

SaMD에 대한 휴리스틱 기반 사용적합성 평가 가이드라인 개발

김종엽^{1,2} · 김정현² · 김재호^{2,4*} · 정명진^{2,3,4*}

¹성균관대학교 삼성융합의과학원 의료기기산업학과

²삼성서울병원 미래의학연구원 AI연구센터

³성균관대학교 삼성융합의과학원 디지털헬스케어학과

⁴성균관대학교 의과대학 데이터융합미래의학교실

Development of Guideline for Heuristic Based Usability Evaluation on SaMD

Jong Yeop Kim^{1,2}, Junghyun Kim², Zero Kim^{2,4*} and Myung Jin Chung^{2,3,4*}

¹Department of Medical Device Management and Research, SAIHST, Sungkyunkwan University, Seoul 06355, Korea

²Medical AI Research Center, Research Institute for Future Medicine, Samsung Medical Center, Seoul 06355, Korea

³Department of Digital Health, SAIHST, Sungkyunkwan University, Seoul 06355, Korea

⁴Department of Data Convergence and Future Medicine, Sungkyunkwan University School of Medicine, Seoul 06351, Korea

(Manuscript received 31 October 2023 ; revised 5 December 2023 ; accepted 6 December 2023)

428

Abstract: In this study, we have a goal to develop usability evaluation guidelines for heuristic-based artificial intelligence-based Software as a Medical Device (SaMD) in the medical field. We conducted a gap analysis between medical hardware (H/W) and non-medical software (S/W) based on ten heuristic principles. Through severity assessments, we identified 69 evaluation domains and 112 evaluation criteria aligned with the ten heuristic principles. Subsequently, we categorized each evaluation domain into five types, including user safety, data integrity, regulatory compliance, patient therapeutic effectiveness, and user convenience. We proposed usability evaluation guidelines that apply the newly derived heuristic-based Software as a Medical Device (SaMD) evaluation factors to the risk management process. In the discussion, we also have proposed the potential applications of the research findings and directions for future research. We have emphasized the importance of the judicious application of AI technology in the medical field and the evaluation of usability evaluation and offered valuable guidelines for various stakeholders, including medical device manufacturers, healthcare professionals, and regulatory authorities.

Key words: SaMD, Heuristic method, International standard, Usability evaluation

I. 서 론

*Corresponding Author : Myung Jin Chung, MD.

Medical AI Research Center, Research Institute for Future Medicine, Samsung Medical Center, Seoul, Republic of Korea
Tel: +82-2-3410-2519

E-mail: mj1.chung@samsung.com

This research was supported by a grant of the Information and Communications Promotion Fund (ICT promotion fund) through the National IT Industry Promotion Agency (NIPA-H1204-23-1001), funded by the Ministry of Science and ICT (MSIT), and Future Medicine 20*30 Project of the Samsung Medical Center (SMC1210791), Republic of Korea.

식품의약품안전처에서는(이하 “식약처”) 의료기기 제조소에서 의료기기 설계에 적용할 수 있는 사용적합성 평가 가이드라인을 발간한다[1]. 기존 의료기기는 Hardware(이하 “H/W” 기반의 기기들이 주류를 이루고 있으며, 그에 따라 H/W 의료기기에 대한 사용적합성 평가 가이드라인이 제시되고 있다. 하지만 최근에는 AI 기반의 독립형 소프트웨어들이 Software as a Medical Device(이하 “SaMD”) 또는 소프트웨어 의료기

기의 범주로 분류되고 있다. 이에 따라 SaMD에 대한 사용 적합성 평가 가이드라인의 필요성은 대두되고 있다. 그중에서도 안전성 및 유효성을 보장할 수 있는 휴리스틱 원칙 기반의 사용적합성 평가 가이드라인은 환자의 안전과 의료 제품의 품질을 도모하기 위하여 필수적으로 고려되어야 한다.

본 연구에서는 새롭게 의료기기로 분류되는 소프트웨어 의료기기(SaMD)에 대한 휴리스틱 원칙 기반 사용적합성 평가 가이드라인의 개발을 위해 휴리스틱 원칙 기반의 의료용 H/W와 일반 비의료기기 Software(이하 “S/W”)의 평가 영역을 조사하고 선행 사례를 참고하여 참조표준을 설정하였다. 그리고 기존의 의료용 H/W와 일반 비의료기기 S/W에 대한 사용적합성 평가 사례를 분석하여 휴리스틱 원칙에 부합하는 평가 영역과 항목들을 조사 및 분석하여 Gap analysis를 실시하였다. 비교분석을 통해 정리된 평가 영역 및 평가 항목에 대한 심각도 분석을 수행하였으며, 심각도가 높게 평가되는 항목들을 필수항목으로 포함하는 사용성 평가 요소들을 제시하여 휴리스틱 원칙을 적용한 위험관리 프로세스를 제시하였다. 이 연구는 SaMD를 위한 사용적합성 평가 가이드라인을 수립하였다.

II. 연구 방법

1. 휴리스틱 기반 의료용 H/W와 일반 비의료기기 S/W에 대한 Gap analysis

본 연구의 목적은 휴리스틱 기반 의료용 H/W와 일반 비의료기기 S/W 간의 차이를 이해하고 두 개의 간극을 확인한다. 휴리스틱 원칙을 적용한 SaMD의 평가요소를 추출하기 위하여 휴리스틱 기반 의료용 H/W와 일반 비의료기기 S/W를 연구대상으로 하여, 문헌검토를 통해 Gap analysis를 수행하였다. SaMD의 특성이 반영된 평가요소를 추출하기 위하여 키워드로는 ‘Heuristics’, ‘Artificial Intelligence Medical devices’, ‘Usability’ 선택하여 의료용, H/W 관련 논문 13건, 일반 비의료기기 S/W 관련 논문 16건을 조사하여 정리하였다[2-30]. 이 두 분야의 평가 영역 및 항목을 수집, 분석 그리고 나열하여 정리하였다. 최종적으로 닐슨의 휴리스틱 원칙을 철저하게 지킨 의료용 H/W문헌 4건과 일반 비의료기기 S/W문헌 5건을 선정하여 평가요소에 대한 공통점과 차이점(표 1)와 같이 정리 하였다.

2. 휴리스틱 원칙 기반 SaMD의 평가요소 분석

닐슨의 휴리스틱 원칙을 기반으로 하여 SaMD의 평가요소를 추출 및 분석하기 위하여, 휴리스틱 원칙 10가지(Visibility of system status(A), Match between system and the realworld(B), Usercontrol and freedom(C), Consistency and standards(D), Error prevention(E), Recognition rather

than recall(F), Flexibility and efficiency of use(G), Aesthetic and minimalist design(H), Help users recognize(I), diagnose, and recover from errors(J), Help and documentation(K))를 나열하고, 선정된 논문 9건에서 정리한 내용을 각각 대표하는 문장으로 재정리하여 통합정리하였다. 그리고 이를 의료용 H/W와 일반 비의료기기 S/W의 평가영역과 평가항목을 공통 평가영역과 공통 평가항목으로 재분류하였다. 더불어, 임상 분야에서 사용되는 평가항목을 식별하는 포괄성을 확보하기 위해, 의료용 H/W와 일반 비의료기기 S/W의 매칭되는 평가항목끼리 재정립하여 SaMD의 적합한 평가 영역과 항목을 확인하였다.

3. 휴리스틱 기반 SaMD 평가항목의 심각도 평가

(1) 심각도 등급 설정 및 잠재적 위험도 분석과 정량화
평가요소의 적합한 근거와 당위성을 확보하기 위하여 국제표준화기구(International Organization for Standardization, ISO), 국제전기기술위원회(International Electrotechnical Commission, IEC), 한국산업표준규격(Korea Industrial standard, KS), 식품의약품안전처 가이드라인을 참조근거로 활용하였다[31]. 심각도의 의미는 특정 의료기기에 대한 위험 허용 기준의 조정이 필요한 위험 허용 기준의 적절성이며, 심각도의 등급을 정량화하기 위하여 정성적 가능성 수준, 반 정량적 가능성 수준, 5가지 정성적 심각성 수준을 수치화하여 반 정량적 5×5 위험 매트릭스로 허용할 수 없는 위험, 추가 위험 통제조사, 중요하지 않거나 무시할 수 있는 위험수준을 정량화하였다.

(2) 평가 영역 및 항목에 대한 위험분석 및 위험평가

위험평가를 수행하기 위하여, 휴리스틱 원칙을 근거한 평가 사례 논문 9건으로부터 정리된 평가 영역과 평가 항목에 대하여 의료 인공지능 연구원 및 SaMD 기반 인허가 연구원을 포함한 5명으로부터 설문지를 통한 위험평가를 실시하였다. 실시된 위험평가로부터 얻어진 정량화 된 반 정량적 가능성 레벨과 정성적 심각성 수준을 각각 1점부터 5점까지 평가하였으며, 두 점수를 곱하여 심각도 점수를 산출하였다. 산출된 심각도 점수는 평가점수를 합산하여 평가자 수로 나눈 평균점수를 반영하였다. 그 범위는 최저 1점부터 최고 25점으로 그 점수가 25점에 가까울수록 심각도 점수의 크며, 위험수준이 높다라는 것을 의미한다. 이를 통해 SaMD의 필요한 요소를 반영하고, 각 평가항목은 해당 분야의 표준적인 테스트 및 평가방법을 사용하여 평가 할 수 있도록 명확한 근거를 마련하였다.

4. 휴리스틱 기반 SaMD 사용적합성 평가 가이드라인 개발

본 연구의 목적은 SaMD의 제조업체, 개발자, 규제당국자, 사용자 등에게 평가 기준을 제공하여 사용자 안전성, 데이터 무결성, 규제준수, 환자 치료 효과, 사용자 편의성에 대한 사용적합성 평가가이드라인을 개발하는 것에 대한 목표로 한

다. 이러한 가이드라인은 위험관리 프로세스는 전반적으로 동일하다는 점에서 휴리스틱 원칙을 기반한 새로운 평가영역 및 항목에 대한 Gap analysis이 적용된 새로운 위험관리 프로세스를 제시함에 있다.

통된 30개의 평가영역과 44개의 평가항목을 추출하였다. 분석결과 이러한 겹분석을 통하여 각 평가영역에 대한 주요한 공통점 및 차이점을 분석한 것을 (표 1)에 정리하였다.

III. 연구 결과

1. 의료용 H/W와 일반 비의료기기 S/W의 Gap analysis 결과

10가지의 휴리스틱 원칙을 기반으로 의료용 H/W 기반 문헌 4건과 일반 비의료기기 S/W 기반 문헌 5건으로부터 공

2. 휴리스틱 기반 SaMD의 평가요소 분석결과

휴리스틱 원칙 기반하여 의료용 H/W와 일반 비의료기기 S/W 문헌에서 제시한 평가영역 및 평가항목을 정리하고, 각 세부항목 별로 동일한 항목별로 정리하여 공통 평가 영역 및 항목과 통합 평가 영역 및 항목으로 재정리하여 (표 2)와 같이 정리하였다. 공통 평가요소는 공통된 요소를 중복제거

표 1. 의료용 H/W와 일반 비의료기기 S/W의 평가의 공통영역 분석결과

Table 1. Commonalities and differences of the evaluation of medical H/W and general non-medical device S/W

휴리스틱 원칙	평가요소		
	의료용 H/W	일반 비의료기기 S/W	공통점 및 차이점
1 Error prevention (A)	에러예방에 대한 제품의 표준화, 아이콘, 평가기준, 여러 가지의 척도, 통제, 비정상적인 상황	데이터셋(기본값 제공, 샘플 포함)에 대한 오류 보완, 분석도구가 오류 예방에 도움이 되는지에 대한 여부, 피드백	에러예방은 범용적인 사용성 원리를 강조하며, 의료용 H/W 및 비의료기기 S/W 평가는 각각 의료 분야와 데이터 중심의 특화된 평가 내용을 다룬다. 전문성, 평가 도구, 평가 내용에서 차이가 있으며, 공통적으로 에러 예방과 사용자 경험 개선을 강조한다.
2 Recognition rather than recall (B)	사용자가 쉬운 사용, 특정 아이콘이 직관적, 작업수행의 여러 가지의 측면, 사용자 경험 고려, 다양한 평가지표	문제해결 프로세스 단순화, 알기 쉬운 용어, 직관적인 사용방법	사용자에게 정보를 기억하지 않고 인지할 수 있도록 하는데 중점을 두며, 의료용 H/W 및 비의료기기 S/W 평가에서는 각각 의료 분야와 데이터 중심의 평가 내용에 특화되어 있습니다. 공통으로는 사용자 부담 감소와 직관적 디자인을 강조하고 있으나, 전문성 요구와 평가 도구 및 방법에서 차이난다.
3 Consistency and standards (C)	일관적인 시스템 상태 및 설정(되돌리기 포함), 그래픽 표시, 지표, 핵심내용, 피드백	일관적인 사용성의 원칙(디자인, typographic, 그래픽디자인 포함), 설계, 표준, 쉬운 사용, 학습가능성	의료용 H/W는 특히 의료 분야에 특화된 평가 기준을 갖추며, 일반 비의료기기 S/W는 디자인과 학습가능성을 강조하여 사용자 경험의 일관성을 추구한다.
4 Visibility of system status (D)	현재상태에 대한 피드백 정보제공, 복구방법	현재상태에 대한 다양한 인터페이스와 척도, 경고, 표준화, 문서화, 사용자에게 명확한 전달, 표준화	의료용 H/W는 안전성 강화를 위해 주로 피드백과 복구 방법에 중점을 두며, 비의료기기 S/W는 다양한 인터페이스와 표준화를 강조하여 명확한 정보 전달과 일관된 사용자 경험을 추구한다.
5 User control and freedom (Navigation) (E)	복구가능성에 대한 사용자 모델 오류, 사용자 행위, 시스템 선택 및 변경, 가능성 제시	복구가능성에 대한 안내, 통제, 상호작용	의료용 H/W는 주로 의료 장비의 안전성 강화를 위해 사용자 모델 오류 및 시스템 변경에 중점을 두고, 비의료기기 S/W는 사용자의 실수 복구를 강조하여 안내, 통제, 상호작용을 개선한다.
6 Flexibility and efficiency of use (F)	효율적인 사용자 경험, 작업수행, 긴급상황, 시스템 재시작, 수치화	효율적인 사용, 도구, 척도, 템플릿, 도움, 사용방법, 사용방식, 난이도, 사용 방법	의료용 H/W는 의료 환경에서의 특수한 상황을 고려하여 효율성을 개선하고, 비의료기기 S/W는 다양한 측면에서 사용자 경험을 향상시키기 위해 노력한다.
7 Match between system and the real world (G)	익숙함으로부터 사용자의 오해, 평가요소의 판단기준	익숙함으로부터 만족과, 알기 쉬운 용어사용	의료용 H/W는 익숙하지 않은 상황에서의 사용자 오해를 최소화하고 익숙함을 제공하고, 비의료기기 S/W는 사용자가 제품을 자연스럽게 이해하고 조작할 수 있도록 노력한다.
8 Aesthetic and minimalist design (H)	핵심정보의 인식, 기록, 원칙, 내용	핵심정보의 제공 및 전달, 작성(사용설명서 포함)	의료용 H/W는 주로 의료 정보 전달과 기록에 중점을 두고 디자인을 간결하게 유지하며, 비의료기기 S/W는 사용자에게 명확한 정보를 제공하고 간결한 사용설명서를 작성하는 데 중점을 둔다.
9 Help users recognize, diagnose, and recover from errors (Recovery) (I)	사용자와 인터페이스의 상호작용, 시스템의 효과적인 활용, 명확한 에러 문구와 알람 제공	명확한 에러문구 및 정보 그리고 사용시나리오 작성	의료용 H/W는 주로 사용자와 인터페이스 상호작용, 효과적인 시스템 활용, 명확한 에러문구 및 알람 제공을 평가하고, 비의료기기 S/W는 명확한 에러문구와 사용시나리오 작성에 중점을 두어 사용자가 에러를 인식하고 복구할 수 있도록 노력한다.
10 Help and Document (J)	다양한 평가방법	평가방법, 목적 및 내용의 명확한 기술이 되도록 문서화, 다양한 도움 서비스 제공	의료용 H/W는 다양한 평가 방법을 통해 성능과 안전성을 평가하고 향상시키고, 비의료기기 S/W는 평가 방법 및 문서화를 강조하여 명확한 사용자 지원과 다양한 도움 서비스를 제공한다.

하고, 공통 평가영역은 29건 공통 평가항목은 57건, 통합 평가영역은 69건, 통합 평가항목은 112건으로 정리되어 선정되었다. 전체항목 112건에 대해 5 × 5 위험도 분석 매트릭

스에 따라 심각도 분석을 수행하였다. 이러한 평가영역과 항목에 SaMD 사용적합성 평가할 수 있는 근거를 마련하였다.

표 2. 문헌검토를 통해 도출된 휴리스틱 기반 평가 영역 및 항목

Table 2. Heuristic-based evaluation areas and items derived from literature review

휴리스틱 원칙	평가영역	평가항목
Error prevention (A)	에러예방을 할 수 있도록 표준화 할 수 있는가	-사용자는 비표준화 된 제품을 선택하지 않았는가 -많은 샘플을 처리할 때 동일한 출력 파일을 잘못 설정된다면, 분석이 다시 실행되어지는가
	사용자는 아이콘의 에러예방을 찾을 수 있는가	-사용자는 특정 치료 아이콘을 찾을 수 있는가
	에러에 대한 평가기준은 명확하거나, 모호하거나 일반적으로는 않은 메시지 등의 질문으로 구성되어 있는가	-에러예방에 대한 평가질문을 포함하였는가
	에러에 대한 평가기준은 다양한 유형의 에러 등의 질문으로 구성되어 있는가	-에러예방에 대한 인터페이스의 평가기준이 있는가
	정상적이지 않은 상황에서 오류예방을 할 수 있도록 작동이 되는가	-비정상작동시에도 에러를 대응할 수 있는가
	에러예방을 할 수 있도록 기본값 및 조건을 설정하여 여러가지의 척도가 있는가	-에러예방에 대한 여러가지(성능, 프로세스, 오류복구율 등)를 평가할 수 있는 기준이 있는가 -에러예방 시, 우선순위를 두는가, 에러 예방 시, 도움되는 조건이나 기본설정을 제공하는가, 사용자들에게 오류예방에 관한 경고를 주어 오류를 방지하는가
	오류예방을 하기 위해 분석에 필요한 유연한 통제나 도구가 있는가	-오류예방에 적절한 인터페이스를 고려하였는가 -분석에 필요한 필수 분석 요소들이 설정이 되었는가
	오류해결에 도움이 되는 피드백이 있는가	-사용자는 경보음이 울리면 특정 버튼을 누를 수 있는가 -화면에 메시지가 뜰 경우, 다른 인터페이스에 방해가 되지 않는가 -에러발생 시, 사용자에게 유용한 정보제공이 되어지는가
	데이터셋에 대한 오류를 보완하여 작동성을 향상시킬 수 있는가	-테스트를 위한 데이터 세트를 제공되면 소프트웨어의 운용성을 향상되는가
	Recognition rather than recall (B)	사용자에게 이해하기 쉬운 정보를 제공하는가
특정 아이콘은 직관적인가		-사용자들이 쉽게 아이콘을 바라보는가
작업수행의 프로세스를 간결하게 만들 수 있는가		-사용자의 작업수행 능력을 향상시키기 위하여 여러가지의 고려사항이 적용되었는가 -사용자에게 설치프로세스를 간결하게 제공할 수 있는 방법이 있는가
사용성 경험을 개선사항을 고려하였는가		-사용자가 고려하는 평가요소로 작업 단계의 프로세스와 명확한 피드백이 있는가 -사용자의 문제해결과정을 올바른 조치로 이어질 수 있도록 직관적으로 확인이 가능한가
다양한 평가 지표를 수집하기 위한 측정 도구나 방법을 사용하는가		-사용자는 다양한 평가방법이 있는가
Consistency and standards (C)	시스템은 일관적으로 기능의 실행해주는가	-사용자가 선택한 버튼이 어떤 시스템에서도 지원이 되는가
	그래픽의 표시는 정확한가	-사용자들을 위한 그래픽 표현이 정확한가 -그래픽 디자인의 원칙이 있는가
	사용자 경험의 제품의 일관적인 원칙이 있는가	-사용자는 제품에 일관적인 기준이 있는가 -일관성에 대한 검증은 대표 설계자와 자신의 설계가 동일한가
	시스템의 확인 설정은 일관하도록 설정이 되어 있는가	-확인설정에는 시스템과 인간이 동일한 설정으로 적용하는가
	개인정보가 표시 될 때, 자동을 시스템을 되돌리는 기능이 일관적으로 실행되는가	-특정 단축키를 누르면 즉시 기본화면으로 들어가는가
	typographic은 일관적으로 설계되어있는가	-typographic의 원칙이 있는가
	명확한 경고문구와 즉각적인 피드백은 일관성 있는가	-Rate 또는 Volume에 대해 기준점이 있는가 -사용자의 올바른 조치를 잘한다면 혼란스럽지 않은가
	사용성의 원칙은 일관적인가	-사용성전문가는 휴리스틱 원칙을 따르는지 판단하는 방법이 있는가

표 2. (continued)
Table 2. (continued)

휴리스틱 원칙	평가영역	평가항목
Consistency and standards (C)	디자인의 원칙은 일관적으로 설계되어있는가	-디자인 원칙이 있는가
	인터페이스의 표준은 일관적인가	-표준 검증방법은 인터페이스가 표준을 준수하는지 검사를 하는가
	사용자는 데이터표준화 작업을 하였는가	-선호하는 데이터 형식이 있는가
	누군든지 사용할 수 있도록 일관적인 설치 및 삭제가 할 수 있는가	-소프트웨어는 누구에게나 설치 및 제거가 가능한가
	제품의 학습가능성은 일관적인가	-내부 및 외부로 일관적으로 학습 용이성을 향상시킬 수 있는가 -단일 제품 또는 관련 제품 내에서 일관성을 유지하는가 -산업의 관행을 준수하는가
Visibility of system status (D)	명확한 피드백을 전달하는가	-사용자에게 가이드를 제공하는가 -명확한 전달과 피드백을 제시하여, 사용자와 상호작용을 하는가
	현재상태의 정보를 제공하는가(알람)	-휴리스틱 원칙에 위배되는 시스템의 현재상태. 현재상태에서 수행 가능한 작업 등의 미해결 질문이 있는 경우에 발생하는가 -경고음이 울리면 눈에 확 띄는가
	현재상태의 시스템 정보현황을 전달하는가	-시스템의 현재 상태는 어떻습니까? -소프트웨어의 활동상태가 적합한가
	현재 상태(오류)에 대한 복구방법이 있는가	-사용자는 화면제한 시간을 인지할 수 있는가 -사용자는 복구방법을 실행할 수 있는가
	현재상태를 다양한 척도로 보여주는가(정의, 시스템의 상태, 표준화)	-그리드 정의, 유연성, 레이아웃, 시스템 피드백 등의 원칙이 있는가
User control and freedom (Navigation) (E)	현재상태의 인터페이스를 다양한 척도로 보여주는가(메뉴, 양식, 글꼴 등)	-시각적 구성 스타일, 메뉴, 피드백 등의 원칙이 있는가
	사용자는 현재 상태를 문서화하였는가	-로그파일은 현재상태의 기록이 문서화하는데 도움을 주는가
	복구에 대한 가이드가 있는가	-사용자가 시스템을 복구할 수 있는 참고사항이 있는가 -프롬프트는 방향제시를 잘해주는가
	휴리스틱 기반으로 시스템을 평가하여 복구가능성 설계의 고려사항으로 채택하였는가	-사용자가 설계된 시스템을 평가하기위한 휴리스틱 방법을 사용하였는가
	되돌리기(실행 및 재시작)의 복구가능성이 명확한가	-전원버튼을 활용하지 않을 때, 특정 기능을 사용이 가능한가 -소프트웨어는 실행 취소 및 다시 실행할 수 있는가
Flexibility and efficiency of use (F)	사용자에게 주의해야 신호를 주어야하는가	-시스템은 사용자에게 주의를 주는가
	어떤 사용자라도 소프트웨어를 통제하기 쉬운가	-소프트웨어는 초보자 그리고 숙련자에게 모두에게 쉽게 컨트롤을 할 수 있는 인터페이스를 가지고 있는가
	사용자에게 작업수행에 수월한 기준이 있는가	-효율적인 기준은 어떤 것으로 이루어져있는가 (예를 들면, 단축키, 명령 키, 기본값 등) -사용자가 옵션을 설정하고, 참조, 주석 등을 정의할 수 있는 템플릿을 구성하였는가
	특정모드에서 작업수행이 불가능함을 강조하는가	-사용자는 특정모드에서 수행이 불가능한가
	사용자의 유연성과 효율성을 위하여 긴급상황에서도 통제불가능한 것이 있는가	-유연성과 효율성을 위하여 긴급한 상황에서 사용자의 판단은 허용되는가
Flexibility and efficiency of use (F)	사용자가 기계의 유연성과 효율성을 위하여 시스템 재시작을 해야 하는 문제가 있는가	-시스템 설계는 효율적으로 하였는가
	모델을 정량화 시킬 수 있는가	-모델 간 도메인의 기능은 유연한가 -모델 내 도메인의 함수는 효율적으로 하였는가
	사용하기가 좋은가	-빠르고 효율적으로 클릭과 읽기 작업을 수행할 수 있는가
	다음은 효율적으로 나타내고 있는가	-유용하고 효율적인 척도가 있는가
	효과적으로 도움을 주는 것이 유용한가	-소프트웨어는 사용자에게 유용한가 -소프트웨어는 사용자에게 효과적인가 -소프트웨어는 사용자에게 생산성향상에 도움이 되는가

표 2. (continued)
Table 2. (continued)

휴리스틱 원칙	평가영역	평가항목
Flexibility and efficiency of use (F)	사용하기에 유연하고 쉬운가	-소프트웨어는 사용자에게 유연한가 -소프트웨어는 사용자가 사용하기에 쉬운가 -소프트웨어는 사용자가 사용하기에 간편한가
	누구든지 사용하기에 쉬운가	-사용자는 소프트웨어를 배우는 것이 쉬운가 -사용자는 소프트웨어 사용에 빠르게 배웠는가 -사용자는 소프트웨어의 사용법을 쉽게 기억하는가
	사용자의 사용방식이 효율적인가	-사용자에게 키보드 단축키나, 터치 제스처와 같은 것을 제공하는가 -개별사용자를 위하여 콘텐츠와 기능을 맞추어 제공하는가
	효율적인 도구를 사용하고 있는가	-사용자가 제품의 작동방식을 선택할 수 있도록 설정을 허용하였는가 -도구는 언제부터 사용하는 것이 효율적이고 유용한가
	사용자가 오해할만한 요소가 있는가	-화면의 숫자 키패드는 다른 입력창과 비슷한가
Match between system and the real world (G)	사용자가 알기 쉬운 용어나 시스템이 익숙한가	-판단 기준은 다음과 같다: 사용자 모델이 시스템 이미지와 일치합니까; 시스템이 제공하는 조치가 사용자가 수행하는 조치와 일치하는지 여부; 시스템의 개체가 작업의 개체와 일치합니까? -사용자가 이해하기 쉬운용어를 사용하는가, 사용자가 용어/개념을 이해할 때, 친숙한 단어와 일치한다고 생각하는가, 사용자 조사는 익숙한 용어 및 중요한 정의에 대한 개념을 도움받을 수 있는가> -시스템은 사용자에게 만족을 주는가(사용환경)
	내가 원하는 방식으로 사용자를 만족시켜주는가	-시스템은 사용자에게 만족을 주는가(만족성) -시스템은 사용자에게 만족을 주는가(성능)
Aesthetic and minimalist design (H)	핵심정보를 이해하기 쉽게 하여 작성하였는가	-환자데이터를 식별하고, 데이터입력을 핵심정보를 기록할 수 있는가 -사용자는 더 나은 경험을 얻을 수 있도록 핵심을 짚어주는가 -정보를 핵심적인 척도에 따라서 구분이 가능한가 -이해하기 쉽게 얼마나 내용을 명확하게 작성하였는가
	사용자들은 핵심정보를 기록할 수 있는가	-사용자는 환자데이터의 핵심정보를 기록 할 수 있는가
	핵심적인 원칙이 있는가	-정보를 핵심적인 척도에 따라서 구분이 가능한가
	사용설명서는 명확하게 작성되었는가	-사용설명서에는 사용자가 쉽게 이해할 수 있도록 명확하게 기술되었는가 -사용하기 편하게 사용설명서에는 명확하게 작성되어있는가
	중요한 정보를 전달해주는가	-문제에 대한 중요한 정보를 전달해주는가 -UI는 내용과 시각적 디자인에 핵심요소를 집중시켰는가
Help users recognize, diagnose, and recover from errors (Recovery) (I)	디자인의 필수요소는 핵심정보를 제공하는가	-사용자들에게 불필요한 정보를 전달하고 있는가 -주요 목표를 지원하기 위해 우선순위를 지정하였는가
	사용자 경험을 향상시키기 위하여, 명확한 에러문구, 핵심정보에 대한 내용을 다루었는가	-명확한 에러문구를 고려하기 위하여 피드백 수준, 명확성 등의 특성이 반영되었는가 -정형화된 사용성 검증은 명확한 절차에 따라 휴리스틱 평가에 사용 되는가
	다양한 요소(사용자, 개발자, 인적요소)를 논의하며, 시나리오를 통해 단계적으로 나아가고 있는가	-다양한 요소(사용자, 개발자, 인적요소)를 논의하며, 시나리오를 통해 단계적으로 나아가고 있는가
	표준 언어 사용, 사용자 정의 가능한 언어, 그리고 사용자 중심 커뮤니케이션은 사용자가 시스템을 더 효과적으로 이해하고 활용할 수 있는가	-명확한 에러문구를 평가하는 방법으로는 전문용어, 정의등이 있는가 -View personality에 명확한 정보를 얻을 수 있는가
	명확한 에러문구가 있는가	-실수를 방지하려면 명확한 에러예방 할 수 있도록 사용 시나리오를 커버하고 지나치게 작성을 하지 않았는가
에러 발생시, 작업환경에서 명확한 피드백을 받을 수 있는가	-인터페이스의 hard press는 작업환경에 명확한 피드백을 받을 수 있는가	

표 2. (continued)
Table 2. (continued)

휴리스틱 원칙	평가영역	평가항목
Help and Document (J)	제품(hw, sw)의 평가방법으로써 어떤 측면을 고려하여 평가하는가	-도움 및 서비스의 평가방법은 어떠한 것들이 포함되어 있는가(민감한 도움, 작업 중심의 방식 등) -어떤 시스템이나 소프트웨어의 사용자 지원 및 도움말 시스템이 어떻게 평가하고 있는가
	안내서에는 도움이 되는 내용이 잘 기술되어 있는가	-라이선스 문서는 소프트웨어 사용 및 배포 지침을 명시하고 있어 도움을 줄 수 있는가
	가이드에는 도움이 되는 내용이 잘 기술되어 있는가	-코드의 새로운 버전이 배포될 때, 변경사항은 명확하게 하여, 이전 버전으로부터 식별된 영향이 문서화 되어 있는가
	코드설명은 제3자가 도움을 받을 수 있도록 명확하게 작성되어 있는가	-사용자들은 업데이트 단계 현황과 같은 워크플로우에 대해 정보(날짜와 시간 포함)를 받음으로서 다음 실험을 계획하고 평균 분석 기간이 자신의 요구사항에 부합하는지 여부를 결정할 수 있는가
	워크플로우에 대한 도움이 되는 피드백이 문서화로 작성되어 있는가	-분석단계의 출력이 간결하게 유지되거나, 고급 사용자의 인텍스 생성 알고리즘의 출력과 같은 정보가 자세히 제공되는가
	도움서비스에 대한 내용의 목적이 표시되는가	-가이드에는 기본 설치 및 사용 지침이 포함되어 있어 도움을 줄 수 있는가
	사용자가 필요로하는 도움말을 제공하는가	-도움말 문서가 검색하기 쉬운가 혹은 가능한 경우 사용자가 필요로 할 때 바로 제공할 수 있는가
	사용자에게 다양한 도움서비스를 제공하는가	-기존 오류메시지를 시각적(굵은 글씨, 빨간색 스트 등)으로 표현하고 있는가
		-사용자가 이해할 수 있는 언어로 작성하고 있는가(전문용어 사용하지 않기)
	표준작업정보가 명확하게 작성되어 있는가	-사용자에게 즉시 오류를 해결할 수 있는 바로가기와 같은 해결책을 제공하고 있는가
		-기능 검사는 일반적인 작업을 수행 및 평가하기 위해 단계로 나뉜 지식이나 경험을 나열하고 있는가

434

3. 휴리스틱 기반 SaMD 평가항목의 심각도 평가결과

(1) 심각도 등급 설정과 잠재적 위해도 분석과 정량화 결과 새롭게 추출한 평가영역과 평가항목이 국제규격 기반 KS P ISO/TR 24971:2020에서 제시한 3가지 위험 매트릭스를 통하여 확인하였다. 정성적 가능성 수준은 빈번, 가능성 높음, 간헐적, 가능성 낮음, 가능성 희박의 순으로 설정하여 5가지 반 정량적 가능성 수준은 각 1/100,000 or less, 1/100,000 ~1/10,000, 1/10,000~1/1,000, 1/100 or more으로 (표 3), (표 4) 그리고 (표 5)와 같다.

정성적 심각성 수준은 무시가능, 경미, 심각/중대, 위독, 파국적/지명적으로 설정하여 잠재적 위해도 평가를 위해 반 정량적 수준과 정성적 심각성 수준을 반 정량적 5 × 5 위험 매트릭스로 구현하여, 위험도에 따라 허용할 수 없는 위

표 3. 정성적 가능성 수준
Table 3. The level of qualitative likelihood

공통용어	가능한 설명
빈번	자주, 빈번, 항상
가능성 높음	발생 가능성 높음
간헐적	발생할 수 있지만 빈번하지 않음
가능성 낮음	발생 가능성 낮음
가능성 희박	발생 가능성 희박함

표 4. 반 정량적 가능성 수준의 예시

Table 4. Example of five semi-quantitative probability levels

Common terms	Examples of probability range
Frequent	$\geq 10^{-3}$
Probable	$< 10^{-3}$ and $\geq 10^{-4}$
Occasional	$< 10^{-4}$ and $\geq 10^{-5}$
Remote	$< 10^{-5}$ and $\geq 10^{-6}$
Improbable	$< 10^{-6}$

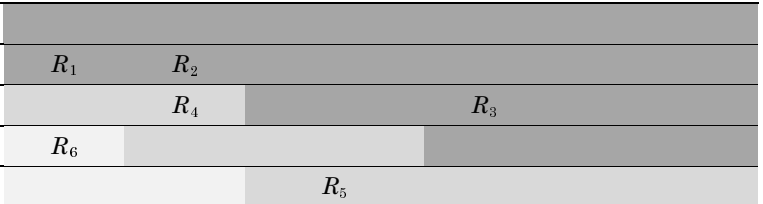
험, 추가 위험 통제 조사, 중요하지 않거나 무시할 수 있는 위험으로 분류하여 (표 6)와 같이 정리하였다. (표 6)에서 심각도와 가능성수준을 고려하여 1~25점중 9점 이상이거나, 반 정량적 가능성 수준이 4점 이상인 항목들을 허용할 수 없는 위험 평가 영역으로 최종선정 하였다.

반 정량적 검사는 측정되는 양의 임상적으로 유용한 근사치를 제공하여, 값은 통상적으로 순위척도(ordinal scale)를 기반으로 규정되거나, 검량한계(quantity limit)로 보고되며, 숫자(규정 값) 범위 사이 또는 규정 값보다 크거나 작음, 적정 농도 또는 계단식 희석, 또는 상대적(+3, +2, +1 또는 미량)으로 표현된다. 정성적 검사는 분석 물질의 존재 유무를



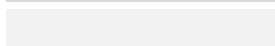
표 5. 5가지 정성적 심각성 수준의 예시
Table 5. Example of five qualitative severity levels

Common terms	Possible description
Catastrophic/Fatal	Results in death
Critical	Results in permanent impairment or irreversible injury
Serious / Major	Results in injury or impairment requiring medical or surgical intervention
Minor	Results in temporary injury or impairment not requiring medical or surgical intervention
Negligible	Results in inconvenience or temporary discomfort

표 6. 반 정량적 5 × 5 위험 매트릭스의 예시
Table 6. Example of a semi-quantitative 5 × 5 risk matrix

Semi-quantitative probability levels	Qualitative severity levels				
	Negligible (1)	Minor (2)	Serious/Major (3)	Critical (4)	Catastrophic/Fatal (5)
Frequent(5)					
Probable(4)					
Occasional(3)					
Remote(2)					
Improbable(1)					

*Key

	Unacceptable risk
	investigate further risk control
	insignificant or negligible risk

결정하고 결과는 양성, 음성 또는 불확실 등으로 보고된다. 판정 기준치(cut-off value) 값과 관련 데이터베이스는 양성 또는 음성 결과를 정의하였다[33].

(2) 평가영역 및 평가항목에 대한 위험분석 및 위험평가 결과 위험평가를 수행하기 위하여, 10가지의 휴리스틱 원칙에

대해 식별된 위험은 일반적으로 휴리스틱 원칙에 정의된 내용으로 설정하고, 이에 대한 SaMD 평가에 적용 될 위험 우선순위를 설정하여 10가지 원칙에 대해 모두 중요한 위험으로 인식하여 위험도평가에 포함하여 (표 7)와 같이 결과를 정리하였다.

휴리스틱 원칙을 근거한 평가사례 논문 9건으로부터 정리

표 7. 휴리스틱 원칙을 적용한 잠재적 위험사례 식별
Table 7. Identified potential risk cases with heuristic principles

N	Evaluation (subject)	식별된 위험	위험 우선순위 (낮음->높음)
1	A	잘못된 입력에 따른 잠재적인 오류	중요한 위험
2	B	사용자의 기억력에 의존하는 기능	중요한 위험
3	C	일관성 부족 및 표준 미준수	중요한 위험
4	D	시스템 상태 부적절한 표시	중요한 위험
5	E	사용자가 제어를 잃는 경우	중요한 위험
6	F	사용자 효율성 부족 및 작업의 유연성 부족	중요한 위험
7	G	사용자가 이해하기 어려운 용어 또는 상황	중요한 위험
8	H	사용자 인터페이스가 복잡하거나 혼란스러운 경우	중요한 위험
9	I	사용자가 오류를 파악하고 복구하는 데 도움이 필요한 경우	중요한 위험
10	J	도움말 및 문서가 부족한 경우	중요한 위험

된 평가영역 96건의 평가항목 132건에 대하여 의료 인공지능 연구원 및 SaMD 기반 인허가 연구원을 포함한 5명으로부터 설문을 통한 위험평가를 실시하여 위험점수의 평균값, 평균심각도를 (표 8)와 같이 결과를 정리하였다.

이는 리커르트 5점 척도를 사용하여 각 잠재적 위험 시나리오에 대해 1~5점까지의 숫자를 할당하여 정량화에 목적이 있다. 1점은 매우 낮은 위험을 나타내고 5점은 매우 높은 위험을 의미한다. 이러한 위험평가를 통해 얻은 점수로부터 심각도 평가의 평균과 환산점수를 산출하였다. 산출된

심각도 평균점수는 33점 이상인 평가요소로부터 평가항목을 확인하고, 심각도 점수에 따른 표 6에 따라 평가영역은 Unacceptable risk(85개), investigate further risk control (29개), insignificant or negligible risk(20개)의 영역으로 분류되었다. 이 중에서 추가 위험 통제조사 이상에 드는 평가영역만을 선정하여 최종 평가영역으로 설정하였다. 최종 설정된 평가영역을 휴리스틱 원칙에 따라 5가지 유형으로 분류하여 휴리스틱 원칙별로 정리하였다.

본 연구에서는 위험사례와 관련된 시나리오를 식별하고

표 8. 휴리스틱 원칙을 적용한 영역 내 평가항목의 위험점수의 평균값

Table 8. Average risk scores of evaluation items in the area to which heuristic principle is applied

N	Evaluation (subject)	위험점수의 평균값		위험수준 (1~25점)
		식별된 위험	평가영역	
1	Error prevention (A)	잘못된 입력에 따른 잠재적인 오류	분석도구에 오류예방이 있을만한 것이 있는지 확인하였는가	24.4
			⋮	⋮
			데이터셋에 대한 오류를 보완하여 작동성을 향상시킬 수 있는가	10.6
2	Recognition rather than recall (B)	사용자의 기억력에 의존하는 기능	사용자들이 알기쉬운 정보(도움말, 정보, 가이드)를 제공하는가	8
			⋮	⋮
			문제해결 프로세스가 단계별로 직관적인가	9.6
3	Consistency and standards (C)	일관성 부족 및 표준 미준수	인터페이스의 표준은 일관적인가	14.4
			⋮	⋮
			프로젝트의 설계는 일관적인가	18.4
4	Visibility of system status (D)	시스템 상태 부적절한 표시	현재상태의 경고를 제공하는가	13
			⋮	⋮
			현재상태의 인터페이스를 다양한 척도로 보여주는가(메뉴, 양식, 글꼴 등)	6.8
5	User control and freedom (Navigation) (E)	사용자가 제어를 잃는 경우	현재 상태에 대한 안내가 잘 되어 있는가	10.2
			어떤 사용자라도 소프트웨어를 통제하기 쉬운가	9.6
			상호작용하는 기능이 복구가능성이 명확한가	11.4
6	Flexibility and efficiency of use (F)	사용자 효율성 부족 및 작업의 유연성 부족	사용하기가 좋은가	12
			⋮	⋮
			효과적으로 도움을 주는 것이 유용한가	16.6
7	Match between system and the real world (G)	사용자가 이해하기 어려운 용어 또는 상황	화면에 표시된 인터페이스는 사용자가 오해할 수 있는가	15.4
			내가 원하는 방식으로 작동해주는 것은 만족을 주는가	4.6
			사용자에게 친숙한 용어나 개념을 사용하는가	14.4
8	Aesthetic and minimalist design (H)	사용자 인터페이스가 복잡하거나 혼란스러운 경우	중요한 정보를 전달해주는가	2.4
			⋮	⋮
			이해하기 쉽게 얼마나 내용을 명확하게 작성하였는가	23
9	Help users recognize, diagnose, and recover from errors(Recovery) (I)	사용자가 오류를 파악하고 복구하는 데 도움이 필요한 경우	명확한 에러문구를 간단하게작성하였는가	13.6
			⋮	⋮
			표준작업정보가 명확하게 작성되어있는가	15.2
10	Help and Document (J)	도움말 및 문서가 부족한 경우	도움서비스에 대한 내용의 목적이 표시되는가	12.8
			⋮	⋮
			어떤 시스템이나 소프트웨어의 사용자 지원 및 도움말 시스템이 어떻게 평가하고 있는가	10.6

각 시나리오의 잠재적 위해도도를 평가하였다. 위험도를 정량화하기 위해 반정량적 레벨은 가능성 희박을 1점, 빈번을 5점으로 평가하였고, 정성적 심각성 수준을 무시 가능부터 파국적/치명적 수준을 1~5점으로 평가하였으며, 반 정량적 가능성 수준의 점수와 정성적 심각성 수준에 점수를 곱하여 위험도를 정량화 하였으며, 이는 (표 6)와 같이 25개의 구간으로 분류하여 산정위험을 $R_1 \sim R_{25}$ 설정하였다[32-33]. ISO 24971:2020와 같이 5×5 matrix를 활용하여 위험평가를 진행하였다[34]. 이는 각 평가항목의 잠재적인 위험을 정량화하여, 사용자와 환자의 안전을 보장하기 위한 위험관리 프로세스를 개선하기 위한 근거로 활용된다[35].

SaMD의 필요한 평가요소는 종합적인 평가 영역 및 항목의 평가를 통하여 필요한 요소를 반영해야 하고, 각 평가항목은 해당 분야의 표준적인 테스트 및 평가방법을 사용하여 평가 할 수 있다. SaMD는 사용자의 안전을 어떻게 보장하고, 명확하게 예후예측 진단 및 보조하여 사용자의 안전을 확보하고, 데이터를 기반으로 의사결정을 내리므로 데이터의 무결성은 핵심적인 평가항목이며, 국내외 규격을 준수하여 제품의 규제측면을 평가하고, 이러한 평가항목을 통하여 환자의 치료효과를 어떻게 개선하는지 확인하고, 이를 효과적으로 활용할 수 있는지를 평가하기 위한 평가영역으로는 사용자안전성, 데이터 무결성, 규제준수, 환자 치료효과, 사용자 편의성으로 선정하였다.

선정된 SaMD의 5가지 평가요소와 휴리스틱 원칙 기반의 종합적인 평가 영역 및 항목을 적용한 영역 분류표를 (표 10)와 같이 구성하고, 총 50가지의 휴리스틱 원칙을 기반한 SaMD의 위험을 고려사항을 제시하였다.

4. 휴리스틱 기반 SaMD 기반의 사용적합성 평가 가이드라인

휴리스틱 기반 SaMD의 사용적합성 평가 가이드라인은 환자 안전과 치료 효과 향상을 목표로 한다. 이러한 가이드는 SaMD의 제조업체, 개발자, 규제당국자, 사용자 등에게

평가 기준을 제공하여 사용성(a), 안전성(b), 데이터 무결성(c), 규제 준수(d), 환자 치료 효과(e), 사용자 편의성(f) 등을 종합적으로 평가하는 도구로 사용하였다. 또한 가이드라인의 개발 프로세스는 평가대상의 특성을 고려하여, 선정된 평가 영역 및 항목에 대하여 국제규격 기반의 기준을 설정하고, 이를 분석하여 개선사항 수립에 적용될 수 있다. 평가요소를 정량화하여 사용적합성 평가를 수행한다. 평가결과와 피드백을 고려하여 가이드라인을 개선 및 보완 할 수 있도록 개선계획을 수립 시 도움을 받을 수 있다.

ISO 14971:2019의 요구사항에 따라 위험관리의 세부적인 분석표에 휴리스틱 기반 평가요소를 적용한 것은 (그림 1)과 같이 총 9가지 프로세스 중에서 위험분석 프로세스에 적용하였다.

개발 단계에 따른 위험 관리 활동은 (표 12)와 같으며, 이는 사용자와 환자의 안전을 보장하기 위한 품질관리 프로세스를 개선하기 위한 근거로 활용된다(그림 1).

본 연구에서 개발된 휴리스틱 원칙 기반의 사용적합성 평가 요소들은 개발 단계에서 요구되는 위험 관리 활동 프로세스에 적용하였으며, 그 프로세스는 (그림 1)과 같다. 위험 관리 프로세스는 일반적으로 위험분석(Risk analysis), 위험평가(Risk evaluation), 위험통제(Risk control), 전체 잔여위험평가(Evaluation of overall residual risk). 위험관리 검토(Risk management review), 생산 및 생산 후 활동(Production and post-production activities)으로 구성되는데, 그 중 위험분석 단계에서 휴리스틱 기반 SaMD 위험 분석(Heuristic-Based SaMD risk estimation) 절차를 추가하는 것으로 소개하였다.

결론 및 고찰

본 연구에서는 휴리스틱 기반 SaMD의 사용적합성 평가 가이드라인에 대한 결과를 정리하고, 부족한 내용과 추가적

표 9. SaMD 사용적합성 평가의 유형 정의
Table 9. Defined type of SaMD usability assessment

N	공통 용어	가능한 설명
1	사용자 안전성 (1)	위험이 사용자의 안전에 직접적인 영향을 미치는 경우 심각도가 높을 수 있다. 이러한 위험은 인명을 위협하거나 환자 치료에 영향을 미칠 수 있다.
2	데이터 무결성 (2)	의료용 소프트웨어의 데이터 무결성을 위협하는 위험은 심각도가 높을 수 있다. 예를 들어, 환자 기록의 왜곡 또는 손실은 큰 문제가 될 수 있다.
3	규제 준수 (3)	규제의 요구사항을 위반하는 위험은 심각도가 높을 수 있다. 규제 당국에 대한 법적 위반으로 이어져 영향을 미칠 수 있다.
4	환자 치료효과 (4)	의료용 소프트웨어의 성능 및 사용성이 환자 치료 효과에 직접적인 영향을 미치는 경우 해당 위험은 높은 심각도를 가질 수 있다.
5	사용자 편의성 (5)	사용자가 소프트웨어를 효과적으로 사용하지 못하면 진단 및 치료에 영향을 미칠 수 있으므로 사용자 편의성 관련 위험은 고려되어야 한다.

표 10. 휴리스틱 원칙을 적용한 평가유형에 따른 영역 분류표

Table 10. Area classification table according to evaluation type applying heuristic principle

휴리스틱 원칙	사용자 안전성	데이터 무결성	규제 준수	환자 치료효과	사용자 편의성
A	<ol style="list-style-type: none"> 1. 에러에 대한 평가기준은 명확하거나, 모호하거나 일반적이지 않은 메시지 등의 질문으로 구성되어 있는가 2. 에러에 대한 평가기준은 다양한 유형의 에러 등의 질문으로 구성되어 있는가 3. 정상적이지 않은 상황에서 오류예방을 할수있도록작동이되는가 4. 에러예방을할수있도록기본값및조건을설정하여여러가지의척도가있는가 5. 오류해결에도움이되는피드백이 있는가 	<ol style="list-style-type: none"> 1. 에러예방을 할 수 있도록 표준화 할 수 있는가 2. 데이터셋에대한오류를 보완하여작동성을향상시킬수있는가 			<ol style="list-style-type: none"> 1. 사용자는 아이콘의 에러 예방을 찾을 수 있는가 2. 오류예방을하기위해분석에필요한유연한통제나도구가있는가
B	<ol style="list-style-type: none"> 1. 사용자에게 이해하기 쉬운 정보를 제공하는가 2. 특정아이콘은직관적인가 3. 작업수행의프로세스를간결하게만들수있는가 4. 사용성경험을개선사항을고려하였는가 5. 다양한평가지표를수집하기위한측정도구나방법을사용하는가 				
C	<ol style="list-style-type: none"> 1. 명확한 경고문구와 즉각적인 피드백은 일관성 있는가 2. 개인정보가표시될때,자동을시스템을되돌리는기능이일관적으로실행되는가 3. 누군든지사용할수있도록일관적인설치및삭제가할수있는가 				<ol style="list-style-type: none"> 1. 시스템은 일관적으로 기능의 실행해주는가 2. 그래픽의표시는정확한가 3. 사용자경험 제품의 일관적인 원칙이있는가 4. 시스템의확인설정은일관하도록설정되어있는가 5. typographic은일관적으로설계되어있는가 6. 사용성의원칙은일관적인가 7. 디자인의원칙은일관적으로설계되어있는가 8. 인터페이스의표준은일관적인가 9. 사용자는데이터표준화작업을하였는가 10. 제품의학습가능성은일관적인가
D	<ol style="list-style-type: none"> 1. 명확한피드백을전달하는가 	<ol style="list-style-type: none"> 1. 현재 상태(오류)에 대한 복구방법이 있는가 	<ol style="list-style-type: none"> 1. 현재상태의 시스템 정보현황을 전달하는가 	<ol style="list-style-type: none"> 1. 현재상태의 정보를 제공하는가(알람) 	<ol style="list-style-type: none"> 1. 현재상태를 다양한 척도로 보여주는가(정의, 시스템의 상태, 표준화) 2. 현재상태의인터페이스를 다양한척도로보여주는가(메뉴, 양식, 글꼴등) 3. 사용자는현재상태를문서화하였는가
E	<ol style="list-style-type: none"> 1. 복구에 대한 가이드가 있는가 2. 휴리스틱기반으로시스템을평가하여 복구가능성설계의고려사항으로채택하였는가 3. 되돌리기(실행및재시작)의복구가능성이명확한가 4. 사용자에게주의해야신호를주어야하는가 5. 어떤사용자라도소프트웨어를통제하기쉬운가 				

표 10. (continued)
Table 10. (continued)

휴리스틱 원칙	사용자 안전성	데이터 무결성	규제 준수	환자 치료효과	사용자 편의성
F	<ol style="list-style-type: none"> 1. 특정모드에서 작업수행이 불가능함을 강조하는가 2. 사용자의유연성과효율성을위하여긴급 상황에서도통제불가능한것이있는가 3. 사용자가기계적유연성과효율성을위하여시스템재시작을해야하는문제가있는가 				<ol style="list-style-type: none"> 1.사용자에게작업수행에수월한 기준이 있는가 2. 모델을정량화시킬수있는가 3. 사용하기가좋은가 4. 다음은효율적으로나타내고있는가 5. 효과적으로도움을주는것이유용한가 6. 사용하기에유연하고쉬운가 7. 누구든지사용하기에쉬운가 8. 사용자의사용방식이효율적인가 9. 효율적인도구를사용하고 있는가
G	<ol style="list-style-type: none"> 1. 사용자가 오해할만한 요소가 있는가 2. 사용자가알기쉬운용어나시스템이익숙한가 3. 내가원하는방식으로사용자를만족시켜주는가 				
H	<ol style="list-style-type: none"> 1. 사용자들은 핵심정보를 기록할 수 있는가 2. 사용설명서는명확하게작성되었는가 3. 중요한정보를전달해주는가 4. 디자인의필수요소는핵심정보를제공하는가 				<ol style="list-style-type: none"> 1. 핵심정보를 이해하기 쉽게 하여 작성하였는가 2. 핵심적인원칙이있는가
I	<ol style="list-style-type: none"> 1. 명확한 에러문구가 있는가 2. 에러발생시, 작업환경에서명확한피드백을받을수있는가 				<ol style="list-style-type: none"> 1. 사용자 경험을 향상시키기 위하여, 명확한 에러문구, 핵심정보에 대한 내용을 다루었는가 2. 표준언어사용,사용자정의가능한언어,그리고사용자중심컴퓨터인터페이스는사용자가시스템을 더효과적으로이해하고활용할수있는가
J	<ol style="list-style-type: none"> 1. 워크플로우에 대한 도움이 되는 피드백이 문서화로 작성되어져있는가 2. 표준작업정보가명확하게작성되어 있는가 				<ol style="list-style-type: none"> 1. 안내서에는 도움이 되는 내용이 잘 기술되어 있는가 2. 가이드에는도움이되는내용이 잘기술되어있는가 3. 코드설명은제3자가도움을받을수있도록명확하게작성되어져있는가 4. 도움서비스에대한내용의목적 이표시되는가 5. 사용자가필요로하는도움말을 제공하는가 6. 사용자에게다양한도움서비스를제공하는가

인 가능성에 대해 관찰하였다. 그 결과 SaMD 분야에 의료 기기의 안전성 및 유효성을 제공하는 중요한 도구로 사용적 합성 평가 요소와 가이드라인을 제시하여 그 필요성을 강조 하였다.

향후 휴리스틱 기반의 사용적합성 평가 가이드라인이 어떻게 의료 분야의 안전성과 효율성 향상에 도움을 줄 수 있는지 구체적인 예시와 사례를 통해 설명할 필요가 있다. 의

료 전문가와 환자 측면에서 사용성평가의 중요성과 이점을 강조하며, 사용성평가 가이드라인을 통해 환자의 치료 효과 개선과 사용자의 편리성을 어떻게 증진시킬 수 있는지도 명확히 전달해야 한다.

더불어, 향후 연구 방향에 대한 제언은 AI 기술의 진보와 의료 환경의 변화를 고려하여, 의료 분야에서 중요한 역할을 하는 특정 질병과 관련된 사항, 그리고 어떤 측면에 중

표 11. 개발 단계에 따른 사용적합성 평가 가이드라인 프로세스의 개선사항

Table 11. Improvements to the usability evaluation guideline process during the development phase

프로세스 단계	개선사항
평가 대상 정의	휴리스틱 기반 SaMD의 용도, 환경, 사용목적 등을 정의 및 평가대상의 특성 고려
평가항목 및 평가영역 선정	휴리스틱 기반 SaMD의 특성에 맞게 사용성, 안전성, 데이터 무결성, 규제 준수, 환자 치료 효과, 사용자 편의성 등을 선정
평가방법 설계	평가항목 및 평가영역에 대한 평가방법 설계를 하고, 국제 규격 기반 기준 설정
평가 및 결과 분석	휴리스틱 기반 SaMD의 사용적합성을 평가하고 결과를 분석하여 개선사항 수립
가이드라인 개선 및 보완	평가결과와 사용자 피드백을 토대로 가이드라인을 개선 및 보완

표 12. 개발 단계에 따른 위험 관리 활동 요약[35-36]

Table 12. Summary of risk management activities by development stage[35-36]

Development step	Process	Risk management step	
Development proposal	-	-	-
Development plan (Design and development input)	Risk analysis	Step 1	Establish intended use and reasonably foreseeable misuse
		Step 2	Identify characteristics related to safety
		Step 3	Identify hazards and hazardous situations
		Step 4	Estimate the risk(s) for each hazardous situation
	Risk Evaluation	Step 5	Is risk control required?
Development progress (Design and development review, validation, validation process)	Risk Control	Step 6-7	Identify appropriate risk control measures and Is risk control practicable?
		Step 7.1	Do the benefits outweigh the risk?
		Step 8	Implement and verify the identified risk control measures
		Step 9	Is the residual risk acceptable?
		Step 10	Are new hazards or hazardous situations introduced or existing risk affected?
		Step 11	Have all identified hazardous situations been considered?
	Evaluation of Overall Residual Risk	Step 12	Is the overall residual risk acceptable in relation to the benefits?
Development completed, design transfer (Design and development output)	Risk management report	Step 13	Review the execution of the risk management plan and prepare the risk management report
Post-production management	Production and post-production activities	Step 14	Collect production and post-production information

440

점을 두어야 하는지 제시해야 하며, SaMD에는 사용자안전성, 사용자편의성을 필수적으로 적용하고, 그 외 데이터무결성, 규제준수, 환자치료효과에 대해서는 고려해야하는 사항으로 추가적인 연구가 필요하다. 또한, 현실적인 도입 및 활용 방안을 구체화하고, 새로운 표준과 가이드라인 개발에 대한 필요성을 강조해야한다.

마지막으로, 의료진, 의료기기 제조업체, 규제과학자, 규제당국과 같은 다양한 이해관계자에게 어떻게 유용하게 적용할 수 있는지에 대한 명확한 설명이 필요하다. 휴리스틱 기반 SaMD 사용적합성 평가의 장점을 강조하고 현실적인 도입 및 활용방안을 제시함으로써, 연구의 결과를 강조할 수 있을 것이다.

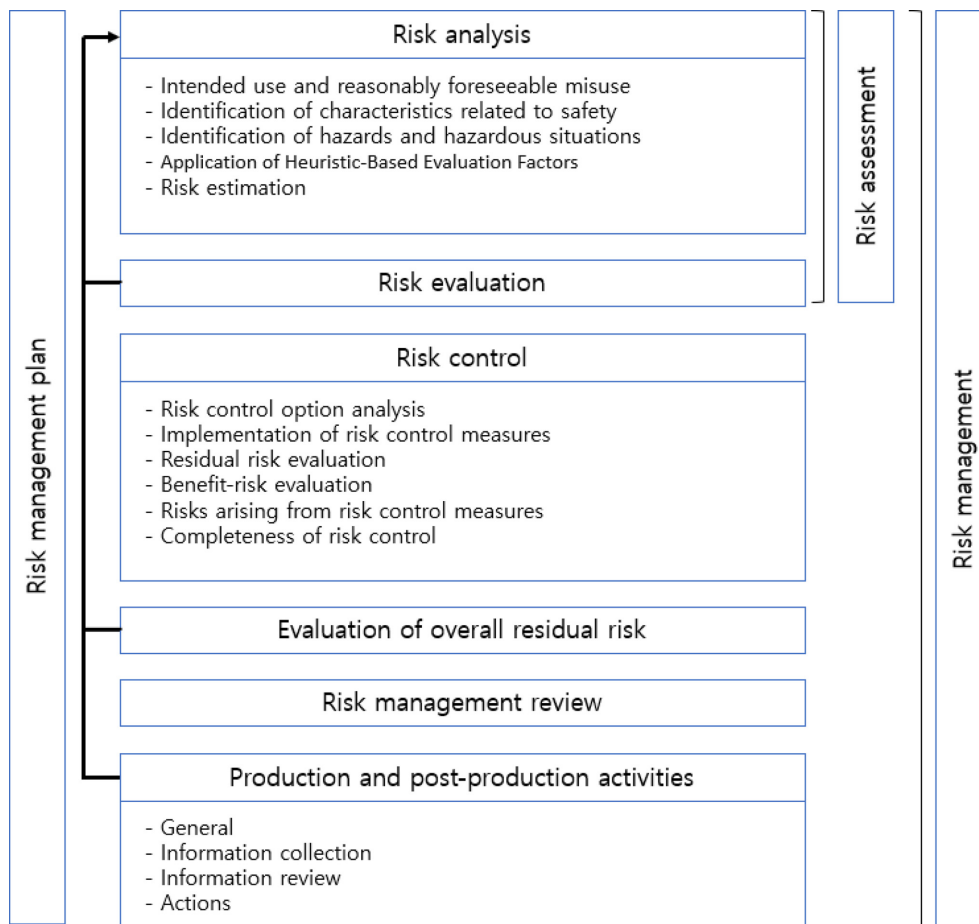


그림 1. 휴리스틱 원칙이 적용된 위험관리 프로세스 도해[37]

Fig. 1. Illustration of the risk management process with applied heuristic principles[37]

References

- [1] 식품의약품안전처, 의료기기GMP 기준 중 사용적합성 적용 가이드라인 (민원인안내서)- 심혈관영상분석소프트웨어, 2022, pp. 1.
- [2] Thankam Paul Thyvalikakath, Comparative study of heuristic evaluation and usability testing methods, 2009;143:322-327.
- [3] Tiffani J. Bright, Effect of Clinical Decision-Support Systems : A Systematic Review, 2012;157:20-43.
- [4] Nancy Stagers, Chapter 47. Patient Safety and Health Information Technology: Role of the Electronic Health Record, 2008 Apr. Chapter 47.
- [5] Pascale Carayon, Human Factors and Usability for Health Information Technology: Old and New Challenges, 2019;71-7.
- [6] Mark J. Graham, Heuristic evaluation of infusion pumps: implications for patient safety in Intensive Care Units, 2004;73: 771-779.
- [7] Patricia Flatley Brennan, Nursing Needs Big Data and Big Data Needs Nursing, 2015;47:5,477-484.
- [8] Jiajie Zhang, TURF: Toward a unified framework of EHR usability, 2011;1056-1067.
- [9] Andre W. Kushniruk, Cognitive and usability engineering methods for the evaluation of clinical information systems, 2004;56-76.
- [10] Nigel bevan, Measuring Usability as Quality of Use, 1995; 115-150.
- [11] Jiajie Zhang, Using usability heuristics to evaluate patient safety of medical devices, 2003;23-30.
- [12] John Brooke, SUS - A quick and dirty usability scale, 2011.
- [13] J Sauro, A Practical Guide to the System Usability Scale: Background, Benchmarks & Best Practices, 2011.
- [14] Aaron Bagor, Determining what individual SUS scores mean: Adding an adjective rating scale., 2009;114-123.
- [15] J Nielsen, Usability engineering, 1994, Chapter 6.
- [16] L Rosenfeld, information architecture for the world wide web, 2002;202.
- [17] Jonathan Lazar, Improving web accessibility: a study of webmaster perceptions, 2004;269-288.
- [18] B Shneiderman, Designing the user interface strategies for effective human-computer interaction, 2010;466.
- [19] A dix, Human-computer interaction, 2003.
- [20] James R. Lewis, The System Usability Scale: Past, Present, and Future, 2018;34:7,577-590.
- [21] Mckay, R.J., A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code, 2000;56-61.

- [22] JN stetson, A Comparison of Questionnaires for Assessing Website Usability, 2004.
- [23] Jeff sauro, Quantifying the User Experience: Practical Statistics for User Research, 2012;314:0-471.
- [24] Jakob Nielsen, Usability Inspection Methods, 1994;413-414.
- [25] Jakob Nielsen, Heuristic Evaluation of User Interfaces, 1990; 249-256.
- [26] Andre W. Kushniruk, Usability Testing in Medical Informatics: Cognitive Approaches to Evaluation of Information Systems and User Interfaces, 1997;2018-222.
- [27] John W. Beasley, Information Chaos in Primary Care: Implications for Physician Performance and Patient Safety, 2011; 24(6):745-751.
- [28] Alice C. Li, Integrating Usability Testing and Think-Aloud Protocol Analysis with "Near Live" Clinical Simulations in Evaluating Clinical Decision Support, 2012;761-772.
- [29] Farah Magrabi, An Analysis of Computer-Related Patient Safety Incidents to Inform the Development of a Classification, 2010;663-670.
- [30] Sylvia J Hysong, Understanding the Management of Electronic Test Result Notifications in the Outpatient Setting, 2011;11:22.
- [31] You Rim Kim, A Study on Safety, Performance and Clinical effectiveness Test Guideline of Versatile Ophthalmic Laser System, 2019;40:250-259.
- [32] e나라표준인증, Medical devices-Guidance on the application of KS P ISO 24971, 2020.
- [33] e나라표준인증, Medical devices-Guidance on the application of KS P ISO 14971, 2021.
- [34] ISO, Guidance on the application of KS P ISO 24971:2020; 2021.
- [35] Il Kyu Kang, ISO14971:2019 detailed analysis and Post Market Surveillance Application Method – Focusing on IVDR requirements (“a direct quotation”), 2022;43:199-213.
- [36] Min Woo Kang, Risk management application plan according to ISO14971:2019-Analysis through risk management of VR medical devices. Medical Device Engineering and Management, The Graduate School, Yonsei University, 2020;10-13.
- [37] ISO, Guidance on the application of KS P ISO 24971:2020, 2021.