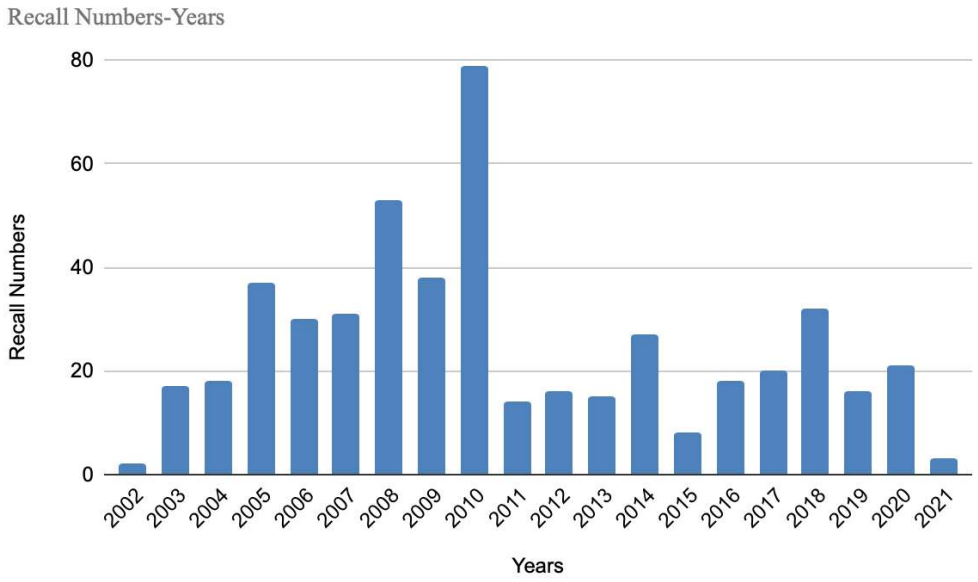

Figure 1. Yearly Distribution of Defibrillator Recalls from 2002 to 2021

Available defibrillator types are Dc-Defibrillator, High Energy, (Including Paddles), Tester, Defibrillator, Dc-Defibrillator, Low-Energy, (Including Paddles), Implantable Cardioverter Defibrillator (Non-Crt), Automated External Defibrillators (Non-Wearable), Atrial Defibrillator, Auxiliary Power Supply (Ac Or Dc) For Low-Energy Dc-Defibrillator, Defibrillator, Implantable, Dual-Chamber, Wearable Automated External Defibrillator, Automatic Implantable Cardioverter, With Cardiac, Over-The-Counter Automated External Defibrillator, Permanent Defibrillator Electrodes. This broad selection ensured that a wide array of defibrillator-related recalls could be analyzed.

3.2. Classification Steps

To begin the classification process, we used predefined root causes from the "Pumps Total Product Life Cycle Guidance for Industry and FDA Staff" [9]. All 271 defibrillator recalls were analyzed according to the handwritten manufacturer recall reason field. This stage of the study took approximately 10 person-days to complete.

For the second level classification a deeper analysis is needed. Existing predefined root cause options from FDA Recall Search Database [10] were used as the second level classification labels. This study lasted about 10 person-days and noted that there are different classifications of root causes compared to those provided to the FDA.

The third level of classification focused specifically on recalls identified as software-related. For this, we applied a taxonomy known as "SW91 - A Framework for Taxonomy Based Testing Using Classification of Defects in Health Software" [12]. This allowed for a more precise classification of software-related defects, and this stage of the study took about 15 person-days.

After completing the classification process, all data and results were reviewed by two academic experts with over 20 years of experience in software development. Based on their feedback, approximately 7% of the classifications were adjusted, and the results were subsequently updated to ensure accuracy.

3.3. Validation of the Classification

The defect classification process was initially conducted by the first author of this research. To ensure accuracy and reliability, the classification was reviewed and validated by two academic experts in the software field, both with over 15 years of experience. During the validation phase, each expert reviewed a subset of 60 defects out of the total 271 defects classified. This subset included 30 defects that were common to both experts, facilitating a direct comparison of their assessments. Consequently, 90 unique defects and associated classification were reviewed collectively. It was found that there was a 70% agreement in the experts' classifications. For the remaining 30% where discrepancies occurred, discussions were held among all three individuals until a consensus was reached. Subsequently, the first author revised the classifications for the remaining 180 items based on the insights gained from these discussions. This comprehensive review and revision process ensured refining the classification outcomes and enhanced both the accuracy and reliability of the results.

4. Results and Analysis

4.1. First Level Classification: Insights from the first level classification are summarized in Table 3. The category with the highest number of recalls was "Hardware," highlighted in bold, which is expected since much of the data was collected during periods when software played a less critical role in device functionality. The second most frequent category was "Software," also highlighted in bold. Notably, the proportion of recalls attributed to software has shown an increasing trend in recent years, reflecting the growing importance of software in medical devices.

Table 3. Level 1 classification results

Count of Level 1 FDA	
Level 1 FDA	**Counts**
Biological and Chemical Sources (BCS)	17
Electrical (ELC)	20
Environmental (ENV)	7
Hardware (HW)	**106**
Mechanical (MC)	16
Operational (OP)	15
Software (SW)	**58**
Use	9
Not Applicaible (NA)	23
Grand Total	**271**

As seen from the results, the data at this level of classification is not highly detailed but is still crucial for understanding the primary cause of the issues. The counts provide a general overview of where most problems originate, which can help guide further investigation and corrective actions.

4.2. Second Level Classification: The results of the second level classification are presented in Table 4, with the most frequently occurring categories highlighted in bold. During this process, 53 recalls were reclassified differently from the recall reason field listed in the FDA database, indicating inconsistencies in how FDA data is categorized. The *"Software Design*" category accounted for 49 recalls, making it the second most common reason after *"Non-Conforming Material/Component",* which had the highest count. The third most frequent category was "*Device Design"* which covered various design-related aspects. Due to the prominence of software-related issues, an additional, more detailed classification was conducted for errors categorized under "Software" at the third level. The "Number of Different Root Causes" column in the table represents the number of errors we identified as distinct from the root cause specified in the FDA database.

Table 4. Level 2 classification results

Level 2 Classification Types and Their Counts		Level 1 Classification Types and Counts									
		BCS	**ELC**	**ENV**	**HW**	**MC**	**OP**	**SW**	**Use**	**NA**	**Number of Different Root causes**
Component Change Control	1	-	-	-	1	-	-	-	-	-	-
Component Design/Selection	31	-	1	-	30	-	-	-	-	-	4
Device Design	**47**	**-**	**12**	**3**	**18**	**8**	**2**	**-**	**2**	**2**	**6**
Employee Error	6	-	-	-	5	1	-	-	-	-	1
Environmental Control	1	-	-	1	-	-	-	-	-	-	
Environmental Control / Device Design	1	-	-	1	-	-	-	-	-	-	1
Environmental Control / Design	2	-	-	2	-	-	-	-	-	-	2
Labeling Design	1	-	-	-	-	-	-	-	1	-	-
False and Misleading Labeling	10	2	-	-	-		8	-	-	-	8
Material / Component Contamination	7	7	-	-	-	-	-	-	-	-	-
Mixed-up of Materials / Components	1	-	-	-	1	-	-	-	-	-	-
Nonconforming Material / Component	**58**	**1**	**6**	**-**	**50**	**1**	-	-	-	-	**11**
Other (end of service life)	1	-	-	-	-	-	1	-	-	-	1
Other (guide change)	2	-	-	-	-		1	-	1		1
Packaging	2	2	-	-	-	-	-	-	-	-	-
Packaging Change Control	2	-	-	-	-	-	-	-	2	-	-
Packaging Process Control	1	-	-	-	-	-	-	1	-	-	-
Premarket Approval	7	-	-	-	-	-	-	-		7	-
Process Change Control	1	-	-	-	-	-	-	1	-	-	-
Process Control	23	3			1	1	2	-	2	14	-
Process Design	6		1			5	-	-	-	-	-
Software Change Control	3	-	-	-	-	-	-	3	-	-	2
Software Design	**49**	**-**	**-**	**-**	**-**	**-**	**-**	**49**	**-**	**-**	**11**
Software Design Change	3	-	-	-	-	-	-	3	-	-	
Sw Manufacturing / Sw Deployment	2	-	-	-	-	-	-	3	-	-	1
Use Error	1	-	-	-	-	-	-	-	2	-	1
Vendor Change Control	2	2	-	-	-	-	-	-	-	-	-
(left blank)	-	-	-	-	-	-	-	-	-	-	50
Grand Total	**271**	17	20	7	**106**	16	15	**58**	9	23	50

4.3. Thrid Level Classification: The final classification was applied to the "Software" category from the second level classification using the "SW91 - Classification of Defects in Health Software" [12]. The results are summarized in Table 5, with the most frequent categories highlighted in bold. Specifically, the categories **Software Change Control**, **Software Design**, **Software Design Change**, and **Software Manufacturing/Deployment**, as outlined in

Table 4, were analyzed at Level 3.Out of 57 identified defects, 50 recalls were successfully classified, as shown in Table 5. However, 7 recalls, although labeled as software defects, lacked sufficient detailed information for Level 3 classification.

The most common category was “Control State”, indicating the importance of managing software state transitions effectively. “Hardware Usage” was another frequently noted category, which is expected for medical devices that rely heavily on hardware integration. The third most prevalent category was “Dead End Code”, highlighting a common issue where developers leave sections of code without proper exit paths or conditions, leading to scenarios where the software cannot proceed. Additional category ratings can be found in the Table below.

Table 5. Level 3 classification results

SW91 Mapping	Count of SW91 Mapping
2.2.2 Requirement Scope	2
3.10 Algorthm Selection	3
3.11.1 Interrupts/exceptions	1
3.11.2 Hardware Usage	**9**
3.12 Failure to Capture Design	2
4.1.2.1 Scalar Precision	1
4.1.4.4 Data Symbolic Value	1
4.10.3 Naming, Data definition, declarations	1
4.2.2.2 Reference outside declared bounds	1
4.2.9 Cleanup	1
4.3.1.2 Dead End Code	**6**
4.3.3.2 Loop Iteration Values	1
4.3.4 Control State	**14**
4.9 Missed Design Translation	2
5.1.2 Test Case Completeness	1
5.2.4 Test Result Verification	1
6.1 Release Version or Configuration	2
7.5.5 Manufacturer does not test change from third party	1
Not Available for Classification	7
Grand Total	**50**

5. Discussion

Previous classification studies [11][13] has highlighted significant gaps in the way recall data is currently categorized, particularly concerning the reasons behind these recalls. The lack of standardization and precision in recall classification hinder the root causes of defects, making it challenging for manufacturers and regulators to fully understand and address recurring issues. The literature includes various studies examining medical device recalls, the decisions leading to them, and their implications for different stakeholders [16]. For instance, Aaliya Parvin, Sudheer, and Kamaraj [16] focus on Class I medical device recalls from 2022 to 2023. While their study does not specifically address software defects, they note that software malfunctions accounted for 3.3% of recalls, underscoring the critical importance of software integrity in medical devices. Similarly Zuckerman, Brown, and Nissen [17] analyzed high-risk medical device recalls from 2005 to 2009 to evaluate the FDA approval processes used. They categorized the medical devices based on medical specialty. The results showed that 78% of recalled devices were cleared through the less rigorous audit process or were exempt from regulatory review, with cardiovascular devices being the most commonly recalled (31%).

By adopting predefined software-related classification labels, organized into two or three levels as proposed in this study, it becomes possible to bring much-needed clarity to the underlying causes of errors reported during recalls. This structured approach provides a consistent framework that allows for more accurate identification of defect patterns, which can be crucial in improving the reliability and safety of medical devices. Additionally, such a system enables manufacturers to use recall data proactively throughout the product development stages, helping to identify potential risks early on and implement corrective measures before the product reaches the market. For instance, Rajaram et al. [13] propose an approach for implementing taxonomy-based testing within a medical device software organization. Their method utilizes SW91 as the source for defect taxonomies. When combined with our three-level approach, it becomes possible to leverage FDA data to offer preventive solutions, addressing potential issues before they escalate into critical recalls.

The multi-level taxonomy developed in this study not only facilitates more precise and comprehensive recall classification but also establishes a robust foundation for future research efforts. By offering a clearer understanding of defect types and their origins, it paves the way for developing predictive models that can anticipate and prevent recurrent issues in medical device development. Enhancing the precision of classification contributes directly to improving product safety, enabling more effective quality control and risk management practices. Moreover, this refined approach can aid regulatory bodies in better monitoring and managing compliance across the industry, strengthening oversight and encouraging higher standards of safety. Ultimately, this study aims to lead to the production of more reliable and effective medical devices, benefiting both manufacturers and end-users by reducing the likelihood of failures and ensuring safer healthcare outcomes.

5.1. Suggestions for Recall Reason Classification

Improving root cause descriptions in recall databases starts with implementing standardized templates and terminology. By defining a set of standard terms and categories, medical device manufacturers can ensure consistency when documenting recall reasons. Instead of relying on free-text fields, the recall database system could use drop-down menus with predefined options, reducing the variability in language and phrasing. Additionally, integrating semantic search tools and Natural Language Processing (NLP) can help address inconsistencies by allowing users to search for recall reasons based on similar meanings rather than exact wording. For instance, a search for “software failure” could also retrieve entries listed as “software malfunction” or “coding error”, making it easier to identify and analyze related recalls across different manufacturers. NLP can further assist by automatically categorizing and suggesting root cause descriptions, streamlining the data entry process.
To enhance data analysis, machine learning can be used to generate reports that highlight common patterns in the recall data, providing manufacturers with insights on frequently observed defects. This feedback loop can help standardize how these issues are described and classified in future entries, leading to clearer and more consistent root cause documentation.

5.2. Limitations of the Study

The results from the 3rd level classification provided a clearer picture of patterns in recall causes. However, it is important to highlight that while the SW91 analysis aimed to provide a detailed classification, it has not captured all possible defect variations because of device limitation. This limitation indicates that the absence of certain defects in the dataset does not imply they do not exist, but not observed within this limited dataset. To address this concern, future research should be performed with similar classification efforts to a wider range of medical devices ensuring a more comprehensive understanding of defect patterns in the medical device domain.

Additionally, the three-level approach may face challenges in practical application. The granular classification of defects requires detailed data, which may not always be available or consistently reported by manufacturers or regulators. Lastly, the effectiveness of this approach for stakeholders depends on their ability to understand and apply the classification results. Differences in expertise and available resources among stakeholders, such as medical device manufacturers and healthcare providers, may limit its impact on improving device safety and reliability. Gathering input from various stakeholders would offer different perspectives and enhance the defect classification process.

6. Conclusions

In this study, we developed a three-level classification approach to address the existing gaps in the categorization of medical device recalls. This methodology was specifically applied to the FDA recall dataset for defibrillators, covering 271 cases. Proper classification is essential, as it provides a clearer understanding of the underlying causes of recalls, enabling the implementation of more effective preventive measures. By accurately identifying and categorizing defects, manufacturers can address root issues more systematically, leading to improved product safety and reliability.

One of the key contributions of this research is the identification of the most frequently observed recall reasons within defibrillator recalls. By highlighting these common causes, our study offers valuable insights that can help guide safety improvements in medical device design and development. Additionally, our research emphasizes the importance of adopting different levels of classification to detect various types of defects. The three-level approach allows for a more nuanced analysis, where no single level is deemed superior to the others; instead, each provides unique insights that contribute to a comprehensive understanding of defect patterns.

This study also suggests new methods for enhancing recall reason classification, including the integration of NLP and machine learning techniques in recall databases. These advanced approaches can streamline and automate the classification process, leading to greater accuracy and efficiency.

As a future direction, we plan to incorporate large language models (LLMs) into the defect classification process, utilizing them for refined defect class verification based on recall data. This step would help further improve the precision of classification, supporting more reliable safety protocols and contributing to the development of safer medical devices. Additionally; future studies could explore the economic impacts of improved recall classifications. By quantifying the potential cost savings and efficiency gains from enhanced classification processes, research could provide valuable insights into the economic benefits that more accurate and systematic recall management could offer to healthcare providers and manufacturers.

Acknowledgement

We would like to thank Assoc. Prof. Dr. Aysu Betin Can for her assistance in verifying the results of defect classifications.

References

[1] Subhan, Arif. 2022. Medical Device Recalls. Journal of Clinical Engineering, 47.2 (2022): 61-62.

[2] U.S. Food and Drug Administration, Recalls, Corrections and Removals of Devices, FDA, [Online]. Available: https://www.fda.gov/medical-devices/postmarket-requirements-devices/recalls-corrections-and-removals-devices. [Accessed: Dec. 15,-2024].

[3] Institute of Medicine. 2011. Public Health Effectiveness of the FDA 510(k) Clearance Process: Measuring Postmarket Performance and Other Select Topics: Workshop Report. Washington, DC: The National Academies Press.

[4] U.S. Food and Drug Administration, Recalls, Market Withdrawals, & Safety Alerts, Accessed Sep. 27, 2024. [Online]. Available: https://www.fda.gov/safety/recalls-market-withdrawals-safety-alerts.

[5] H. Alemzadeh, R. K. Iyer, Z. Kalbarczyk and J. Raman, 2013. Analysis of Safety-Critical Computer Failures in Medical Devices, in IEEE Security & Privacy, vol. 11, no. 4, pp. 14-26, July-Aug. 2013.

[6] European Commission. EUDAMED, [Online]. Available: https://health.ec.europa.eu/medical-devices-eudamed/overview_en. [Accessed: Oct. 22, 2024].

[7] Therapeutic Goods Administration, Recall Actions, [Online]. Available: https://www.tga.gov.au/safety/recall-actions. [Accessed: Oct. 22, 2024].

[8] Medical Device Incident Reporting (MDIR) Guide, [Online]. Available: https://www.tga.gov.au/resources/resource/user-guide/medical-device-incident-reporting-mdir-guide. [Accessed: Oct. 22, 2024].

[9] U.S. Food and Drug Administration, "Infusion Pumps Total Product Life Cycle: Guidance for Industry and FDA Staff," 2018. [Online]. Available: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/infusion-pumps-total-product-life-cycle. [Accessed: Dec. 15, 2024] .

[10] U.S. Food and Drug Administration, "Device Classification Resources," [Online]. Available: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfres/res.cfm. [Accessed: Sep. 27, 2024]

[11] Z. Bliznakov, K. Stavrianou, and N. Pallikarakis, 2014. Medical Devices Recalls Analysis Focusing on Software Failures during the Last Decade, in XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013, L. Roa Romero, Ed. IFMBE Proceedings, vol. 41. Springer, 2014.

[12] ANSI/AMMI SW91:2018, Classification of Defects in Health Software, American National Standards Institite, 2018.

[13] H. Rajaram, H. Ketheswarasarma, et al., 2019. A Framework for Taxonomy Based Testing Using Classification of Defects in Health Software—SW91, European Conference on Software Process Improvement, Cham: Springer International Publishing, 2019.

[14] T. S. Iwashita, Y. K. Kitamura, and S. W. Nagao, 2020. Influence of COVID-19 pandemic on bystander interventions, emergency medical service activities, and patient outcomes in out-of-hospital cardiac arrest in Osaka City, Japan, *Resuscitation*, vol. 157, pp. 58-64, 2020. doi: 10.1016/j.resuscitation.2020.10.017.

[15] U.S. Food and Drug Administration, Classification of Devices, Accessed Sep. 27, 2024. [Online]. Available: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfPCD/classification.cfm.

[16] MJ. Aaliya. Parvin., & Kamaraj, R. (2024). A Comprehensive Analysis of Class I Medical Device Recalls: Unveiling Patterns, Causes and Global Impacts. *Cureus*, *16*(8), e67542.

[17]Zuckerman, D. M., Brown, P., & Nissen, S. E. (2011). Medical device recalls and the FDA approval process. *Archives of internal medicine*, *171*(11), 1006-1011.