Quantitative Comparison of the 21 CFR Part 820 and ISO 13485:2016

Eung-Seok Song

The Graduate School
Yonsei University
Department of Medical Device Engineering and Management

Quantitative Comparison of the 21 CFR Part 820 and ISO 13485:2016

A Master's Thesis
Submitted to the Department of Medical Device Engineering and Management,
the Graduate School of Yonsei University
in partial fulfillment of the requirements for the degree of
Master of Medical Devices Regulations

Eung-Seok Song

December 2024

This certifies that the Master's Thesis of Eung-Seok Song is approved

Thesis Supervisor Sung-Uk Kuh

Thesis Committee Member Won-Seok Jang

Thesis Committee Member Jin-Sung Kim

The Graduate School
Yonsei University
December 2024

ACKNOWLEDGEMENTS

의료기기 규제 전문가로 성장할 수 있게 도움을 준 연세대학교 의료기기산업학과와 온코소프트, 그리고 가족과 친우에게 감사드립니다. 새로운 길에 발을 디디면서 한 걸음을 뗄 때마다 두려웠지만, 많은 분들께서 도움을 주셔서 학과 수업에 집중하고, 배운 지식을 업무에 적용할 수 있었습니다.

먼저, 논문이 완성될 수 있도록 방향성을 알려주시고, 부족한 점이 채워지도록 가르쳐 주신 구성욱 교수님, 장원석 교수님, 김진성 교수님께 감사드립니다. 논문을 완성해가며 지식 체계를 바로잡을 수 있었습니다. 그리고 항상 많은 것을 배울 수 있도록 지도해주시고, 저를 성장시켜 주신 교수님들께 감사드립니다.

항상 사랑과 믿음으로 감싸주시고, 응원해주신 우리 가족이 있어서 행복하고, 감사드립니다.

마지막으로 의료기기 규제 전문가가 될 수 있도록 도움을 주고, 응원해주신 연세대학교 의료기기산업학과, 온코소프트, 가족, 친우 그리고 고객을 위해 열정을 꺼뜨리지 않고, 인내하며 끊임없이 발전하여 긍정적 영향을 끼치는 인사가 되겠습니다. 감사합니다.

2024년 12월

송 응 석 올림

TABLE OF CONTENTS

List of Figures iii

List of Tables iv

ABSTRACT v

I. Introduction 1

1. Background 1

2. National Quality Management Systems 3

2.1. Republic of Korea 3

2.2. Japan 4

2.3. U.S.A. 5

3. NLP 7

3.1. Neural Network 10

3.2. Text representation, Vectorization 11

3.3. Transformer 13

4. Objective 16

II. Material & Method 18

1. 21 CFR Part 820 & ISO 13485:2016 18

2. Sentence-BERT 27

3. AAMI TIR 102:2019 31

4. Cosine Similarity 33

5. NDCG 34

III. Result 36

IV. Discussion 46

V. Conclusion 48

References 50

Abstract In Korean 53

List of Figures

<Figure 1. Example of documentation structure diagram> 4
<Figure 2. Natural Language Processing structure 1> 7
<Figure 3. Natural Language Processing structure 2> 8
<Figure 4. Neural Network components> 10
<Figure 5. CBOW and Skip-gram> 12
<Figure 6. The Transformer – model Architecture> 14
<Figure 7. The 7 Subsystems and interrelation of Quality System> 18
<Figure 8. 21 CFR Part 820 files> 26
<Figure 9. ISO 13485:2016 files> 26
<Figure 10. Overall pre-training and fine-tuning procedures for BERT> 27
<Figure 11. Siamese Network> 28
<Figure 12. Siamese Network at inference> 29
<Figure 13. Triplet Network> 30
<Figure 14. Comprehensive Result 36
<Figure 15. Cosine Similarity Result 37
<Figure 16. NDCG Result 37

List of Tables

<Table 1. Contents of MHLW MO169> 5
<Table 2. 21 CFR Part 820 Sections> 19
<Table 3. ISO 13485:2016 Clauses> 23
<Table 4. Comparison from 21 CFR 820 to ISO 13485:2016> 32
<Table 5. Similarity 21 CFR Part 820 Sections with ISO 13485:2016 Clauses> 38

ABSTRACT

Quantitative Comparison of the 21 CFR Part 820 and ISO 13485:2016

Previously, a qualitative comparison approach was used to identify similarities and differences between 21 CFR Part 820 and ISO 13485:2016. This approach aimed to integrate regulations and standards into a quality management system that fully encompasses all requirements. In contrast, this study quantitatively analyzes the similarities between 21 CFR Part 820 and ISO 13485:2016 and proposes ISO 13485:2016 requirements that can replace 21 CFR Part 820 requirements.

Korea and Japan established standards for the facilities and quality of medical devices by integrating additional requirements based on ISO 13485:2016. These countries also defined additional requirements to regulate organizations and individuals involved in medical device production or related activities. On the other hand, the United States regulates organizations and individuals engaged in such activities under 21 CFR Part 820.

Natural language processing (NLP) is a process and outcome of human-computer communication used for translation, search, comparison, analysis, extraction, and generation. In the past, NLP relied on word occurrence frequencies and transformed vector values. However, modern NLP techniques can now determine whether 'John' in a sentence refers to a person’s friend or the biblical figure. This advancement is attributed to the advent of Neural Networks, which improved the ability of computers to understand complex contexts, and Word2Vec, which can learn vectorized word representations. Additionally, the development of Transformer models allows for the parallel processing of all input values while maintaining their sequential order through attention mechanisms.

This study employs Sentence-BERT, which integrates the Transformer architecture to generate sentence embeddings. Cosine Similarity, Normalized Discounted Cumulative Gain (NDCG), and AAMI TIR 102:2019 were used to compare 21 CFR Part 820 and ISO 13485:2016. Cosine Similarity measures semantic similarity, with values closer to 1 indicating very high similarity and values closer to 0 indicating dissimilarity. However, it lacks predefined thresholds for interpreting similarity values. To address this limitation, this study introduces AAMI TIR 102:2019 and NDCG. AAMI TIR 102:2019 provides bidirectional mappings between 21 CFR Part 820 and ISO 13485:2016, while NDCG is primarily used to evaluate the quality of ranked search results. NDCG assesses how well the order of search results aligns with user expectations, with higher relevance scores for items appearing at the top of the list indicating better quality. In this study, relevance scores were assigned based on mappings provided in AAMI TIR 102:2019, with a score of 1 for

matches and 0 for mismatches. Finally, to overcome the lack of threshold values in NLP analysis results, ranges for cosine similarity and NDCG values were established.

Cosine similarity scores were categorized into three ranges: very high similarity (1 ~ 0.8), moderate similarity (0.8 ~ 0.6), and dissimilarity (less than 0.6). Similarly, NDCG scores were classified into high agreement (1 ~ 0.8), partial agreement (0.8 ~ 0.6), and low agreement (less than 0.6) based on the alignment of cosine similarity scores with annotations from AAMI TIR 102:2019.

Sections 820.30, 70, 72, 90, and 180 demonstrated very high similarity (greater than 0.8) with ISO 13485:2016 clauses and high agreement with AAMI TIR 102:2019 annotations. These sections can be replaced by the corresponding ISO 13485:2016 clauses when establishing each section. Sections 820.20, 22, 25, 40, 50, 60, 75, 80, 86, 100, 120, 130, 150, 160, 170, 184, 198, and 200 showed moderate similarity (0.8 ~ 0.6) with ISO 13485:2016 clauses, indicating partial agreement with AAMI TIR 102:2019 annotations. These sections appear replaceable with corresponding ISO 13485:2016 clauses but require further review. However, sections 820.5, 65, 140, 181, 186, and 250 showed dissimilarity (less than 0.6) with ISO 13485:2016 clauses and low agreement with AAMI TIR 102:2019 annotations. These sections cannot be replaced by matching ISO 13485:2016 clauses, and the requirements of these sections must be retained as is to establish a new quality management system.

As a result of quantitatively analyzing and evaluating the similarities between the sections of 21 CFR Part 820 and the clauses of ISO 13485:2016, this study concluded that 5 sections of 21 CFR Part 820 can be replaced with corresponding ISO 13485:2016 clauses to implement a quality system for manufacturers in Korea to export or manufacture medical devices in the United States. 18 sections of 21 CFR Part 820 appear replaceable with ISO 13485:2016 clauses but require appropriate review. Finally, 6 sections of 21 CFR Part 820 cannot be replaced at all with ISO 13485:2016 clauses, and the requirements of these sections must be adopted as is to establish a new quality system.

---

21 CFR Part 820, ISO 13485:2016, Natural Language Processing, Semantic Similarity, Neural Network, Word Embedding, Vector, Transformer, AAMI TIR 102:2019

I. Introduction

1. Background

The study titled "Comparison of the Quality System Requirements of Code of Federal Regulation Part 820 and International Standard ISO 13485" focuses on integrating 21 CFR Part 820 and ISO 13485:2003 to develop a harmonized quality management system for sterile medical device companies without the missing of the regulations or the international requirements. This study found similarities and differences between 21 CFR Part 820 and ISO 13485:2003, based on the historical context of the establishment of 21 CFR Part 820 (1996). The establishment of 21 CFR Part 820 (1996) involved harmonization with ISO 9001 and ISO 13485:1996 to ensure alignment of medical device regulations with internationally recognized quality management standards.

With the increasing complexity of manufacturing processes across all industries, globally recognized quality management system standards began to emerge in the late 1980s. A quality management system specifically tailored for medical device manufacturers, recognized by an international organization, was first published in 1996. Over time, as indicated by the ISO Survey 2006, approximately 8,000 ISO 13485 certifications were issued across 67 countries by the end of 2006. In contrast, the U.S. regulates medical device quality systems through 21 CFR Part 820. Medical device manufacturers operating in the U.S. are required to comply with these regulations, with the FDA conducting inspections to ensure regulatory adherence. Therefore, this study aims to develop an integrated quality management system that combines ISO 13485:2003 and 21 CFR Part 820, enabling manufacturers to distribute medical devices both in the U.S. and internationally[1)].

Prior to developing a harmonized quality management system, this study analyzes the article by Gallifa, J., "The new ISO 13485:2003. Detailed comparison with FDA Quality System Regulations and ISO 9001:2000," published in 2005 (pp. 1-48). The article presents a table aligning the individual section of 21 CFR Part 820 with the clauses of ISO 13485:2003 and describing their differences. Based on this analysis, the study facilitates an integrated quality management system specifically designed for sterile medical device companies, which provides a table that groups similar sections of 21 CFR Part 820 and

clauses of ISO 13485:2003 into categories. The table would help U.S. manufacturers integrate ISO 13485:2003 requirements into their compliance with the Quality System Regulation (QSR), ensuring no omissions in regulatory or international requirements[1].

In contrast, this study does not aim to address the gap between 21 CFR Part 820 and ISO 13485:2016 but instead focuses on conducting a quantitative analysis to identify how ISO 13485:2016 clauses can potentially substitute for specific 21 CFR Part 820 sections. By analyzing the similarities between individual 21 CFR Part 820 sections and all ISO 13485:2016 clauses, this study seeks to provide manufacturers with practical insights into how ISO 13485:2016 can be utilized to streamline their efforts in meeting U.S. regulatory requirements. The primary goal is not regulatory harmonization but to reduce the additional workload for manufacturers who have already implemented ISO 13485:2016 based quality management systems.

While this study enhances the understanding of similarities between the regulation and the standard, its primary focus is on providing manufacturers with actionable insights to facilitate informed decision making. By contributing an objective, data-driven perspective on the substitutability of specific clauses, this study seeks to support manufacturers in streamlining their quality management processes, particularly for those who have already implemented ISO 13485:2016. This approach not only advances the understanding of alignment but also reduces the additional workload for meeting U.S. regulatory requirements.

2. National Quality Management Systems

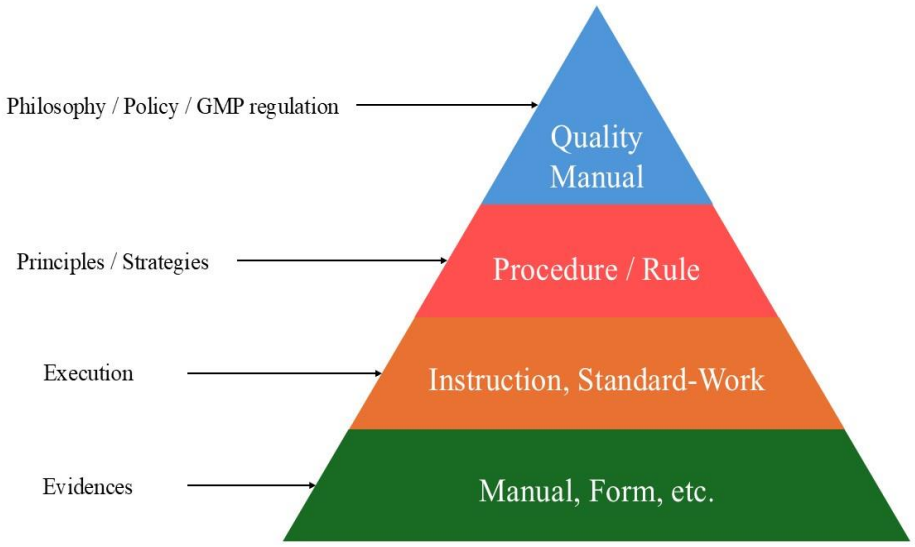
2.1. Republic of Korea

In the Republic of Korea, a person who intends to manufacture a medical device must obtain a manufacturing business permission[2]. Furthermore, those who obtain a manufacturing business permission must obtain a manufacturing permission, manufacturing certification, or manufacturing notification for the medical device to be manufactured[2]. In addition, those intending to obtain a manufacturing business permission and those intending to obtain a manufacturing permission, manufacturing certification, or manufacturing notification must comply with the Standards for Facilities and Manufacturing and Quality Management System specified in Annexes 2 of the Enforcement Rule of the Medical Devices Act[3]. The method of auditing the Facilities and Manufacturing and Quality Management System is set out in Article 5 - Standard for GMP Audit of Good Manufacturing Practice (GMP) for Medical Devices and Annexes 2 of the Enforcement Rule of the Medical Devices Act[3].

In December 2017, the Ministry of Food and Drug Safety (MFDS) published the Medical Device GMP International Quality Management Guide in response to the transition from ISO 13485:2003 to ISO 13485:2016. This guide was designed to support domestic medical device companies in securing international competitiveness and facilitating exports. It compared the contents of Good Manufacturing Practice (GMP) for Medical Devices Annexes 2 - The Standard for GMP Audit for Medical Devices with the revised international standard ISO13485:2016 (updated from ISO 13485:2003). It described changes from the previous The Standard for GMP Audit for Medical Devices and provided additional considerations for manufacturing sites[4]. Finally, A Comprehensive Guide to Medical Device GMP (Revision 8), which provides a detailed explanation of Good Manufacturing Practice (GMP) for Medical Devices, states that Annexes 2 of the Enforcement Rule of the Medical Devices Act, which is described in Article 5 - Standard for GMP Audit of Good Manufacturing Practice (GMP) for Medical Devices, reflects ISO 13485:2016. This means that those who intend to manufacture medical devices in Korea must establish a quality management system that complies with ISO 13485:2016 and regulatory requirements[5]. Figure 1. Example of documentation structure diagram illustrates the classification of documents used in the quality management system into types based on their scope of application, decision-making stage, and level of importance, and

provides a simplified schematic representation of the structure of each document, as described[5].

<Figure 1. Example of documentation structure diagram[5]>



2.2. Japan

In the Japan, a person who intends to manufacture or is participate in one or more stages of the life cycle of a medical device, including design and development, production, storage, and distribution, or engages in the business of manufacturing and marketing (Marketing Authorization Holder (MAH)) is required to establish a quality management system[6), 7)]. This requirement is stipulated in the Ministerial Ordinance on Standards for Manufacturing Control and Quality Control for Medical Devices and In-Vitro Diagnostics (MHLW MO169), which was initially established in 2004 to harmonize the quality management system requirements with ISO13485:2003. Then, in 2014, The second chapter of MHLW MO169 was revised to align more closely with ISO 13485:2003. And the additional requirements to ISO13485:2003 were moved to the third chapter of the ordinance (MHLW MO169). Later, in 2021, The second chapter of MHLW MO169 was further revised to align with ISO13485:2016. Furthermore, the Pharmaceuticals and

Medical Devices Agency (PMDA) provides a table (Table 1. Contents of MHLW MO169) which shows harmonization Japanese medical devices quality management system requirements with ISO 13485:2016[8].

<Table 1. Contents of MHLW MO169[8]>

Chapters: Title	Contents
Chapter 1: General	Scope, Definition
Chapter 2: Basic requirements	Requirements harmonized with ISO 13485
Chapter 3: Additional Requirements	Additional requirements
Chapter 4: Requirements for Biological Medical Devices etc.	Product specific requirements
Chapter 5: Requirements for Radioactive IVDs	
Chapter 5-2: Requirements for R-SUD	
Chapter 6: Provisions Applied Mutatis Mutandis for manufacturers etc.	Provisions applied mutatis mutandis

2.3. U.S.A.

In the U.S.A., any person who intends to design, manufacture, package, label, store, install, or service all finished devices – which are any device or accessory to any device that is proper for use or able to function, whether or not is packaged, labeled, or sterilized intended for human use – and imports them domestically or exports them from other countries to U.S.A., is governed by 21 CFR Part 820[9].

The 21 CFR Part 820 is referred to as Current good manufacturing practice (CGMP) and quality system regulation. Historically, the CGMP (Current Good Manufacturing Practice) requirements for devices in 21 CFR Part 820 were first authorized under Section 520(f) of the Federal Food, Drug, and Cosmetic Act (the Act). Over time, 21 CFR Part 820 has steadily evolved to become the current regulatory standard in U.S.A. The requirements in this regulation are intended to guarantee that finished devices will be safe and effective, and compliant with the Federal Food, Drug, and Cosmetic Act[9].

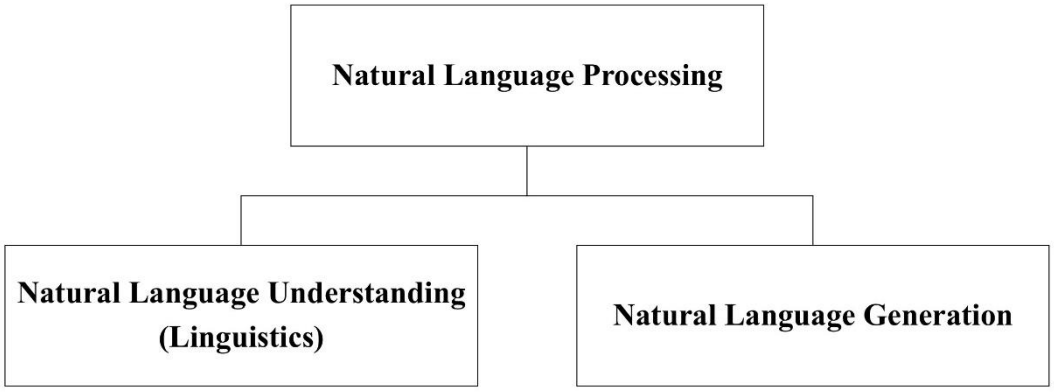
The CGMP requirements mandate that manufacturers establish and maintain a quality system that is appropriate for the specifications of all finished devices which are produced in compliance with regulatory requirements and other applicable standards. However, if some organizations, as described above, do not engage in some parts mentioned, then they only need to comply with requirements related to the activities they perform. The executive layer is responsible for ensuring that the quality system effectively meets these requirements and that all finished devices are consistently manufactured and distributed in compliance with specifications, ensuring their safety and effectiveness[9].

The Quality System Regulation (QSR), a codified regulation, adopts an "umbrella" approach and provides a general framework. Because there are a lot of manufacturers where produce many types of medical devices, the QSR applies broadly to them producing medical devices but does not specify in detail how they should establish a quality system suited to their state-of-the-art devices. All manufacturers are required to follow the framework in a manner that is applicable to their organizational procedures, specific products, and operations. An appropriate quality system ensures that medical devices are consistently safe and effective[10].

3. NLP

Natural Language Processing (NLP) is a service for users who do not have enough time to learn or master a new language. NLP is a branch of artificial intelligence and linguistics that is dedicated to enabling computers to understand sentences or words written in human language. It was created to facilitate users' tasks and satisfy the desire to communicate with computers in natural language, and it is divided into two parts (Figure 2. Natural Language Processing structure 1) and developed: natural language understanding and natural language generation[11].

<Figure 2. Natural Language Processing structure 1[11]>

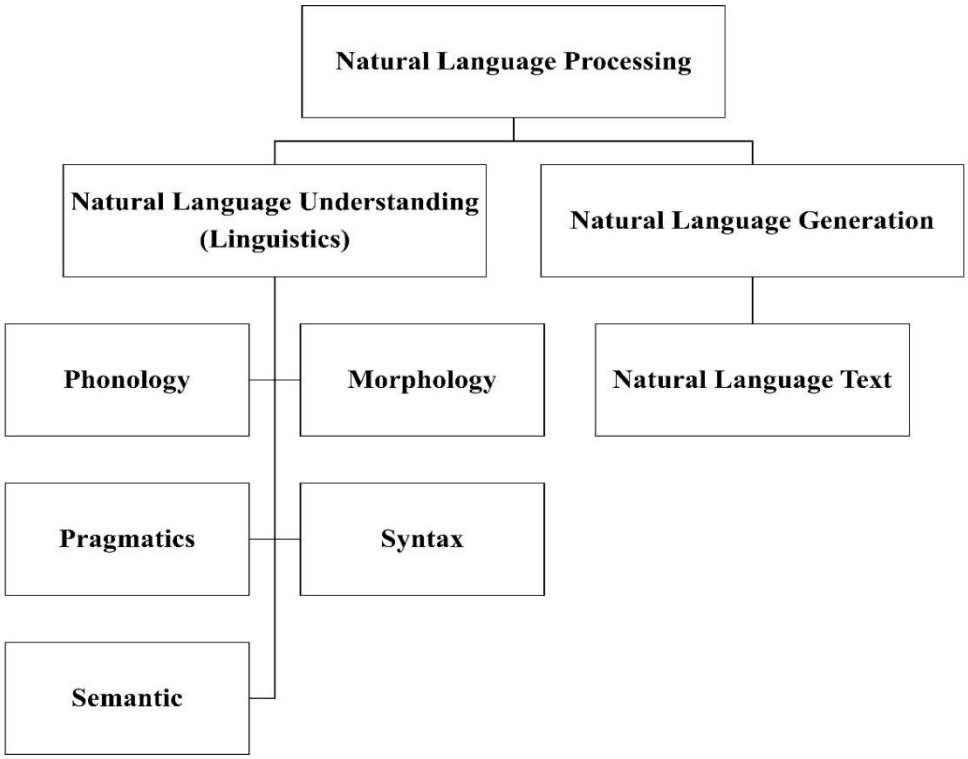


In the existing literature, most of the research on NLP has been conducted by computer scientists, and experts from various academic fields such as linguists, psychologists, and philosophers have also shown interest. One of the most interesting aspects of NLP is that NLP is not only useful as a technical tool, but also has academic and intellectual value in deepening the understanding of human language. The field of NLP is related to various theories and techniques that deal with natural language problems that communicate with computers. It is used for automatic text summarization and generation, text classification and evaluation, machine translation, question answering, contextual relationship analysis,

information extraction and recognition, etc. These tasks are also directly applied to real-life situations such as language translation, customer feedback analysis, chatbots, and scanned bill extraction[11]. One notable example is that natural language processing technologies have been studied in the legal field to improve the understanding of real-world situations for legal professionals such as lawyers, and to lower the barrier to entry into the legal field for those who need legal services or practitioners in the industry[12), 13)]. These research activities show that NLP tasks are closely connected while being used independently.

Figure 3. Natural Language Processing structure 2 expands on the part about natural language understanding among the two categories of natural language processing. This natural language understanding allows machines to understand natural language and extract and analyze concepts, entities, emotions, keywords, etc. For example, it is used in the customer management field of a company or organization to understand problems reported by customers verbally or in writing. In the context of Figure 3, this study will focus specifically on explaining the terminology and concepts related to semantic[11].

<Figure 3. Natural Language Processing structure 2[11)]>



At the semantic level, the most important task is to figure out the proper meaning of a sentence. In the case of humans, we rely on our knowledge of language and the concepts present in that sentence to understand the meaning of a sentence, but machines solve the problem of understanding the meaning of a sentence through semantic processing. Semantic processing figures out the possible meanings of a sentence by processing the logical structure of the sentence and recognizing the most relevant words to understand the interactions between words or other concepts in the sentence[11]. It differs from textual entailment (TE) which is characterized by unidirectional equivalence. Semantic processing is not affected by differences in vocabulary or syntax used to convey the same meaning and has the characteristics of bidirectional and progressive equivalence[14]. For example, even if a sentence does not consist of actual words, we understand that it is about “animals” if it contains related concepts such as “tiger”, “lion”, “fox”, or “penguin”. In the same context as “the moon and the sun are more similar than clouds and the sun”. This level of processing also resolves semantic ambiguity in words that have multiple meanings. For example, the word 'organ' as a noun can mean either a human organ or a musical instrument. At the semantic level, words are examined for their dictionary interpretation or interpretation derived from the context of the sentence. For example, there is a sentence, “John is good and diligent.” This sentence is talking about John, a biblical character, or a person named John. Therefore, the appropriate interpretation is considered by looking between words and words, between phrases, to figure out the appropriate meaning of the sentence. Therefore, the semantic similarity measure is a methodology that evaluates the degree of the complex semantic relationship between them based on the meaning or semantics of sentences or documents rather than lexical similarity[11), 14)].

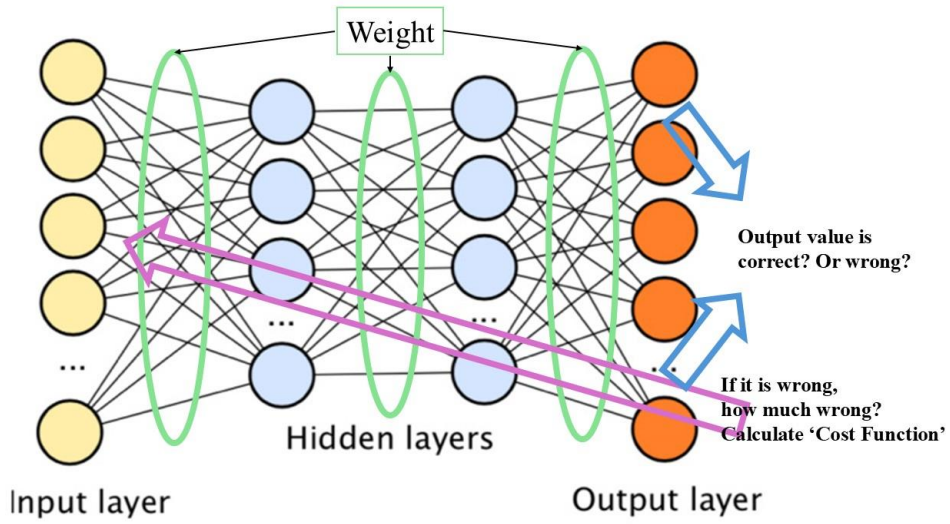
Natural language processing (NLP) initially began with simple rule-based approaches and has made various advancements over time. With improvements of computing performance and data processing capabilities, Neural Network models began to be successfully applied to NLP. Subsequently, embedding techniques such as Word2Vec and GloVe made notable progress in converting text into vector representations, which opened new possibilities for NLP. Then, the Transformer model emerged and demonstrated performance that surpassed previous techniques, and innovative models such as BERT and GPT were developed based on it[15), 16)]. With these advancements, NLP is once again receiving great attention. Following Section 1 NLP, this paper explores the development process of NLP in the order of Neural Network, Word2Vec, and Transformer.

3.1. Neural Network

Neural Network is a model inspired by the human brain, where neurons are interconnected. Rather than being pre-programmed to perform a specific task, Neural Network learns by adjusting the connections (weights and biases) between neurons based on the exposed data to produce outputs close to the correct answer (or labeled value). This process is done by repeatedly updating the connection values between neurons by backpropagating information from the output layer to the input layer based on the loss function to minimize the difference between the final output data and the correct answer[17), 18)].

Typically, as illustrated in Figure 4. Neural Network architecture and components, a Neural Network consists of an input layer, one or more hidden layers, and an output layer, and the neurons in each layer are connected through weights. Increasing the number of hidden layers enables the modeling of complex nonlinear, allowing the network to perform more sophisticated feature extraction. This approach is often referred to as deep learning, where neural networks with multiple hidden layers are used to uncover patterns in complex datasets. Additionally, researchers have developed specialized architectures such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to solve a variety of complex problems[17), 18)].

<Figure 4. Neural Network architecture and components>



Weight(w) plays a role in determining the importance of information in the connections between nodes. The greater the weight of a specific connection, the greater the influence that connection has on the output value. And it is expressed green ovals in Figure 4. Neural Network architecture and components[17), 18)].

Bias(b) helps the model express the data distribution more flexibly by adjusting the basis of the activation function by adding a constant value to the value calculated by the weight[17), 18)].

Cost function($J(\theta)$) measures the difference between the predicted output and the correct answer, guiding the model during training[17), 18)].

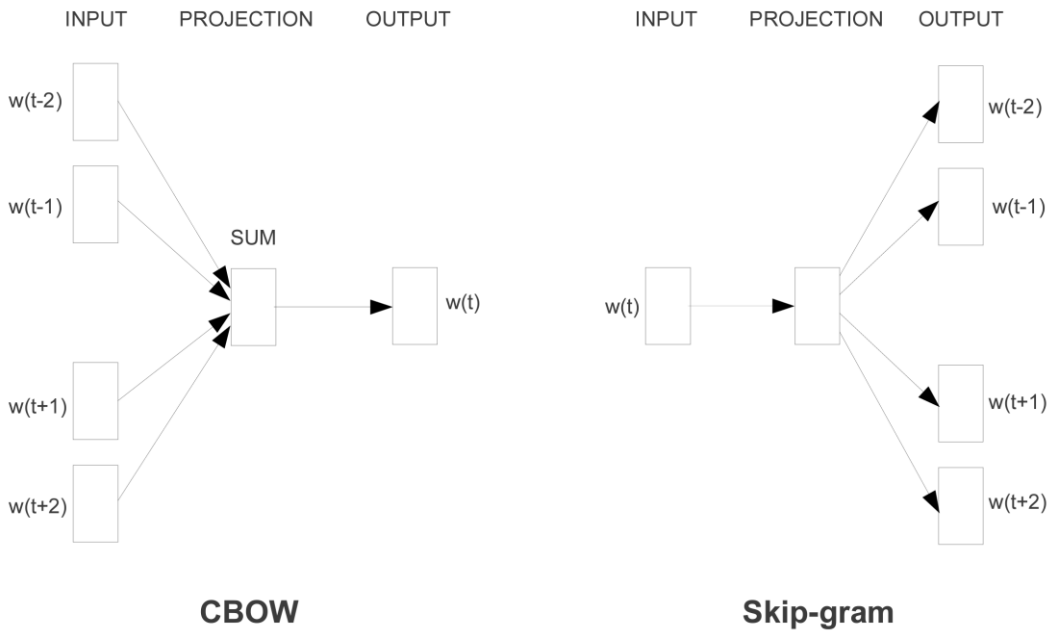
Backpropagation explains the process of propagating the error, calculated by the Cost Function, backward through the network from the output layer to the input layer. This process adjusts the weights and biases iteratively, aiming to minimize the error and improve the model's accuracy over time. And it is expressed as a pink arrow in Figure 4. Neural Network architecture and components[17), 18)].

3.2. Text Representation, Vectorization

Previous word representation methods used in natural language processing focused either on the number of words (Bag of Word) or on the frequency and rarity of words (TD-IDF), so they only processed individual words in isolation, and did not capture the similarity between words[19)]. While N-gram, a statistical language modeling technique, has the advantage of being able to represent language with relatively low computational complexity for a large amount of data, but when the data set is not large enough, the performance depends on the quality of the data set[20)].

Therefore, to overcome these limitations, Skip-gram (Figure 5. CBOW and Skip-gram), an efficient method that can learn high-quality word vectors from a large amount of unstructured text data, was introduced[21)]. For reference, CBOW (Figure 5. CBOW and Skip-gram), which has opposite input and output directions from Skip-gram, predicts the target word based on its surrounding words. The learning goal of Skip-gram is to find useful representations that can predict the surroundings of the target word within a sentence[20)].

<Figure 5. CBOW and Skip-gram[20)]>



In both CBOW and Skip-gram models, w(t) represents the target word within a sentence, while w(t-1, t-2, t+1, t+2) denote the words surrounding the target word – the preceding one to two words and the following one to two words, respectively. These words serve as the input or output of the model depending on the method. Both models utilize a hidden layer, referred to as the projection layer, which works similarly to the principles of neural networks to learn weights and find the correct representation during the learning process (Figure 5. CBOW and Skip-gram)[19)].

Unlike most neural networks previously used to learn word vectors, Skip-gram can efficiently learn from a large amount of data using a simple neural network architecture. The learned vectors explicitly encode many linguistic regularities and patterns. Although the word forms are different, the semantic meaning is captured in numerical form, and the nuance and contextual information of the word are preserved as much as possible, so that the model can effectively understand and utilize these features[20)].

Rather than relying on hand-crafted features, researchers began using large corpora that represented the encapsulated semantic relationships between distributed representations in vector spaces. This approach later led to the development of advanced word embedding models, such as ELMo[16)].

3.3. Transformer

This architecture overcomes the memory limitations caused by computation in existing models by using the Attention mechanism. For example, the number of operations required to connect signals between arbitrary input or output positions in the ConvS2S and ByteNet models increases linearly and logarithmically, respectively, with the distance between positions. In contrast, the Transformer leverages the Attention mechanism to model global dependencies regardless of the distance of the input or output sequence[22]. On the other hand, RNN and Long Short-Term Memory (LSTM) models process inputs sequentially, so they cannot effectively handle words with different morphological forms when the input time and order are far apart. In contrast, the Transformer enables parallel processing, which significantly improves the learning speed, and can effectively handle semantic similarity even in long sequences[23].

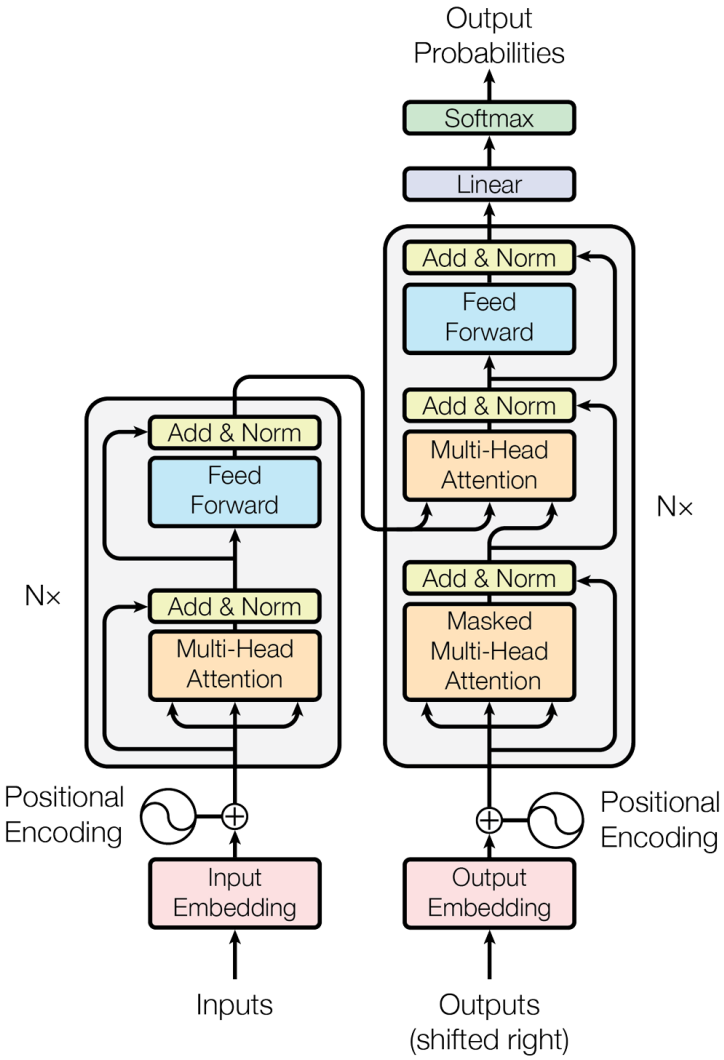
The Transformer model has an encoder and decoder with multiple layers of attention, point-wise feedforward networks. The encoder converts the input word into a fixed-length continuous representation (vector). This continuous representation, known as word embedding, encodes the order and semantic information of the input by representing a word as a point in an n-dimensional vector space[23, 24]. It learns the relationship between each word in the input sequence using multiple layers of attention and combines the output of each layer with the input using residual connections. The vector representation at each position is nonlinearly transformed in a high-dimensional space through a fully connected feedforward network, and the vector generated in this way is passed to the decoder[22, 24].

The decoder has a similar structure to the encoder but has an additional sublayer that performs multi-head attention over the output of the encoder. The decoder takes as input both the sequences fed from the encoder and the sequences generated within the decoder itself. The decoder uses auto-regressive processing to apply masked self-attention so that each position ensures to attend only to the preceding positions in the sequence. At the same time, it performs parallel attention processing over the output of the encoder to integrate contextual information from the input sequence[22, 24].

Finally, the Attention mechanism learns the relationship between the output and the vectors called query (the criteria to search for), key (the feature of the data), and value (the value to be returned in the end), and maps them. The dot product between the query vector and the key vector is computed and scaled by $\sqrt{d_k}$ to prevent the values from becoming excessively large. Then, the softmax function is applied to obtain the weights for

calculating the weighted sum of the value vector. The weights are then multiplied by the value vector to produce the final output. In this way, the model pays more attention to important words by giving higher weights to relevant words than to irrelevant ones. The resulting vector is fed to a fully connected linear layer before being passed to the second sublayer of the encoder for further processing. This process can be performed in parallel due to the multiple layer structure[22), 24)].

<Figure 6. The Transformer – model Architecture[22)]>



Finally, as shown in Figure 6. The Transformer – model Architecture, it consists of an encoder-decoder architecture with multiple layers of attention and feedforward networks. The Nx labels in figure 6. The Transformer – model Architecture, indicate the repeated layers in both the encoder and decoder, which enable the model to learn increasingly complex patterns and relationships in the input sequence. Positional Encoding ensures that sequence order information is preserved by adding position-based values to the input embeddings, enabling the Transformer to process inputs in parallel[22].

The encoder converts input embeddings into continuous vector representations, capturing both semantic and positional information using self-attention and feedforward sublayers. The decoder receives these representations along with its own generated outputs, using masked attention to ensure it predicts words sequentially while attending to relevant contextual information. The Attention mechanism calculates the importance of each word in the input by computing dot products of query, key, and value vectors, followed by scaling and applying the softmax function. By focusing more on relevant words, the Transformer achieves efficient and accurate sequence-to-sequence learning[22].

4. Objectives

The study titled "Comparison of the Quality System Requirements of Code of Federal Regulation Part 820 and International Standard ISO 13485" aims to develop a harmonized quality management system by integrating 21 CFR Part 820 and ISO 13485:2003 for sterile medical device companies. This integrated quality management system ensures full compliance with both regulatory and international requirements. To achieve this, the study provides a table that groups similar sections of 21 CFR Part 820 and clauses of ISO 13485:2003 into categories, forming a robust system that prevents any omissions of the regulations or international requirements.

In contrast, this study quantitatively analyzes the similarities between 21 CFR Part 820 and ISO 13485:2016 to provide analytical insights for manufacturers. By offering an objective assessment of the alignment between the regulation and the standard, this study identifies opportunities where ISO 13485:2016 can be utilized to reduce the effort required for manufacturers to meet 21 CFR Part 820 requirements. Unlike prior studies focusing on integration or harmonization, this study emphasizes minimizing the additional workload for compliance.

To address this objective, this study first investigates the quality management system (QMS) frameworks implemented by nations adhering to ISO 13485:2016 and 21 CFR Part 820. Specifically, it examines the regulatory landscapes of the Republic of Korea and Japan, which incorporate additional requirements into their QMS based on ISO 13485:2016. These systems require companies that manufacture medical devices or are participating in one or more stages of the medical device life cycle consistently produce effective and safe medical devices that meet regulatory and customer requirements. In contrast, the United States, which implements its own regulations under 21 CFR Part 820, requires companies that manufacture finished medical devices and components or are involved in one or more stages of their life cycle consistently produce effective and safe medical devices and components in compliance with its regulations. Despite the different regulatory foundations, the goal remains the same to ensure the consistent production of safe and effective medical devices.

Additionally, this study explores natural language processing (NLP) methods to enable a semantic comparison between ISO 13485:2016 clauses and 21 CFR Part 820 sections. Advancements in NLP, such as neural networks, word embedding techniques, and Transformer models, have significantly enhanced the ability to analyze human language

and compare texts. Leveraging these advancements, this study employs the Sentence-BERT model as a quantitative NLP-based approach to identify specific ISO 13485:2016 clauses that can potentially substitute for 21 CFR Part 820 sections. Moreover, the study utilizes the cosine similarity calculation method applied in Sentence-BERT. The rationale for using cosine similarity lies in its defined range of similarity values, which makes interpretation straightforward, and its focus on directional similarity rather than vector magnitude, enabling an effective comparison of semantically similar texts. Since the Sentence-BERT paper also uses cosine similarity to evaluate semantic similarity between sentences, this methodology aligns well with the analytical objectives of this study. By doing so, the study supports manufacturers with quality management systems based on ISO 13485:2016 in facilitating the consistent production and distribution of safe and effective medical devices within the U.S. regulatory environment.

To enhance the robustness of this similarity analysis, the study incorporates the Normalized Discounted Cumulative Gain (NDCG) metric, a widely used evaluation method for ranking systems, and the bidirectional mapping methodology outlined in AAMI TIR 102:2019 by experts from the AAMI organization. By combining cosine similarity metrics with expert insights from AAMI, this approach ensures a comprehensive evaluation of the alignment between 21 CFR Part 820 and ISO 13485:2016. The study also establishes score ranges for similarity and NDCG to complement the NLP based evaluations, addressing the limitations of similarity analysis models that lack defined thresholds.

Therefore, this study aims to provide practical metrics and insights for companies seeking to export medical devices to the United States or manufacture them within the United States. By minimizing changes to existing quality systems while meeting U.S. regulatory requirements, the study supports manufacturers in streamlining compliance processes and reducing additional burdens.

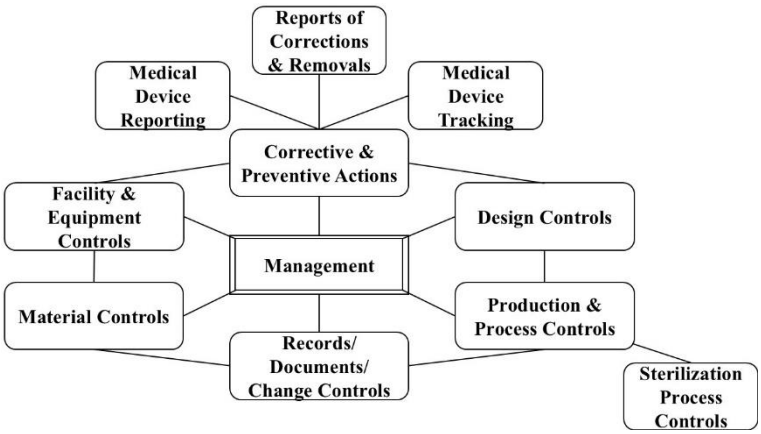
II. Material & Method

1. 21 CFR Part 820 & ISO 13485:2016

This subchapter introduces 21 CFR Part 820 and ISO 13485:2016 and presents two tables of the sections of 21 CFR Part 820 and the clauses of ISO 13485:2016.

Before reviewing 21 CFR Part 820, this study visualizes the relationships between its components using Figure 7. The 7 Subsystems and interrelation of Quality System.

<Figure 7. The 7 Subsystems and interrelation of Quality System[25)]>



The 21 CFR Part 820 is referred to as Current good manufacturing practice (CGMP) and quality system regulation. It applies to any person who intends to design, manufacture, package, label, store, install, or service all finished devices – which are any device or accessory to any device that is proper for use or able to function, whether or not is packaged, labeled, or sterilized intended for human use – and imports them domestically or exports them from other countries to U.S.A. It also mandates manufacturers establish and maintain a quality system that is appropriate for the specifications of all finished devices which are produced in compliance with regulatory requirements and other applicable standards. Next, this study will examine the sections and descriptions of 21 CFR Part 820 through Table 2. 21 CFR Part 820 Sections.

<Table 2. 21 CFR Part 820 Sections[9)]>

Section & Name	Description
Subpart A – General Provisions	
820.1 Scope	This regulation explains the scope of its application in terms of geographic regions, product categories, manufacturing processes, and the circumstances under which exemptions or exclusions from its application may be granted.
820.3 Definition	Definitions of words used in 21 CFR Part 820.
820.5 Quality system	The manufacturers shall establish and maintain a quality system.
Subpart B – Quality System Requirements	
820.20 Management responsibility	Build organizational structure, policy, system procedures, and review system.
820.22 Quality audit	Assure that the quality system complies with requirements.
820.25 Personnel	Manufacturers shall have sufficient employees with necessary educations.
Subpart C – Design Controls	
820.30 Design controls	Design and Development planning, realization(execution), testing, design transfer, documentation.
Subpart D – Document Controls	
820.40 Document controls	Document procedures (write, review, approve)
Subpart E – Purchasing Controls	
820.50 Purchasing controls	Assure the purchased materials and products remain as required by manufacturers.

Subpart F – Identification and Traceability	
820.60 Identification	Identifying product during all stages of receipt, production, distribution, and installation to prevent mix-ups.
820.65 Traceability	Implantable devices and devise support or sustain life shall be tracked.
Subpart G – Production and Process Controls	
820.70 Production and process controls	Control complex manufacturing processes, sterilization processes, and cleaning processes.
820.72 Inspection, measuring, test equipment	Testing production processes
820.75 Process Validation	Validate whether the results from production process are met as intended.
Subpart H – Acceptance Activities	
820.80 Receiving, in-process, and finished device acceptance	Set tolerances for importing materials and products, intermediate products in the manufacturing process, and finished goods.
820.86 Acceptance status	Acceptance criteria and results
Subpart I – Nonconforming Product	
820.90 Nonconforming product	Management methods for nonconformance product
Subpart J – Corrective and Preventive Action	
820.100 Corrective and preventive action	Fix current status and prevent what will happened.
Subpart K – Labeling and Packaging Control	
820.120 Device labeling	Record UDI, lot, batch in documents.
820.130 Device packaging	Packaging and shipping procedures.
Subpart L – Handling, Storage, Distribution, and Installation	

820.140 Handling	Ensure that mix-ups, damage, deterioration, contamination, or other adverse effects to product do not occur during handling.
820.150 Storage	Ensure that mix-ups, damage, deterioration, contamination, or other adverse effects to product do not occur during storage.
820.160 Distribution	The approved devices to be distributed. Meet the requirements by customers before delivery.
820.170 Installation	Installation procedures
Subpart M – Records	
820.180 General requirements	Private information shall be secured. Record retention period. Not apply to maintain record.
820.181 Device master record	Shall be approved in accordance section 820.40. Device specifications, production methods and others shall be recorded.
820.184 Device history record	Manufactured date, quantity, acceptance records, UDI, and others.
820.186 Quality system record	Quality System records in accordance with 820.20 and 820.40.
820.198 Complaint files	Complaints shall be processed in accordance with adequate procedures.
Subpart N – Servicing	
820.200 Servicing	Service shall be processed in accordance with adequate procedures.
Subpart M – Statistical Techniques	
820.250 Statistical techniques	Valid statistical techniques required for establishing, controlling, and verifying the acceptability of process capability and product characteristics.

The ISO 13845:2016 standard requires a quality management system for manufacturers involved in any stage of the process of the life cycle of medical devices including design and development, production, storage, distribution, installation, servicing, decommissioning and disposal of medical devices. Manufacturers shall establish a quality management system that consistently produces safe medical devices that perform as intended. Additionally, this standard expects manufacturers to identify their role under their regulations, as well as the specific requirements of their activities and medical devices. The standard specifies requirements for process approach and risk-based approach[26].

The process approach is that the activities which take inputs and transform them into outputs can be considered a process. For organizations to perform effectively, they need to identify and manage a large number of connected processes. The application of a system of processes within organizations, along with the identifications and interactions of these processes and their management to turn out desired results, is referred to as a “process approach”[26].

The risk-based approach is to identify potential hazard and risk in quality management system. Then develop controls to prevent and mitigate risks. The risks in quality management system may transfer to patients by finished products[26].

Furthermore, the standard allows organizations to declare certain requirements of ISO 13485:2016 as not applicable if they do not perform activities corresponding to those requirements. Next, this study will examine the clauses and descriptions of ISO 13485:2016 through Table 3. ISO 13485:2016 Clauses[26].

The sections and clauses listed in Table 2. 21 CFR Part 820 Sections and Table 3. ISO 13485:2016 Clauses are used as input data for this study. However, 21 CFR Part 820.1 Scope, 820.3 Definition, and ISO 13485:2016 1 Scope, 2 Normative References, and 3 Terms and Definitions are excluded. The input data from 21 CFR Part 820 include subpart titles, section titles, and their content, while the input data from ISO 13485:2016 include clause titles (X.X level), subclause titles, and their content. These input data are manually typed and saved into a single Word file. The sections from 21 CFR Part 820 are saved under titles such as '820.5, 820.20, …, 820.250' in the 21 CFR Part 820 files folder (Figure 8. 21 CFR Part 820 files), and the clauses from ISO 13485:2016 are saved under titles such as '4.1.1, 4.1.2, …, 8.5.3' in the ISO 13485:2016 files folder (Figure 9. ISO 13485:2016 files).

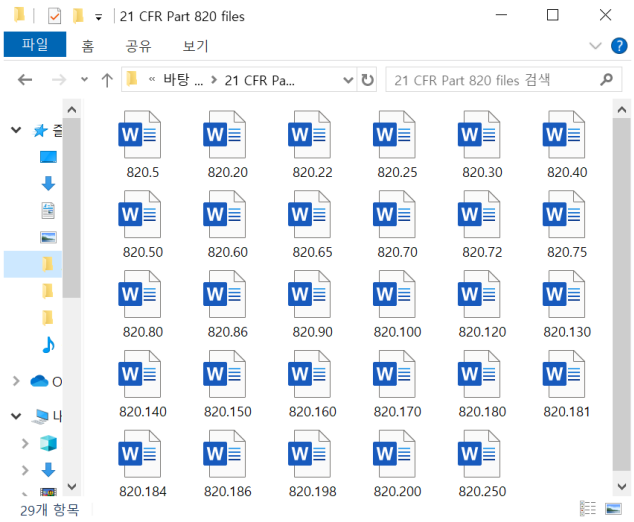
<Table 3. ISO 13485:2016 Clauses[26)]>

Clause & Name	Description
1 Scope	Organizations engage in different stages of the life cycle of medical products, including the design, repair, installation, maintenance and storage of medical devices. Clauses 6, 7 and 8 may not be applicable to
2 Normative references	Clarifies that any references to ISO 9000 refer to ISO 9000:2015.
3 Terms and definitions	Defines terms used throughout this standard.
Chapter 4 Quality Management System	
4.1 Management responsibility	Establish quality management system. Regulatory requirements shall be applicable to organizations. Risk-based approach to quality management system. Monitoring outsourced processes.
4.2 Documentation requirements	Develop Quality manual, quality policy, quality objectives. Develop medical device file (specifications, labelling, instructions, installation, services, and others). Documented procedures and record
Chapter 5 Management responsibility	
5.1 Management commitment	Internal communication, quality policy, guaranteeing quality objectives, performing management review, guaranteeing resources.
5.2 Custom focus	Ensure to meet regulatory requirements and customer requirements.
5.3 Quality policy	Adequate purpose of organization.

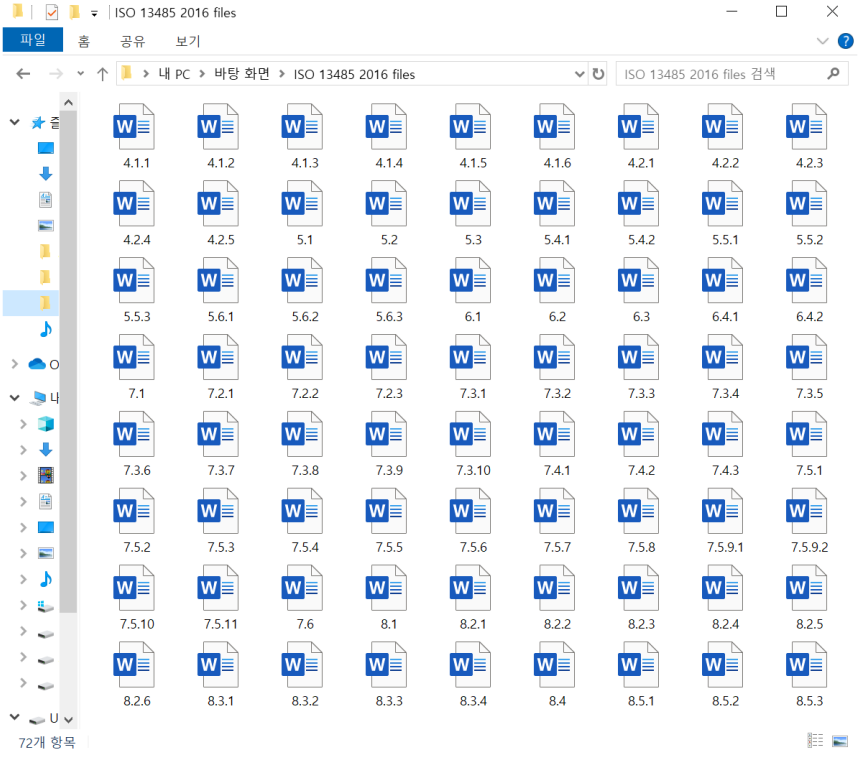
	Demonstrate effectiveness of quality management system. The policy shall be reviewed.
5.4 Planning	Planning to achieve quality objectives.
5.5 Responsibility, authority and communication	Mandate role and responsibility. Designate quality management representative. Internal communication.
5.6 Management review	Review per planned period. Management review input & output.
Chapter 6 Resource management	
6.1 Provision of resources	Ensure resources to implement quality management system, maintain effectiveness, and meet requirements.
6.2 Human resources	Education, Training based on task risk of personnel.
6.3 Infrastructure	Infrastructure for requirements of products.
6.4 Work environment and contamination control	Work environment, contamination and sterilization control.
Chapter 7 Product realization	
7.1 Planning of product realization	Planning product realization (what processes we need to realize products).
7.2 Customer-related processes	Ensure to meet customer requirements. Training users. Communicate with customers.
7.3 Design and development	Design and development procedure. Design and Development planning, realization(execution), testing, design transfer, documentation.
7.4 Purchasing	Assure the purchased materials and products remain as required by manufacturers.

7.5 Production and service provision	Control complex manufacturing processes, installation processes, cleaning processes, service processes, and sterilization processes. Identify and Traceability. Preservation.
7.6 Control of monitoring and measuring equipment	The equipment shall be adequate in process.
Chapter 8 Measurement, analysis and improvement	
8.1 General	Planning monitoring, measuring, analysis, and improvement process to ensure adequacy of product and quality management system.
8.2 Monitoring and measurement	Monitoring customer feedback. Process complaint. Report adverse event to regulatory authority. Internal audit. Monitoring and measuring processes and products.
8.3 Control of nonconforming product	Identify nonconformance products to avoid mix-ups. Take actions to nonconformance products before delivery and after delivery. Remake.
8.4 Analysis of data	Analyze collected data from inside and outside.
8.5 Improvement	Utilize quality policy, quality objectives, audit results, postmarket surveillance, data analysis to improve quality management system. Corrective action. Preventive action.

<Figure 8. 21 CFR Part 820 files>



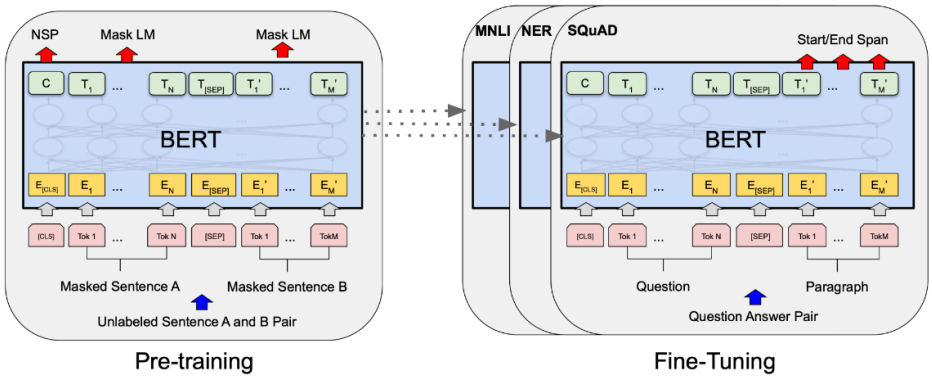
<Figure 9. ISO 13485:2016 files>



2. Sentence-BERT

The BERT model is a pre-trained transformer model that can perform various NLP tasks (question answering, sentence classification, sentence-pair regression). The BERT model uses a cross-encoder where sentence pairs are provided as input and distinguished by [CLS], [SEP] tokens. The transformer network utilized by BERT enables multi-head attention. However, this input requires a huge amount of computation to perform sentence pair regression tasks, making it unsuitable for sentence similarity analysis and clustering, as described in the Sentence-BERT paper. In addition, since the BERT model is structured to receive sentence pairs as input, it is difficult to generate independent sentence embeddings. Many researchers are conducting various studies to derive single sentence embeddings, but it remains a challenge due to the structural limitations of the BERT model[27].

<Figure 10. Overall pre-training and fine-tuning procedures for BERT[36]>

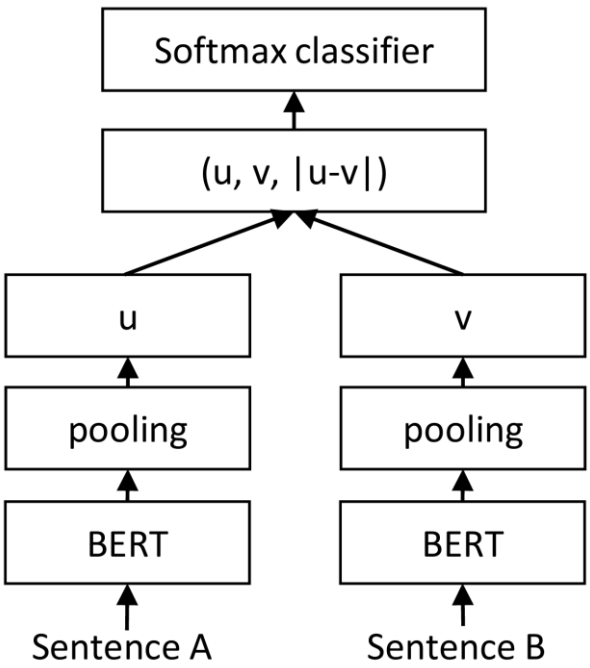


In Figure 10. Overall pre-training and fine-tuning procedures for BERT, sentences A and B are separated using the format '[CLS] Sentence A [SEP] Sentence B [SEP]'. The sentences are converted into vectors based on the BERT vocabulary, segment information is added to distinguish sentence A from sentence B, and positional information is assigned to the input tokens. The embedded sentences pass through the same Transformer network, and the vector value assigned to the [CLS] token is calculated to determine sentence similarity[36]. Due to the simultaneous vectorization of both sentences, generating single sentence embeddings is inefficient, and the computation demand significantly increases as encoding and decoding are performed on the same Transformer network.

To overcome these limitations, Sentence-BERT modifies the pre-trained BERT model using a siamese network and a triplet network. The SBERT model changes the network structure to facilitate large-scale semantic similarity comparison, clustering, and information retrieval (via semantic search) tasks that the BERT model has not been able to handle. The SBERT model adopts the Bi-encoder structure to independently generate sentence embeddings, pre-compute them, and store them, enabling efficient similarity search and analysis even for large-scale datasets. On the other hand, the Cross-encoder receives sentence pairs at once and has high accuracy but is slow to process and inefficient for large-scale tasks. SBERT proves to be a suitable model for these tasks, owing to the Siamese network and triplet network, which enhance the efficiency of the Bi-encoder and improve the accuracy of similarity analysis and information retrieval[27].

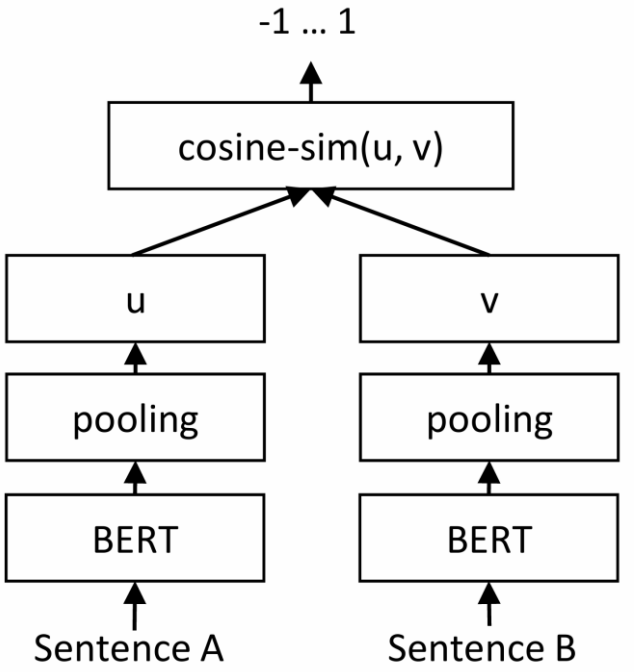
The Siamese network architecture generates fixed-size vectors for input sentences, enabling effective comparison of textual similarity. In Figure 11. Siamese Network, it consists of twin networks, each designed to process a pair of texts while sharing the same weights. This shared structure enables the network to produce consistent feature representations for semantically similar inputs. The network learns to encode semantically meaningful embeddings for each input[27), 28)].

<Figure 11. Siamese Network[27)]>



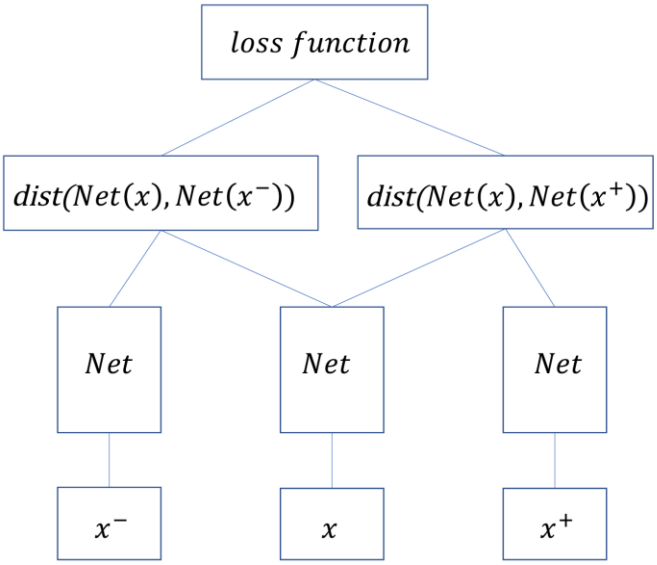
After training, the network processes a pair of sentences through the twin sub-networks independently. The similarity between the pair of sentences is computed using a distance metric, such as Cosine similarity or Euclidean distance. The network outputs a similarity score, typically between 0 and 1, indicating how semantically similar the pair of input sentences is[27), 28)]. Figure 12. Siamese Network at inference explains how the architecture computes similarity scores.

<Figure 12. Siamese Network at inference[27)]>



The triplet network uses three sentences to learn whether two sentences belong to the same class or different classes. First, one sentence is designated as the pivot, while the other two are set as its positive example and negative example, respectively. The positive example is assigned to the same class as the pivot, and the network learns to minimize the distance between their embeddings. Conversely, the negative example is assigned to a different class from the pivot, and the network learns to maximize the distance between their embeddings[27), 29)].

<Figure 13. Triplet Network[29)]>



This process ensures that similar sentences are represented closer together in the vector space, while dissimilar sentences are mapped farther apart, enabling the model to effectively distinguish semantic relationships. As illustrated in figure 13. Triplet Network, the triplet network adjusts embeddings through a loss function, optimizing the relative distances among the pivot, positive, and negative samples. By leveraging this mechanism, the model improves its performance on tasks such as semantic clustering and classification.

The Sentence-BERT model, 'all-mpnet-base-v2', is loaded to create sentence embeddings in a Python 3.10.0 environment with the following installed libraries: 'sentence-transformers' for generating embeddings, 'nltk' for tokenizing text into sentences, and 'docx2txt' for extracting text from Word files.

Text extracted from Word files is tokenized into individual sentences to ensure that each sentence is processed independently for embedding. Using the 'all-mpnet-base-v2' model, these sentences are converted into dense vector representations. To represent the entire document, the sentence embeddings are aggregated by calculating their mean, resulting in document-level embeddings. These document-level embeddings are used for similarity calculations.

3. AAMI TIR 102:2019

The Association for the Advancement of Medical Instrumentation (AAMI) is founded as a nonprofit organization in 1967. It is a diverse community which consists of various professionals united by one important mission – the development, management, and use of safe and effective health technology. AAMI is the primary source of consensus standards in both U.S. and international for the medical device industry as well as practical information, support, and guidance for healthcare technology and sterilization professionals. The AAMI also provides many educational programs for medical industry personnel to review and study various standards e.g. Human Factors for Medical Devices, ANSI/AAMI SW96:2023 – Security Risk Management Guidance, Integrating Risk Management into the Product Life Cycle[30].

FDA has recognized many standards from AAMI e.g. ANSI/AAMI SW96:2023 – Security Risk Management Guidance, AAMI TIR12:2020/(R)2023 Designing, testing, and labeling medical devices intended for processing by health care facilities: A guide for device manufacturers, and others[31].

And for this technical information report, it is organized into Chapter 1 Scope, Chapter 2 Using this technical information report, Chapter 3 Key considerations, which includes subsections such as Definitions. Additionally, it features two tables that provide a bidirectional comparison of 21 CFR Part 820 and ISO 13485:2016 under Chapter 3. The following points of caution are noted[32]:

First, “it is not a word-for-word literal identification of differences; thus, the reader must be familiar with quality management system requirements along with the statutory definitions to apply this report[32].”

Second, “this document provides the basis for interpretation of the associated requirements and applicable U.S. FDA rules. Users must be aware that this analysis provides a comparison of the QS Regulation and the standard only[32].”

Third, “The mapping is provided in two directions purposefully. When evaluating the two quality management systems, the full intent and similarities can only be determined by comparing in both directions. Therefore, both tables should be reviewed in their entirety. An example of this is 21 CFR 820 supplier controls compared to ISO 13485:2016 outsourced suppliers and purchasing controls and the need for quality agreements[32].”

The AAMI TIR 102:2019, titled U.S. FDA 21 CFR Mapping to the Applicable Regulatory Requirements Referenced in ISO 13485:2016 Quality Management System, serves as a resource for evaluating the alignment between the 21 CFR Part 820 and the ISO 13485:2016. This technical information report provides a detailed mapping between individual sections of 21 CFR Part 820 and corresponding clauses of ISO 13485:2016, facilitating a clear understanding of their similarities and differences. The document includes expert annotations that identify ISO 13485:2016 clauses deemed similar to specific sections of 21 CFR Part 820. These annotations are instrumental in assigning relevance scores, with identified matches receiving a score of 1 to indicate alignment and non-matches receiving a score of 0 to indicate no alignment. By leveraging this expert defined mapping, the report ensures that evaluations consider both semantic similarities and expert judgment, creating a balanced framework for assessing the alignment between the regulation and the standard. These dual perspective addresses areas where semantic similarity alone may be insufficient, providing a more comprehensive evaluation of the alignment between regulatory and standard requirements.

To illustrate how AAMI TIR 102:2019 maps regulations and standards, 'Table 4. Comparison from 21 CFR 820 to ISO 13485:2016', was created by referencing and reconstructing a table provided in AAMI TIR 102:2019. This table provides a structured comparison by listing sections from 21 CFR Part 820 alongside their corresponding requirements, the matching clauses from ISO 13485:2016, and the requirements of those clauses. Additionally, the table includes specific considerations for each regulation and standard. If there are notable distinctions, the table elaborates on the explicit requirements of the regulation or standard. Otherwise, it states "No significant difference," indicating alignment without substantial divergence. This structured approach enables readers to clearly understand how AAMI TIR 102:2019 facilitates the alignment of 21 CFR Part 820 with ISO 13485:2016 through detailed mappings and considerations[32)].

<Table 4. Comparison from 21 CFR 820 to ISO 13485:2016[32)]>

	21 CFR Part 820	Requirement	ISO 13485:2016	Requirement	U.S. FDA Quality System considerations	ISO 13485:2016 considerations
1.	820.5 Quality System	Content is copied from 21 CFR Part 820.5	4.1.1 Quality management system, General requirements	Content is copied from ISO 13485:2016 4.1.1	No significant difference.	No significant difference.

4. Cosine Similarity

The generated document embeddings which are semantically meaningful can be compared with cosine similarity. Usually, two document embeddings as referred vector are computed. For each document pair, document A and document B are passed through the BERT-based model, which produces the embeddings *u* und *v*. The similarity of these embeddings is computed using cosine similarity. The two documents are fed through the same model rather than two separate models. Figure 12. Siamese Network at inference explains how the architecture computes similarity scores[27]. However, this study prepared the 'scikit-learn' library for cosine similarity calculation in a Python 3.10.0 environment. Cosine similarity produces values in the range of [-1, 1], where 1 indicates maximum similarity, 0 indicates no similarity, and -1 indicates maximum dissimilarity. For this study, cosine similarity scores are normalized to the range [0, 1] to simplify the interpretation of results, where 1 represents maximum similarity and 0 represents maximum dissimilarity[33].

In particular, the cosine similarity for similar documents is maximized (1.0) and the cosine similarity for dissimilar documents is minimized (0.0) [33]. To evaluate the semantic similarity between the individual 21 CFR Part 820 section and ISO 13485:2016 clauses, the followings criteria are established based on the average cosine similarity scores:

$1 \geq$ the average cosine similarity ≥ 0.8, the 21 CFR Part 820 section is deemed very high similar to the corresponding ISO 13485:2016 clauses, and it is highly considered substitutable.

$0.8 >$ the average cosine similarity ≥ 0.6, the 21 CFR Part 820 section is deemed moderate similar to the corresponding ISO 13485:2016 clauses, and it is moderately considered substitutable.

$0.6 >$ the average cosine similarity, the 21 CFR Part 820 section is deemed dissimilar to the corresponding ISO 13485:2016 clauses, and it is not considered substitutable.

5. NDCG

This metric was developed as people deliberated how to have a good ranking function and how to design it to assess its performance as the role of ranking, which is necessary in search engines, recommendation systems, and expert search, emerged. While classification and regression (similarity) are simple and have natural performance measures, designing an optimal ranking evaluation method was very difficult compared to them. However, NDCG can assign a relevance rating to each retrieved document, which has a distinct advantage over most ranking measures that only allow binary relevance[34]. In addition, NDCG is different from other measures that uniformly weigh all rankings, including a discount function based on rank. Therefore, this function accounts for the user's preference for results with higher ranks, indicating that results ranked closer to the top are considered more important than those ranked lower. The NDCG value is expressed as DCG/IDCG between 0 and 1[35].

$$\text{NDCG} = \frac{DCG}{IDCG} \quad (1)$$

After the relevance score is calculated in advance, the DCG value adjusts the relevance score according to the position of the desired result in the output ranking list. If the user's desired results are ranked at the top, the DCG value increases, and if the desired results are ranked lower, the DCG value decreases according to their rank[35].

$$\text{DCG} = \sum_{i=1}^{n} \frac{rel_i}{\log_2(i+1)} \quad (2)$$

Therefore, if the user's desired results are ranked lower, the ranked relevance value decreases logarithmically. Similarly, after the relevance score is pre-calculated, the IDCG value is the ideal rank DCG value where the most relevant items are placed at the top[35].

$$\text{IDCG} = \sum_{i=1}^{n} \frac{rel_i}{\log_2(i+1)} \quad (3)$$

In this study, relevance scores are assigned using expert evaluations provided in AAMI TIR 102:2019. Specifically, ISO 13485:2016 clauses that experts have identified as similar to individual 21 CFR Part 820 sections are assigned a relevance score of 1, while all other ISO 13485:2016 clauses are assigned a relevance score of 0. Using these relevance scores and the cosine similarity values calculated between each individual 21 CFR Part 820 section and all ISO 13485:2016 clauses, this study derives NDCG values. The NDCG values provide an objective measure of how well the cosine similarity rankings align with

the expert selected rankings from AAMI TIR 102:2019, ensuring that the evaluation accounts for both semantic similarity and expert opinion.

The following criteria are established based on the NDCG value:

$1 \geq$ NDCG value ≥ 0.8, the agreement between the AAMI expert annotations (Expert-Labeled) and the calculated cosine similarity demonstrates high agreements. The ranking of ISO 13485:2016 clauses, based on their assigned relevance to each 21 CFR Part 820 section as provided in AAMI TIR 102:2019, aligns closely with the cosine similarity calculations. This high alignment indicates that the matched clauses are highly substitutable.

$0.8 >$ NDCG value ≥ 0.6, the agreement between the AAMI expert annotations (Expert-Labeled) and the calculated cosine similarity demonstrates partial agreements. The ranking of ISO 13485:2016 clauses, based on their assigned relevance to each 21 CFR Part 820 section as provided in AAMI TIR 102:2019, partially aligns with the cosine similarity calculations. This moderate alignment suggests that the matched clauses are moderately substitutable

$0.6 >$ NDCG value, the agreement between the AAMI expert annotations (Expert-Labeled) and the calculated cosine similarity demonstrates low agreements. The ranking of ISO 13485:2016 clauses, based on their assigned relevance to each 21 CFR Part 820 section as provided in AAMI TIR 102:2019, shows little to no alignment with the cosine similarity calculations. This low alignment suggests that the matched clauses are not substitutable.

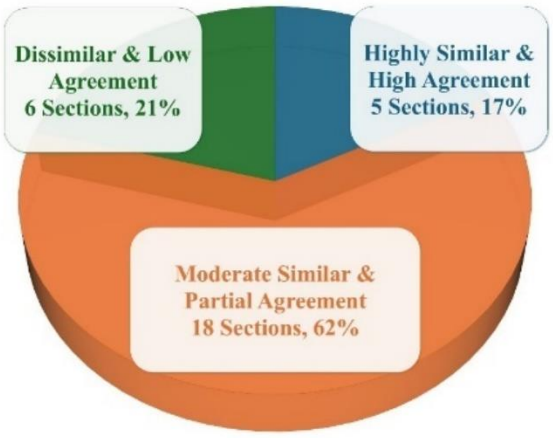
To ensure a conservative evaluation, when the Cosine Similarity and NDCG values are used in combination, in cases where discrepancies exist between the two metrics, the lower value is prioritized to avoid overestimation of similarity. This approach ensures that the assessment remains cautious and robust.

Therefore, by evaluating both metrics independently, this study identifies cases where semantic similarity and expert judgments may align or diverge. The combined use of both metrics ensures a robust and conservative evaluation, minimizing the risk of overestimating substitutability. This multidimensional approach provides deeper insights into the alignment between regulatory sections and clauses.

III. Results

Among the 29 sections, in Figure 14. Similarity & NDCG between 21 CFR Part 820 and ISO 13485:2016, Sections 820.30, 70, 72, 90, and 180 showed a very high similarity (above 0.8) with matched ISO 13485:2016 clauses (Table 5. Similarity 21 CFR Part 820 Sections with ISO 13485:2016 Clauses), demonstrating a high agreement with the annotations in AAMI TIR 102:2019. These sections can be substituted with the matched ISO 13485:2016 clauses when establishing each section. Additionally, Sections 820.20, 22, 25, 40, 50, 60, 75, 80, 86, 100, 120, 130, 150, 160, 170, 184, 198, and 200 exhibited moderate similarity (between 0.6 and 0.8) with matched ISO 13485:2016 clauses (Table 5. Similarity 21 CFR Part 820 Sections with ISO 13485:2016 Clauses), indicating partial agreement with the annotations in AAMI TIR 102:2019. These sections appear substitutable with the matched ISO 13485:2016 clauses, but further review is required. However, Sections 820.5, 65, 140, 181, 186, and 250 displayed dissimilarity (below 0.6) with matched ISO 13485:2016 clauses (Table 5. Similarity 21 CFR Part 820 Sections with ISO 13485:2016 Clauses), showing low agreement with the annotations in AAMI TIR 102:2019. These sections cannot be substituted with the matched ISO 13485:2016 clauses and must be newly established.

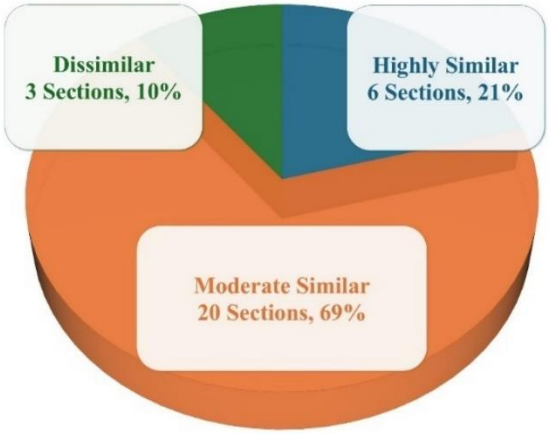
<Figure 14. Similarity & NDCG between 21 CFR Part 820 and ISO 13485:2016>



Next, Cosine similarity and NDCG values were analyzed independently without combining them. Frist, in Figure 15. Similarity between 21 CFR Part 820 and ISO 13485:2016, the sections categorized as very high are 21 CFR Part 30, 50, 70, 72, 90, and

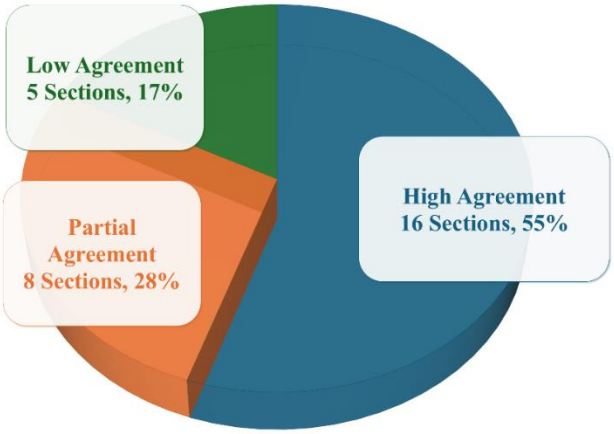
180. The sections categorized as moderate are 21 CFR Part 20, 22, 25, 40, 60, 65, 75, 80, 86, 100, 120, 130, 150, 160, 170, 181, 184, 198, 200, and 250. The sections categorized as dissimilar are 21 CFR Part 5, 140, and 186.

<Figure 15. Similarity between 21 CFR Part 820 and ISO 13485:2016>



Second, in Figure 16. NDCG between 21 CFR Part 820 and ISO 13485:2016, the sections categorized as high agreement with cosine similarity are 21 CFR Part 20, 22, 30, 40, 70, 72, 75, 90, 100, 120, 130, 140, 150, 170, 180, and 184. The sections categorized as partial agreement with cosine similarity are 21 CFR Part 25, 50, 60, 80, 86, 160, 198, and 200. The sections categorized as low agreement are 21 CFR Part 5, 65, 181, 186, and 250.

<Figure 16. NDCG between 21 CFR Part 820 and ISO 13485:2016>



<Table 5. Similarity 21 CFR Part 820 Sections with ISO 13485:2016 Clauses>

Rank	Section	Clauses	Cos. Sim.	Expert Labeled	NDCG
31	820.5.	4.1.1.	0.55	1	0.2
1	820.20.	5.3.	0.86	1	0.84
2		5.1.	0.85	1	
3		4.2.2.	0.83	1	
6		5.5.2.	0.82	1	
7		6.1.	0.82	1	
12		4.2.1.	0.78	1	
17		5.4.2.	0.77	1	
19		5.6.3.	0.76	1	
22		5.6.1.	0.75	1	
25		4.1.1.	0.74	1	
26		5.4.1.	0.74	1	
31		5.5.1.	0.72	1	
33		5.6.2.	0.71	1	
36		8.2.4.	0.70	1	
1	820.22.	8.2.4.	0.85	1	0.83
6		5.6.2.	0.71	1	

Rank	Section	Clauses	Cos. Sim.	Expert Labeled	NDCG
1	820.25.	6.2.	0.80	1	0.72
8		6.4.1.	0.69	1	
19		6.1.	0.65	1	
1	820.30.	7.3.3.	0.86	1	0.98
2		7.3.4.	0.85	1	
3		7.3.2.	0.84	1	
4		7.3.7.	0.84	1	
5		7.3.6.	0.83	1	
6		7.3.5.	0.80	1	
7		7.3.9.	0.77	1	
9		7.3.8.	0.76	1	
13		7.3.10.	0.74	1	
15		7.3.1.	0.73	1	
1	820.40.	4.2.4.	0.86	1	0.81
8		4.2.5.	0.70	1	
1	820.50.	7.4.1.	0.87	1	0.78
6		7.4.3.	0.80	1	
7		7.4.2.	0.79	1	

Rank	Section	Clauses	Cos. Sim.	Expert Labeled	NDCG
1	820.60.	7.5.8.	0.73	1	0.74
6	820.65.	7.5.9.1.	0.65	1	0.38
10		7.5.1.	0.64	1	
1	820.70.	7.5.1.	0.86	1	0.87
2		6.3.	0.85	1	
3		7.5.6.	0.82	1	
4		7.6.	0.82	1	
5		4.1.4.	0.80	1	
1	820.72.	7.6.	0.89	1	1
1	820.75.	7.5.6.	0.86	1	0.8
3		7.5.7.	0.81	1	
27		8.2.5.	0.70	1	
1	820.80.	8.2.6.	0.84	1	0.66
5		7.4.3.	0.75	1	
11		7.1.	0.71	1	
25		7.5.8.	0.67	1	
27		8.3.1.	0.66	1	
68		7.5.10.	0.43	1	

Rank	Section	Clauses	Cos. Sim.	Expert Labeled	NDCG
2	820.86.	7.5.8.	0.65	1	0.63
1	820.90.	8.3.1.	0.90	1	0.88
2		8.3.4.	0.81	1	
13		8.3.2.	0.74	1	
1	820.100.	8.5.2.	0.83	1	0.83
2		8.5.3.	0.82	1	
4		7.5.6.	0.78	1	
5		8.4.	0.78	1	
7		4.1.4.	0.77	1	
23		8.5.1.	0.71	1	
30		7.3.9.	0.70	1	
1	820.120.	7.5.1.	0.75	1	0.96
2		7.5.8.	0.72	1	
3		7.5.11.	0.70	1	
4		8.2.6.	0.67	1	
6		7.5.9.2.	0.65	1	
13		6.3.	0.61	1	

Rank	Section	Clauses	Cos. Sim.	Expert Labeled	NDCG
1	820.130.	7.5.11.	0.66	1	0.85
4		7.5.1.	0.54	1	
1	820.140.	7.5.11.	0.62	1	0.82
7		6.3.	0.48	1	
1	820.150.	7.5.11.	0.77	1	1
2	820.160.	7.5.9.2.	0.71	1	0.63
4		7.5.11.	0.69	1	
7		7.2.1.	0.66	1	
11		7.2.2.	0.64	1	
25		7.1.	0.58	1	
47		7.5.9.1.	0.51	1	
1	820.170.	7.5.3.	0.75	1	1
1	820.180.	4.2.5.	0.83	1	1
2		4.2.4.	0.76	1	
5	820.181.	4.2.3.	0.60	1	0.39
1	820.184.	7.5.1.	0.74	1	0.87
2		7.5.8.	0.68	1	
18		7.1.	0.58	1	

Rank	Section	Clauses	Cos. Sim.	Expert Labeled	`NDCG
30	820.186.	4.2.5.	0.56	1	0.24
40		4.2.4.	0.53	1	
1	820.198.	8.2.2.	0.89	1	0.74
5		8.2.3.	0.71	1	
36		4.2.5.	0.58	1	
1	820.200.	7.5.4.	0.79	1	0.66
7		7.5.8.	0.69	1	
21		8.1.	0.63	1	
41		8.2.3.	0.58	1	
9	820.250.	8.1.	0.69	1	0.3

In Table 5. Similarity Between 21 CFR Part 820 Sections and ISO 13485:2016 Clauses, this study examines the 21 CFR Part 820 sections with the highest and lowest cosine similarity and NDCG scores. Additionally, sections where the difference between cosine similarity and NDCG scores exceeds 0.2 are identified and analyzed to highlight notable discrepancies.

Section 820.30 (Design Control) contains the most sub-sections among the sections of 21 CFR Part 820, and Clause 7.3 (Design & Development) of ISO 13485:2016 similarly includes the second most sub-clauses among its clauses, following Clause 7.5 (Production and Service Provision). Despite the extensive content available for comparison, the high similarity score of 0.83 and NDCG value of 0.98 indicate high agreement with AAMI expert annotations (Expert-Labeled). Although the exact reasons for these results were not qualitatively analyzed in this study, the high agreement may stem from the fact that both sections cover processes involved in the design, development, transfer, and modification of medical devices. The high similarity and NDCG values in such a pivotal area for medical

device creation and production suggest alignment between 21 CFR Part 820 and ISO 13485:2016, which could be advantageous for manufacturers adhering to ISO 13485:2016 who aim to meet U.S. regulatory requirements.

Section 820.72 (Inspection, Measuring, and Test Equipment) specifies procedures, calibration intervals, and acceptance criteria for managing equipment such as measurement or calibration devices and temperature-humidity meters that require accurate and precise readings for product and environment testing. Similarly, ISO 13485:2016 Clause 7.6 (Control of Monitoring and Measuring Equipment) addresses requirements for equipment corresponding to those described in 820.72. Although this study does not qualitatively analyze the reasons for the high similarity score and NDCG value, the lack of overlap in scope and content with other sections or clauses may have contributed to the observed results. Therefore, the high similarity and NDCG values in such a pivotal area for inspection, measuring, and test equipment suggest alignment between 21 CFR Part 820 and ISO 13485:2016, which could be advantageous for manufacturers adhering to ISO 13485:2016 who aim to meet U.S. regulatory requirements.

Sections 820.5 (Quality System), 820.186 (Quality System Record), and 820.250 (Statistical Techniques) address different aspects of quality management. Section 820.5 mandates the establishment of a quality system, while 820.186 requires the maintenance and management of records related to quality system activities and medical device records. Section 820.250 specifies the need for analytical techniques to manage process capability and product characteristics, including defining and applying appropriate sampling methods when used. These three sections have relatively low minimum similarity scores of 0.55, 0.53, and 0.69, respectively, and their NDCG values 0.2, 0.24, and 0.3 are significantly lower than those of other 21 CFR Part 820 sections. Manufacturers comparing ISO 13485:2016 and 21 CFR Part 820 should pay particular attention to compliance with these three regulations.

Interestingly, Sections 820.5, 65, 120, 130, 140, 150, 170, 180, 181, 184, 186, and 250 exhibit significant differences between their NDCG values and cosine similarity scores. Specifically, the differences calculated as (NDCG value - cosine similarity) are as follows: -0.35, -0.27, 0.28, 0.25, 0.27, 0.23, 0.25, 0.20, -0.21, 0.20, -0.21, and -0.39, respectively.

This result indicates a notable inconsistency, as sections with lower cosine similarity scores still achieve high NDCG values. However, it is essential not to focus solely on cases of low similarity, as there are also instances where NDCG values are lower than cosine similarity, highlighting a different type of misalignment. The need to investigate 21 CFR

Part 820 sections where the difference between NDCG and cosine similarity exceeds ±0.2 arises from the following reasons:

First, defined ranges for cosine similarity and substitutability are used. When the average cosine similarity is between 1 and 0.8, the 21 CFR Part 820 section is considered very high similar to the corresponding ISO 13485:2016 clauses and is deemed highly substitutable. If the cosine similarity falls between 0.8 and 0.6, the section demonstrates moderate similarity to the corresponding clauses and remains moderately substitutable. However, when the cosine similarity is below 0.6, the section is considered dissimilar to the corresponding ISO 13485:2016 clauses and is not substitutable. These thresholds allow for consistent interpretation of semantic similarity.

Second, defined ranges for NDCG and agreement with expert annotations (Expert-Labeled) are applied. When the NDCG value is between 1 and 0.8, it reflects a high level of agreement between the calculated cosine similarity and the AAMI expert annotations, indicating high similarity and substitutability. If the NDCG value falls between 0.8 and 0.6, there is partial agreement, suggesting moderate substitutability. For NDCG values below 0.6, low agreement between the metrics indicates dissimilarity, meaning the section is not substitutable.

Third, to ensure a balanced and conservative evaluation, discrepancies where $|\text{NDCG} - \text{Cosine Similarity}| > 0.2$ are prioritized. In such cases, the lower value is considered to avoid overestimating similarity. This approach accounts for both instances where NDCG exceeds cosine similarity and where NDCG falls below cosine similarity, reflecting potential misalignments between semantic similarity and expert annotations. By evaluating both metrics independently, this study identifies cases where semantic similarity and expert judgments either align or diverge. Combining both metrics ensures a multidimensional evaluation framework, offering deeper insights into the alignment between regulatory sections and clauses.

IV. Discussion

This study differs from prior research that analyzed the similarities and differences between 21 CFR Part 820 and ISO 13485:2003 to establish a comprehensive quality management system without omitting any regulatory or standard requirements. Instead, this study quantitatively analyzes the similarities between 21 CFR Part 820 and ISO 13485:2016 to help manufacturers in the Republic of Korea and Japan determine which sections of 21 CFR Part 820 can be substituted with corresponding clauses of ISO 13485:2016 when exporting medical devices to the U.S. or manufacturing them in the U.S.

Based on the results, three main findings can be derived. First, organizations establishing a quality system can expect to replace Sections 820.30, 70, 72, 90, and 180 with similar clauses of ISO 13485:2016, as these sections have both similarity and NDCG scores exceeding 0.8. Second, for Sections 820.20, 22, 25, 40, 50, 60, 75, 80, 86, 100, 120, 130, 150, 160, 170, 184, 198, and 200, the similarity and NDCG scores are both less than 0.8 but greater than 0.6. Therefore, these sections can be potentially replaced by their matched clauses of ISO 13485:2016 but need to be reviewed and adjusted to implement a quality system effectively. Third, Sections 820.5, 65, 140, 181, 186, and 250 cannot be replaced by clauses of ISO 13485:2016 as their similarity and NDCG scores are both below 0.6. Therefore, organizations must establish new processes and procedures for sections 820.5, 65, 140, 181, 186, and 250, then incorporated into the existing quality management system. Lastly Sections 120, 130, 150, 160, 170, and 184, as presented in Table 5. Similarity of 21 CFR Part 820 Sections with ISO 13485:2016 Clauses, indicate moderate similarity ($0.8 > \text{Cosine Similarity} \geq 0.6$) and partial agreement ($0.8 > \text{NDCG} \geq 0.6$) with the matched ISO 13485:2016 clauses. However, they exhibit outlier characteristics compared to other sections due to $|\text{NDCG} - \text{Cosine Similarity}| > 0.2$.

A practical perspective, rather than interconnecting all 29 sections to establish a new quality system to export medical devices to U.S. or manufacture them in U.S. It would be more resource efficient to prioritize the analysis and establishment of the 6 sections with low similarity and NDCG scores both below 0.6. The following work is that organizations review the next 18 sections, and finally address the remaining 5 sections. Particularly, when manufacturers in the Republic of Korea and Japan aim to develop quality systems complying with 21 CFR Part 820, they should carefully review the Cosine Similarity and NDCG values presented in Table 5. Similarity of 21 CFR Part 820 Sections with ISO 13485:2016 Clauses, especially for Sections 120, 130, 150, 160, 170, and 184, which

exhibit outlier characteristics (|NDCG − Cosine Similarity| > 0.2), rather than focusing on Sections 820.20, 22, 25, 40, 50, 60, 75, 80, 86, 100, 198, and 200. Manufacturers must consider whether to substitute the ISO 13485:2016 clauses for Sections 120, 130, 150, 160, 170, and 184 or to integrate the requirements of these sections into their existing quality systems.

It is assumed that these results were caused by three primary limitations in this study. First, the terms defined and used in 21 CFR Part 820 and ISO 13485:2016 were not fully harmonized. Second, relevance scores were evaluated in a binary manner rather than on a multi-level scale. Finally, the SBERT model (all-mpnet-base-v2) was not fine-tuned using data from medical devices, relevant regulations, or publications issued by international organizations. Consequently, only 5 sections (17% of the 29 sections) – Sections 820.30, 70, 72, 90, and 180 – were found to be highly similar to ISO 13485:2016, with agreement levels consistent with the annotations in AAMI TIR 102:2019.

These three factors represent the primary limitations of this study. If the terms defined and used in 21 CFR Part 820 and ISO 13485:2016 were harmonized, and the qualitative similarities between each section of 21 CFR Part 820 and all ISO 13485:2016 clauses were expressed on a multi-level scale, the relevance scores would have been evaluated on a more granular level. Moreover, fine-tuning the SBERT model (all-mpnet-base-v2) using data from medical devices, relevant regulations, or publications issued by international organizations would likely result in more than six sections being identified as highly similar to ISO 13485:2016. In such cases, the high similarity could be reflected in multi-level relevance scores and yield a greater level of agreement when compared with qualitative similarity annotations.

Future research could address these limitations by harmonizing terms between 21 CFR Part 820 and ISO 13485:2016, implementing multi-level relevance scoring, and fine-tuning SBERT models using domain-specific datasets. This would likely result in more than six sections being identified as highly similar to ISO 13485:2016, reflected in multi-level relevance scores and higher agreement levels with qualitative annotations. Such improvements would provide manufacturers in Korea and Japan with more precise guidance on which sections of 21 CFR Part 820 to prioritize when developing quality systems compliant with U.S. regulations.

V. Conclusion

This study quantitatively analyzed the similarities between 21 CFR Part 820 and ISO 13485:2016, providing analytical insights into the potential substitution of clauses between the regulations and standard. It highlights opportunities to leverage ISO 13485:2016 to reduce the efforts required to meet 21 CFR Part 820 requirements. Unlike previous studies that focused on integration or harmonization, this study emphasizes minimizing additional workload for regulatory compliance.

The Republic of Korea and Japan have regulations that integrate additional requirements into their QMS based on ISO 13485:2016. These systems mandate that companies manufacturing medical devices or participating in one or more stages of the medical device lifecycle continuously produce effective and safe medical devices that meet regulatory and customer requirements. In contrast, the United States enforces its own regulations under 21 CFR Part 820, requiring companies manufacturing finished medical devices and components or engaging in one or more stages of the lifecycle to adhere to these regulations to ensure the continuous production of effective and safe medical devices and components.

This study also investigated natural language processing (NLP) techniques to enable semantic comparisons between clauses in ISO 13485:2016 and sections in 21 CFR Part 820. Advances in NLP, including neural networks, word embedding techniques, and Transformer models, have significantly improved the ability to analyze human language and compare texts. Among NLP models, Sentence-BERT was chosen for its strong performance in sentence similarity analysis (Cosine Similarity). Sentence-BERT utilizes Siamese and triplet networks to learn from large corpora, making it well-suited for this task.

To address the limitations of Cosine similarity analysis without a defined threshold and to strengthen the robustness of the study, this study integrated the Normalized Discounted Cumulative Gain (NDCG) metric, widely used in ranking systems, with the bidirectional mapping methodology described by AAMI experts in AAMI TIR 102:2019. By combining the Cosine Similarity metric with the expert annotations of AAMI, this study ensures a comprehensive evaluation of the alignment between 21 CFR Part 820 and ISO 13485:2016. Additionally, the study establishes score ranges for Cosine similarity and NDCG to address the limitations of similarity analysis models that lack defined thresholds.

In conclusion, this study provides practical metrics and insights for companies exporting medical devices to the U.S. or manufacturing them in the U.S. It enables medical device manufacturers in Korea and Japan to minimize changes to their existing quality management systems while meeting U.S. regulatory requirements. By simplifying the compliance process and reducing additional burdens, this study offers a pathway for manufacturers to streamline their regulatory adherence efforts effectively.

References

1. Figueroa, Yazmín. "Comparison of the Quality System Requirements of Code of Federal Regulations Part 820 and International Standard ISO 13485." *Revista Politechne;* (2010).

2. Ministry of health and welfare. “Medical Device Act”. *Ministry of Food and Drug Safety (MFDS)*. 22.Oct.2024.

3. Ministry of Food and Drug Safety (MFDS). “Standards for Facilities and Manufacturing and Quality Management System”. *Medical Device Management Division*. 19.Dec.2023.

4. Ministry of Food and Drug Safety (MFDS). “Medical Device GMP International Quality Management Guide”. *Medical Device Safety Evaluation Division*. Dec.2017.

5. Ministry of Food and Drug Safety (MFDS). “A Comprehensive Guide to Medical Device GMP (Revision 8)”. *Medical Device Management Division*. 09.Nov.2023.

6. Ministry of Health, Labour and Welfare (MHLW). “Ministerial Ordinance on Standards for Manufacturing Control and Quality Control for Medical Devices and In-Vitro Diagnostics (MHLW MO169)”. *Pharmaceuticals and Medical Devices Agency (PMDA)*. 26.Mar.2021.

7. Pharmaceuticals and Medical Devices Agency (PMDA). “QMS regulation in Japan”. *Office of Manufacturing/Quality and Compliance*. 01.Apr.2015.

8. Pharmaceuticals and Medical Devices Agency (PMDA). “Revision of Japanese Medical Device QMS requirements, 2. Contents of MHLW MO169”. *PMDA*, n.d. https://www.pmda.go.jp/english/review-services/regulatory-info/0004.html. Accessed 20.Nov.2024

9. Food and Drug Administration. “21 CFR Part 820”. *Code of Federal Regulations*. 02.Apr.2020. https://www.ecfr.gov/current/title-21/chapter-I/subchapter-H/part-820. Accessed 20.Nov.2024.

10. Food and Drug Administration. “Quality System (QS) Regulation/Medical Device Current Good Manufacturing Practices (CGMP)”. *FDA*. 31.Jan.2024. https://www.fda.gov/medical-devices/postmarket-requirements-devices/quality-system-qs-regulationmedical-device-current-good-manufacturing-practices-cgmp. Accessed 20.Nov.2024.

11. Khurana, Diksha, et al. "Natural language processing: state of the art, current trends and challenges." *Multimedia tools and applications* 82.3 (2023): 3713-3744.

12. 홍정화, et al. "텍스트마이닝 기반 안전사고 관련 규정· 법령의 자동 매칭." 대한산업공학회 춘계공동학술대회 논문집 (2024): 563-570.

13. 박길식, and 김준태. "S-BERT를 이용한 법률 검색." *한국컴퓨터정보학회 논문지* 27.11 (2022): 57-66.

14. Agirre, Eneko, et al. "Semeval-2014 task 10: Multilingual semantic textual similarity." *Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014)*. 2014.

15. Wibawa, Aji Prasetya, and Fachrul Kurniawan. "Advancements in natural language processing: Implications, challenges, and future directions." *Telematics and Informatics Reports* 16 (2024): 100173.

16. Sharma, Kabir. “30 Years of Research on Semantic Similarity Measurement.” OSF Preprints, 16 May 2023. Web.

17. Lee, Jae-Seong. "심층 신경망의 발전 과정과 이해." *Information and Communications Magazine* 33.10 (2016): 40-48.

18. Lee, Seung-Cheol, et al. "Deep learning." *Journal of KSNVE* 27.3 (2017): 19-25.

19. Abhishek Jha. “Vectorization Techniques in NLP [Guide]”. *Neptune Blog*. 11.Aug.2023. https://neptune.ai/blog/vectorization-techniques-in-nlp-guide. Accessed 27.Nov.2023.

20. Mikolov, Tomas. "Efficient estimation of word representations in vector space." *arXiv preprint arXiv:1301.3781* 3781 (2013).

21. Mikolov, Tomas, et al. "Distributed representations of words and phrases and their compositionality." *Advances in neural information processing systems* 26 (2013).

22. Vaswani, A. "Attention is all you need." *Advances in Neural Information Processing Systems* (2017).

23. Zhang, Emma Yann, et al. "From turing to transformers: A comprehensive review and tutorial on the evolution and applications of generative transformer models." *Sci* 5.4 (2023): 46.

24. Nieminen, Matias. "The Transformer Model and Its Impact on the Field of Natural Language Processing." (2023).

25. Food and Drug Administration. “Quality Systems”. *FDA*. 28.Mar.2023. https://www.fda.gov/inspections-compliance-enforcement-and-criminal-investigations/inspection-guides/quality-systems. Accessed 20.Nov.2024.

26. ISO/TC 210. “ISO 13485:2016 Medical devices — Quality management systems —

Requirements for regulatory purposes". *International Organization for Standardization (ISO)*. Mar.2016.

27. Reimers, N. "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks." *arXiv preprint arXiv:1908.10084* (2019).
28. Chopra, Sumit, Raia Hadsell, and Yann LeCun. "Learning a similarity metric discriminatively, with application to face verification." *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*. Vol. 1. IEEE, 2005.
29. Ein-Dor, Liat, et al. "Learning thematic similarity metric using triplet networks." *Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics (ACL), 2018.
30. Association for the Advancement of Medical Instrumentation (AAMI). "About AAMI". *AAMI*. n.d. https://www.aami.org/about-aami/about-aami. Accessed 27.Nov.2024
31. Food and Drug Administration. "Recognized Consensus Standards: Medical Devices". FDA. 07.Oct.2024. https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfstandards/results.cfm. Accessed 20.Nov.2024.
32. Association for the Advancement of Medical Instrumentation (AAMI). "AAMI TIR102:2019; U.S. FDA 21 CFR mapping to the applicable regulatory requirement references in ISO 13485:2016 Quality Management Systems". *AAMI*. 30.Aug.2019
33. Scikit-learn: Machine Learning in Python. "6.8.1. Cosine Similarity." *Scikit-learn User Guide, version 1.5.2*. 09.2024, https://scikit-learn.org/1.5/modules/metrics.html#cosine-similarity. Accessed 20.Nov.2024.
34. Wang, Yining, et al. "A theoretical analysis of NDCG type ranking measures." *Conference on learning theory*. PMLR, 2013.
35. Järvelin, Kalervo, and Jaana Kekäläinen. "Cumulated gain-based evaluation of IR techniques." *ACM Transactions on Information Systems (TOIS)* 20.4 (2002): 422-446.
36. Devlin, Jacob. "Bert: Pre-training of deep bidirectional transformers for language understanding." *arXiv preprint arXiv:1810.04805* (2018).

Abstract in Korean

21 CFR Part 820과 ISO 13485:2016의 정량적 비교

21 CFR Part 820와 ISO 13485:2016을 정성적으로 비교하여 유사점과 차이점을 도출하고, 규제와 표준을 통합하여 요구사항을 모두 포함하는 품질경영시스템을 수립하는 접근 방식이 있었다. 본 연구는 21 CFR Part 820와 ISO 13485:2016의 유사성을 정량적으로 분석하고, 21 CFR Part 820 요구사항을 대체할 수 있는 ISO 13485:2016 요구사항을 도출하는 방법을 제시한다.

한국과 일본은 ISO 13485:2016을 기반으로 추가 요구사항을 통합하여 의료기기의 시설 및 품질에 대한 표준을 수립했으며, 해당 국가에서 의료 기기를 생산하거나 관련 활동에 참여하는 조직과 사람들을 규제하기 위한 추가 요구사항을 제시하고 있다. 반면 미국은 21 CFR Part 820이라는 규정에 따라 의료 기기를 생산하거나 관련 활동에 참여하는 조직과 사람들을 규제하고 있다.

자연어 처리란 번역, 검색, 비교, 분석, 추출 및 생성에 사용되는 인간과 컴퓨터의 의사소통의 프로세스 및 결과이다. 과거에는 자연어 처리가 단어의 발생 빈도와 변환된 벡터 값에 의존했으나, 오늘날의 최첨단 자연어 처리에서는 문장에 "John"이 나오는지, 누군가의 친구 John인지, 아니면 성경 속 인물 John인지 판별할 수 있다. 이는 컴퓨터가 복잡한 맥락을 이해하는 능력을 향상시킨 Neural Network, 벡터화된 단어 표현을 학습할 수 있는 Word2Vec의 출현 덕분에 가능하다. 그리고 Transformers 모델의 출현으로 입력의 순차적 순서를 유지하면서 병렬적인 어텐션 메커니즘을 통해 모든 입력 값을 처리가 가능하다.

본 연구는 Transformer 아키텍처를 통합하고 문장 임베딩을 생성할 수 있는 Sentence-BERT를 사용하였으며, Cosine Similarity와 Normalized Discounted Cumulative Gain(NDCG), AAMI TIR 102:2019을 활용하여 21 CFR Part 820과 ISO 13485:2016을 비교하였다. Cosine Similarity는 의미적 유사성을 측정하는데, 1에 가까운 값은 매우 높은 유사성을 나타내고, 0에 가까운 값은 비유사성을 나타낸다. 그러나 유사성 값에 대한 사전 정의된 임계 값이 없다. 이 제한점을 해결하기 위해 AAMI TIR 102:2019와 NDCG를 도입하였다. AAMI TIR 102:2019는 21 CFR Part 820과 ISO 13485:2016 간의 양방향 매핑을 제공하는 반면 NDCG는 주로 순위가 매겨진 검색 결과의 품질을 평가하는 데 사용되는 지표이다. NDCG는 검색 결과 순서가 사용자가 원하는 결과와 얼마나 일치하는지 평가한다. 검색 결과 목록 상단에 관련성 점수가 높은 항목이 있다면, 검색 결과의 품질이 좋음을 의미한다. 본 연구에서 관련성 점수는 AAMI TIR 102:2019에 제공된 매핑

을 기반으로 할당되었으며, 일치 시 1점, 일치하지 않는 경우 0점을 부여하였다. 마지막으로, 자연어 처리 분석의 결과에 대한 임계 값이 없는 제한점을 해결하기 위해 코사인 유사도와 NDCG 값의 범위를 설정하였다.

Cosine Similarity 점수는 매우 높은 유사도(1 ~ 0.8), 중간적 유사도(0.8 ~ 0.6), 유사하지 않음(0.6 미만)의 세 가지 범주로 분류하였다. 마찬가지로 NDCG 점수는 Cosine Similarity와 AAMI TIR 102:2019의 주석과의 일치도에 따라 매우 높은 일치도(1 ~ 0.8), 부분적 일치도(0.8 ~ 0.6), 낮은 일치도(0.6 미만)의 세 가지 범주로 분류하였다.

섹션 820.30, 70, 72, 90 및 180은 ISO 13485:2016 조항과 매우 높은 유사성(0.8 이상)을 보였으며, AAMI TIR 102:2019의 주석과 높은 일치를 보였다. 이러한 섹션은 각 섹션을 수립할 때 일치하는 ISO 13485:2016 조항으로 대체할 수 있다. 또한, 섹션 820.20, 22, 25, 40, 50, 60, 75, 80, 86, 100, 120, 130, 150, 160, 170, 184, 198, 200은 ISO 13485:2016 조항과 중간 수준의 유사성(0.8 ~ 0.6)을 보였으며, AAMI TIR 102:2019의 주석과 부분적으로 일치함을 나타낸다. 이러한 섹션은 일치하는 ISO 13485:2016 조항과 대체 가능한 것으로 보이지만 추가 검토가 필요하다. 그러나 섹션 820.5, 65, 140, 181, 186 및 250은 ISO 13485:2016 조항과 불일치(0.6 미만)를 보였으며, AAMI TIR 102:2019의 주석과 낮은 일치성을 나타냈다. 이러한 섹션은 일치하는 ISO 13485:2016 조항으로 대체할 수 없으며, 해당 섹션의 요구사항을 그대로 수용하여 새로운 품질 시스템을 구축해야 한다.

본 연구에 21 CFR Part 820의 섹션과 ISO 13485:2016의 조항 사이의 유사점을 정량적으로 분석하고 평가한 결과, 대한민국 의료기기 제조업체들이 미국에 의료기기를 수출하거나 미국에서 의료기기를 제조하기 위해서 구축할 품질 시스템 요구사항 중 21 CFR Part 820의 5개 섹션은 일치하는 ISO 13485:2016의 조항으로 대체하여 품질 시스템을 구현할 수 있고, 21 CFR Part 820의 18개 섹션은 일치하는 ISO 13485:2016의 조항으로 대체 가능하지만 적절성에 대한 검토가 필요하다. 21 CFR Part 820의 6개 섹션은 일치하는 ISO 13485:2016의 조항으로 전혀 대체할 수 없고, 해당 섹션의 요구사항을 그대로 수용하여 새로운 품질 시스템을 구축해야 한다.

---

21 CFR Part 820, ISO 13485:2016, 자연어 처리, 의미적 유사성, Neural Network, 단어 임베딩, 벡터, Transformer, AAMI TIR 102:2019