

의료 데이터 산업을 위한 비정형 데이터 비식별화 정책에 관한 연구*

이 선 진*, 박 태 립**, 김 소 회***, 오 영 은****, 이 일 구*****

요 약

빅데이터 기술이 발전하면서 데이터가 전 산업의 혁신 성장을 가속하는 초연결 지능화 사회로 빠르게 진입하고 있다. 고품질의 다양한 데이터를 보유하고 활용하는 융복합 산업이 새로운 성장 동력으로 자리매김하고 있으며, 다양한 전통 산업군에 빅데이터가 융합되어 데이터 기반의 혁신을 통해 디지털 전환이 이루어지고 있다. 특히 의료 분야에서는 전자의무기록 데이터와 같은 정형 데이터와 CT, MRI 등의 비정형 의료 데이터를 함께 활용함으로써, 질병 예측 및 진단의 정확도를 높이고 있다. 현재 의료 산업에서 비정형 데이터의 중요성과 규모는 나날이 증가하고 있지만, 종래의 데이터 보안 기술과 정책은 정형 데이터 중심이며, 비정형 데이터의 보안성과 활용성에 대한 고려는 미비하다. 향후 빅데이터를 활용한 진료가 활성화되려면 데이터의 다양성과 보안성이 데이터 구축, 유통, 활용 단계에서 내재화되고 유기적으로 연계되어야 한다. 본 논문에서는 국내의 데이터 보안 제도와 기술 현황을 분석한다. 이후 의료 분야에서 비정형 데이터가 활발히 사용될 수 있도록 비식별조치 가이드라인에 비정형 데이터 중심의 비식별 기술과 산업에서의 기술 적용 사례를 추가하고, 비정형 데이터에 대한 개인정보 판단 기준을 수립할 것을 제안한다. 더 나아가 개인정보를 침해하지 않고, 비정형 데이터에 활용할 수 있는 객체 특징 기반의 식별 ID를 제안한다.

A study on the policy of de-identifying unstructured data for the medical data industry

Sun-Jin Lee*, Tae-Rim Park**, So-Hui Kim***, Young-Eun Oh****, Il-Gu Lee*****

ABSTRACT

With the development of big data technology, data is rapidly entering a hyperconnected intelligent society that accelerates innovative growth in all industries. The convergence industry, which holds and utilizes various high-quality data, is becoming a new growth engine, and big data is fused to various traditional industries. In particular, in the medical field, structured data such as electronic medical record data and unstructured medical data such as CT and MRI are used together to increase the accuracy of disease prediction and diagnosis. Currently, the importance and size of unstructured data are increasing day by day in the medical industry, but conventional data security technologies and policies are structured data-oriented, and considerations for the security and utilization of unstructured data are insufficient. In order for medical treatment using big data to be activated in the future, data diversity and security must be internalized and organically linked at the stage of data construction, distribution, and utilization. In this paper, the current status of domestic and foreign data security systems and technologies is analyzed. After that, it is proposed to add unstructured data-centered de-identification technology to the guidelines for unstructured data and technology application cases in the industry so that unstructured data can be actively used in the medical field, and to establish standards for judging personal information for unstructured data. Furthermore, an object feature-based identification ID that can be used for unstructured data without infringing on personal information is proposed.

Key words : Medical industry, Unstructured data, De-identification, De-identification action, De-identification policy

접수일(2022년 06월 29일), 수정일(2022년 09월 06일),
게재확정일(2022년 09월 27일)

★ 본 연구는 2022년도 정부(과학기술정보통신부의 지원으로 한국연구재단의 지원(No. 2020R1F1A1061107)과 2022년도 정부(산업통상자원부의 지원으로 한국산업기술진흥원의 지원(P0008703, 2022년 산업혁신훈인재성장지원사업), 과학기술정보통신부 및 정보통신기획평가원의 ICT혁신인재4.0 사업의 연구결과로 수행되었습니다.
(IITP-2022-RS-2022-00153310).

- * 성신여자대학교 미래융합기술공학과 (주저자)
- ** 성신여자대학교 미래융합기술공학과 (공동저자)
- *** 신세계아이앤씨 정보보안팀 (공동저자)
- **** 라인비즈플러스 viz security (공동저자)
- ***** 성신여자대학교 융합보안공학과 / 미래융합기술공학과 (교신저자)

1. 서 론

4차 산업혁명과 함께 가속화되고 있는 데이터 중심의 디지털 트랜스포메이션은 의료 산업 전반의 서비스 체계를 근본적으로 바꾸고 있다. 과거에는 의료 서비스를 제공받던 사용자들이 인공지능, 사물인터넷, 빅데이터 등의 신기술을 기반으로 언제 어디서나 네트워크에 연결되어 실시간으로 데이터를 생산하고 있다. 이에 따라 의료 산업은 병원 중심의 의료정보(Electronic Medical Record (EMR), Electronic Health Record (EHR))에서 개인 중심의 의료정보(Personal Health Record (PHR)) 기반으로 변화하고 있다[1]. 보건복지부에 따르면 국내 보건의료 데이터 및 인공지능 시장은 2019년 554억 원에서, 2023년 2,456억 원으로 고속 성장할 것으로 예측되며[2], COVID-19로 초래된 팬데믹 상황에 대한 방역에서도 보건의료 정보 활용의 중요성이 입증되었다[3]. 또한, 수술실 폐쇄회로 설치 의무화 법률이 2021년 8월 국회 본회의를 통과하면서, 2023년 하반기부터 본격적으로 시행될 예정이다[4].

이처럼 의료 산업에서의 데이터 관련 수요 및 활용 영역은 점차 높아지고 있지만, 정형 및 비정형 데이터에는 개인정보와 민감정보가 포함되어 있어서 데이터 활용에 제약이 따른다[5]. 일례로 수술실 폐쇄회로 설치를 들 수 있다. 의료인 측에서는 수술 영상 촬영이 의료인의 수술 방법 및 조치를 제한시키고, 의료인의 사기를 떨어뜨릴 수 있다는 점을 들어 우려하고 있는데, 이것은 의료인의 정보 및 행위가 노출되기 때문이다. 의료 산업에서 정형 및 비정형 데이터를 활용하기 위해서는 특정 수준의 프라이버시를 유지하면서 데이터 활용 범위와 응용 분야를 확장·보완해 나가는 것이 중요하다.

오늘날의 데이터 범주에는 기존에 활용도가 높던 정형 데이터뿐 아니라 이미지, 영상, 음성 등의 비정형 데이터가 큰 비중을 차지하고 있다. 정형 데이터는 지정된 필드에 저장되어 곧바로 분석할 수 있는 데이터를 의미하고, 비정형 데이터는 사진, 영상, 음성 등 정해진 형태가 없는 데이터를

의미한다[6]. 비정형 데이터의 중요성이 증가함에 따라 여러 산업 분야에서는 비정형 데이터를 활용한 학술적·경제적 유용성에 주목하고 있다. 의료 분야에서는 Computed Tomography (CT) 혹은 Magnetic Resonance Imaging (MRI)와 같은 고차원 의료 영상이 연간 10억 건씩 촬영되는 것으로 예측되는데[7], 이러한 모든 데이터는 전산 처리되어 네트워크를 통해 유관 부서로 전달되고 있다[8]. 시계열 의료 영상 데이터는 유용한 데이터이지만, 의료 영상 데이터를 빅데이터로 활용하기 위해서는 개인정보와 민감정보 유출에 대한 안전조치와 제도가 마련되어야 한다.

전 세계적으로 의료 데이터 활성화를 위한 규제 혁신과 산업 육성이 추진되면서, 데이터 비식별 제도와 가이드라인이 마련되었다. 국내에서는 데이터 3법 개정으로 전통 산업과 신기술의 결합을 통해 데이터를 활용한 새로운 가치 창출이 기대되고 있다. 하지만 종래의 기술과 정책은 전통 산업에서 주로 활용되어 온 정형 데이터 중심으로 발전되어 왔으며, 비정형 데이터에 대한 고려가 부족한 실정이다. 또한, 비정형 데이터 개방을 위한 보안 기술이 개발되고 있으나 비식별조치의 결과로 데이터의 유용성이 저하되는 문제를 해결하지 못하고 있다. 의료 산업에서 빅데이터 경제가 활성화되려면 다양한 형태의 데이터를 안전하게 활용할 수 있는 기술적, 제도적 기반이 마련되어 데이터 구축, 유통, 활용 단계 전반에 걸쳐 디지털 기술과 데이터 보안 정책이 유기적으로 결합되어 운용되는 환경에서 데이터를 활용해야 한다.

본 논문에서는 앞서 서술한 문제점들을 해결하고자 국내외 데이터 보안 기술과 제도를 분석해 현시점에서의 정책적, 기술적 한계점을 파악한다. 그 후, 의료 산업과 산업보안 기술 활성화를 위한 비정형 빅데이터 비식별화 및 추적 기술 조사를 바탕으로 정책적 개선안을 제시한다.

2. 종래 데이터 비식별 기술·정책 분석

최근 전 산업 분야에서 데이터양이 급증하면서,

증가하는 개인정보 유출 사고와 함께 개인정보보호의 중요성이 유례없이 강조되고 있다. 개인정보 보호를 강화하기 위해 미국, EU, 일본에서는 각각 비식별정보, 가명정보, 익명가공정보를 활용하는 정책을 도입하고 있다. 이 정책에서는 데이터 활용을 위해 비식별조치, 가명화, 익명화 기법을 사용하여 데이터 결합을 촉진하고 있으며[9], 관련 법률의 제정 및 개정도 활발히 진행되고 있다. 이러한 주요 국가들의 비식별 처리에 대한 기준과 절차를 제시하는 주요 법률과 가이드라인의 내용은 사전 검토, 비식별조치, 데이터 결합, 적정성 평가, 사후관리단계로 나뉘며, 정형 데이터 및 비정형 데이터 모두 같은 단계로 비식별되고 비식별 조치 단계에서만 기술적 차이를 보인다.

2.1 사전 검토 단계

사전 검토 단계는 데이터가 개인정보에 해당하는지를 검토하는 단계이다. 개인정보에 해당하지 않으면 별도 조치 없이 자유롭게 활용할 수 있지만, 개인정보에 해당하는 경우에는 비식별조치를 해야 한다[10]. <표 1>은 해외 주요국과 한국에서 정하는 개인정보의 정의와 범위를 비교한 표이다.

<표 1> 국내외 개인정보의 범위

| 구분 | 개인정보의 정의 | 비정형 데이터 언급 범위 |
|----|---|---|
| 한국 | 살아있는 개인에 관한 정보로서 성명, 주민등록번호 및 영상 등을 통하여 개인을 알아볼 수 있는 정보. 다른 정보와 쉽게 결합하여 알아볼 수 있는 것 포함 | 정보의 종류, 형태, 형식 등에 관하여는 특별한 제한이 없음 |
| 미국 | 개인에 관한 정보로 개인의 성명 또는 신분번호, 기호, 지문, 사진 등 개인에게 배정된 신분의 식별을 위한 특기사항 | 개별법마다 다름 |
| EU | 식별되었거나 식별 가능한 자연인 즉 정보 주체에 관련된 모든 정보 | 얼굴 이미지를 민감 데이터로 분류 |
| 일본 | 생존하는 개인에 관한 정보로서 성명·생년월일·기타 기술 등에 기재 혹은 기록되거나 음성, 동작 및 그 밖의 방법을 사용하여 | 개인 식별부호(DNA, 머리의 부위의 위치 및 형상에 따라 정한 용모, 흉채, |

| | |
|--|-------------|
| 표시된 일체의 사항에 의해 특정의 개인을 식별할 수 있는 것(다른 정보와 용이하게 조합할 수 있고, 그에 의해 특정의 개인을 식별할 수 있는 것 포함) | 보행 자세 등) 포함 |
|--|-------------|

<표 1>과 같이 현재 국내 규정에서는 개인정보의 개념과 범위가 상당히 포괄적으로 정의되어 있다. 개인정보보호 관련 법률 또는 해설서에서 개인정보의 범위에 대한 해설이 제공되지만, 모든 분야를 커버할 수 없어 실무적으로 판단하기 어려운 상황이다. 그리고 비정형 데이터에 대한 구체적인 내용은 거의 언급되지 않는 것을 알 수 있다. 해외에서는 일부 비정형 데이터 사례를 개인정보 개념에 포함하고 있으나 국내에서는 정보의 종류, 형태를 아직 세분화되지 않았다.

또한 현재 국내 법률에서는 데이터가 개인정보에 해당하는지 판단할 근거가 미흡하다. 하지만 국내외 모두 현 법률과 가이드라인에서 언급하는 개인정보의 개념이 포괄적으로 명시되어 있어 개인정보의 범위를 구체적으로 설정하기 어렵다[14]. 국내에서는 개인정보보호 법령 및 지침·고시 해설서, 개인정보 비식별조치 가이드라인을 제공하여 개인정보의 판단기준에 대한 해설을 제공하고 있으나, 이는 법률 해석에 그칠 뿐 다수의 실무적 상황에서 보호해야 할 개인정보의 범위를 어떻게 설정할지 구체적인 기준과 사례를 제공하고 있는 것은 아니다. 특히 정형 데이터 위주의 정책은 비정형 데이터에서의 개인정보, 가명정보, 익명정보를 구분하는 기준이 모호하므로 비정형 데이터를 활용하는 데 어려움이 따른다.

2.2 비식별조치 단계

비식별조치 단계는 정보 집합물에서 비식별 기법을 통해 개인을 알아볼 수 없도록 하는 단계이다. 개인 또는 관련된 사물에 고유하게 부여된 값인 식별자는 원칙적으로 삭제하고, 데이터 이용에 필요한 식별자는 비식별조치 후 활용할 수 있다[10].

국제 표준단체 ISO의 ISO/IEC 20889[15]와 국내 비식별조치 가이드라인에 따르면 일반적인 데이터 비

식별조치에는 가명처리, 총계처리, 데이터 삭제, 데이터 범주화, 데이터 마스킹이 있다. 이와 같은 비식별 기법은 데이터 종류에 따라 정형 데이터에 범용적으로 적용될 수 있고, 부분적으로 비정형 데이터에도 적용할 수 있다. <표 2>는 비식별 기법별 적용 가능한 정형 및 비정형 데이터를 나타낸 것으로, 국내 개인정보 비식별조치 가이드라인[10]을 재구성한 것이다.

<표 2> 비식별 기법 분류

| 처리방법 | 세부기법 | 적용가능한 정형 데이터 | 적용가능한 비정형 데이터 |
|---------|---------------------------------------|-------------------------|------------------|
| 가명처리 | 휴리스틱 가명화, 암호화, 교환방법 | 성명, 사용자 ID 기호번호 등 | 텍스트, 이미지, 영상 |
| 총계처리 | 총계처리, 부분총계, 라운딩, 재배열 | 나이, 신장, 소득 등 | 텍스트 |
| 데이터 삭제 | 식별자 삭제, 식별자 부분 삭제, 레코드 삭제, 식별요소 전부 삭제 | 성명, 전화번호, 위치 정보, 소득 등 | 텍스트, 이미지, 영상, 음성 |
| 데이터 범주화 | 감추기, 랜덤 라운딩, 범위 방법, 제어 라운딩 | 나이, 소득, 카드지출액, 서비스 이용 등 | 텍스트 |
| 데이터 마스킹 | 임의 잡음 추가, 공백과 대체 | 사용자 ID, 성명, 생년월일 등 | 텍스트, 이미지, 영상, 음성 |

정형 데이터는 <표 2>의 기법을 통해 비식별조치가 가능하다. 비정형 데이터의 범주에는 이미지, 영상, 음성이 포함되며, 이러한 비정형 데이터는 정형화 과정을 거친 후 비식별 기법을 적용할 수 있다. 비식별 기법 중 가명처리를 할 수 있는 비정형 데이터에는 이미지, 영상이 있다. 비식별화 과정에서 신분증의 개인정보나 중요서류를 암호화하여 저장하기도 하고, Closed-circuit Television (CCTV)에 출현하는 사람의 얼굴을 합성하여 대체하기도 한다. 데이터 삭제 또한 이미지, 영상, 그리고 음성에 적용할 수 있다. 원본 데이터에서 식별자나 속성자에 해당하는 부분 또는 전체를 삭제하여 비식별화 조치한다. 데이터 마스킹은 이미지, 영상, 음성 데이터를 가공할 때 가장 흔히 사용되는 기법의 하나로, 영상에 모자이크와 같이 필터를 걸어서 픽셀에 잡음을 추가하거나 음성 중 식별자

에 해당하는 부분에 임의의 잡음을 추가하여 비식별화한다. 이 중 총계처리와 데이터 범주화는 숫자로 된 데이터를 통계적 기법을 사용하여 가공하기 때문에 비정형 데이터에 적용하기 적합하지 않다.

현재 국내의 개인정보보호법이나 비식별조치 연구는 정형 데이터에 초점이 맞추어져 있어서 의료 산업 대부분을 차지하고 있는 비정형 데이터가 온전히 활용되기 어렵다. 최근 발표된 가명정보 처리 가이드라인의 개인정보 가명처리 기술에서도 정형 데이터만을 예시로 들고 있다. 분류상 비식별조치 기법은 정형 데이터와 비정형 데이터에 모두 적용할 수 있으나, 적용 데이터 종류는 정형 데이터로 제한되어 있다. 해외에서도 비정형 데이터 비식별조치 기법에 관한 연구는 꾸준히 진행되고 있지만, 관련 법률과 정책은 뒷받침되지 못하고 있다. 즉, 의료 데이터 산업의 활성화를 위해서는 각 기법을 사진, 영상, 음성과 같은 비정형 데이터의 적용 방법에 관한 연구가 필요하다.

2.3 데이터 결합 단계

가명처리 된 개인정보가 빅데이터로 활용되기 시작하면서 데이터 결합 과정에서의 비식별조치 중요성이 커지고 있다. 결합 과정에서 재식별 위험이 커지기 때문에 가이드라인에서는 정보 집합물을 결합할 때 서로 같은 알고리즘을 사용해서 임시 대체기에 잡음을 추가하거나 식별자 중 일부를 조합하고 결합대상 정보 집합물도 비식별조치를 수행하여 결합 과정에서 개인을 식별하지 못하도록 처리할 것을 권고한다[10]. 또한, 가명정보 처리 가이드라인에서 데이터 결합 전문기관은 결합키 연계 정보를 통해 복수의 데이터를 결합하며, 데이터 결합 신청자는 반출된 가명정보를 결합 신청 목적에 따라 처리할 수 있다[16].

국내뿐만 아니라 해외에서도 일정 조건을 만족시키면 기업 간 데이터 결합을 허용한다. 미국은 단일화된 법 없이 자율규제 방식을 진행하고 있기에 개별법에서 제한하지 않으면 자유로운 데이터의 이용과 결합을 보장하고 EU에서는 제3의 신뢰기관의 참여와 관리·감독이 보장될 때 데이터 결합을 허용한다[10]. 일본도 데이터의 자유로운 이용과 결합을 허용하고 있으나 개인정보를 위탁받은 자에 대해 철저한 관리

감독이 보장될 때 결합이 가능하다[17].

그러나 데이터 결합 단계에서도 개인정보의 예시를 정형 데이터에만 기초하여 설명하고 있고 데이터의 결합 방식이나, 결합 단계에서의 비정형 데이터 처리 방식에 대한 설명은 부족한 상황이다.

2.4 적정성 평가 및 사후관리 단계

적정성 평가 단계와 사후관리 단계는 결합에 따른 개인정보 재식별 위험성을 평가하고, 정보의 보호조치를 이행하는 단계이다. 데이터의 비식별 수준이 적정하다고 평가될 때는 데이터를 활용할 수 있지만, 그렇지 않으면 추가적인 비식별조치가 필요하다. 데이터를 활용하는 과정에서도 비식별 정보에 필수적인 보호조치를 이행하고 재식별 가능성을 모니터링하는 사후관리가 요구된다[10]. <표 3>은 국내외 적정성 평가 및 사후관리 방법을 비교한 표이다.

<표 3> 국내외 적정성 평가와 사후관리

| 구분 | 적정성 평가 | 사후관리 |
|----|---------------------------|---|
| 한국 | k-익명성 의무, 1-다양성, t-접근성 선택 | 관리적/기술적/물리적 보호조치 수행 |
| 미국 | 전문가 결정방식 | 재식별 리스크 정의하여 사후관리 강조 데이터 제공처에 식별금지 계약 |
| EU | k-익명성, 1-다양성, t-접근성 중 선택 | 관리적/기술적 조치 후 별도분리 재식별에 대한 최소한의 검증 실시 |
| 일본 | k-익명성 등 | 재식별 금지 익명가공정보를 다른 정보와 조합할 수 없음(법제 38조) |

2.4.1 적정성 평가

한국, EU, 일본은 데이터베이스에 같은 값을 가지는 레코드를 적어도 k개가 있도록 하여 개인을 특정할 수 없게 하는 모델인 k-익명성을 공통적으로 적정성 평가에 적용한다[19]. k-익명성 모델 이외의 다른 프라이버시 모델의 적용 방식에 따라 나라별 적정성 평가 방식의 차이가 존재한다. 한국은 개인정보보호 책임자와 외부 전문가로 구성된 적정성 평가단이 k-익명성을 활용하여 식별 가능성을 종합적으로 평가한

다. k-익명성은 최소한의 평가 수단으로 의무화되어 있고 데이터의 특성, 재식별 가능성 등을 종합적으로 판단하여 추가적인 평가모델을 결정할 수 있다[10]. 이와 달리 EU는 적정성 평가 과정에 k-익명성, 1-다양성, t-접근성 프라이버시 모델을 선택적으로 적용한다. 일본은 k-익명성 등의 프라이버시 모델을 통해 적정성 평가를 수행한다.

미국은 개인정보보호에 대한 일반법이 존재하지 않기 때문에 개별적인 분야에서 개인정보를 보호하는 분야별 규율방식을 채택한다[20]. 미국의 보건의료 분야 법률인 Health Insurance Portability and Accountability Act (HIPPA)에 따르면 국내 비식별조치 적정성 평가단과 유사하게 개인정보를 검토하는 전문가를 구성하여 전문가 결정 방식으로 적정성 평가를 진행한다. 반복성, 데이터 소스의 접근 가능성, 구별성을 기준으로 재식별 위험을 최소화했는지 판단한다[21].

2.4.2 사후관리

국내는 개인정보보호법 시행령 및 행정규칙에서 개인정보 및 가명정보의 기술적, 관리적, 물리적 보호조치를 명시하고 있다. 가명정보 처리 가이드라인에 따르면 적정성 검토 및 추가처리 단계에서 재식별 가능성을 검토하고 제3자에게 제공할 때 재식별 위험을 고려하도록 하고 있다. 또한, 가명정보를 처리하는 과정에서 특정 개인을 알아볼 수 있는 정보가 생성된 경우에는 즉시 해당 정보의 처리를 중지하고 지체없이 회수·파기하여야 한다고 규정하고 있다[16]. 이외에도, 비식별 정보의 재식별을 방지하기 위해 비식별 정보 파일 대장을 관리하거나, 이용목적 달성 시 지체 없이 파기하는 관리적 보호조치를 취하고 있으며, 비식별 정보파일에 대한 접근권한 관리 및 접근통제, 접속 기록 관리 등 기술적 보호조치에 대해서도 명시하고 있다[10].

미국은 개인정보 재식별 가능성을 인정하여 의료, 교육 등의 개별 법령에서 정보주체의 개인정보를 재식별하지 않을 것, 제3자가 데이터를 활용할 때 재식별을 방지할 것, 데이터 제공처에서 개인정보를 재식별하지 않도록 계약에 명시할 것을 권고하고 있다[20]. 일본에서도 정보취급자가 익명가공정보를 재식별

하기 위해 다른 정보와 조합하거나 가공방법에 대한 정보 취득을 막고, 익명가공정보에 대한 재식별을 금지하여 익명성을 보존할 수 있도록 조치하고 있다[13]. EU에서는 재식별에 대한 최소한의 검증을 실시하고 재식별될 경우에는 해당 정보를 파기할 것을 권고하고 있다[17].

3. 비정형 빅데이터 비식별조치 기술

본 장에서는 의료 데이터의 대부분을 차지하는 비정형 데이터에 대한 현 비식별조치 기술에 대해 살펴본다.

일반적인 데이터는 앞서 언급한 비식별조치 기술로 가공하여 처리할 수 있지만, 영상, 음성과 같은 데이터는 데이터 특성을 고려하여 알맞은 비식별 기법을 선택해야 재식별 요소를 최소화하면서 활용도를 높일 수 있다. <표 4>는 현재 산업에서 활용되고 있는 대표적 비정형 데이터인 영상 데이터의 비식별 기법을 나타낸 것으로, 영상 데이터 익명화 기술 및 평가방안 보고서[22]를 재구성한 표이다.

<표 4> 영상 데이터 비식별 기법

| 처리방법 | 설명 | 장점 | 단점 |
|---------|--|---|---|
| 이미지 필터링 | 영상의 식별자에 여러 필터를 적용하는 기법 | 간단한 처리 | 딥러닝 기술로 일정 수준 복원 가능 |
| 이미지 암호화 | 영상을 암호화하여 허가된 대상에게만 공개하는 기법 | 원본 영상으로 복원 가능 네트워크를 통해 영상을 전송할 때 사용 | 실시간 처리 어려움 키가 없으면 암호화된 영상 데이터 복호화 불가 |
| 얼굴 합성 | 영상 데이터에 적합하게 k-익명성 모델을 확장 수집한 얼굴 이미지 집합 내에서만 비슷하게 k개 | 수학적으로 보장되는 개인 식별 방지 수준 제공 합성한 얼굴을 대체하는 기법으로 익명화된 데이 | 실시간 영상 처리가 어려움 익명화 단계에서 유용성을 고려하지 않으면 유용 |

| | 의 얼굴을 합성하는 기법 | 터의 활용도 높음 | 성 보장 불가 원본 얼굴로 복원 불가 |
|-------|---|---|---|
| 인 페인팅 | 영상에서 특정한 부분을 제거하고 생긴 공백 또는 손상된 부분을 채우는 기법 | 제거된 대상에 대해서 어떠한 시각적 정보도 남지 않으면 영상 내에서 특정한 목적과 무관한 사람 등 제거 | 실시간 처리 어려움 제거 영역이 큰 경우 복구 불가 데이터 유용성 하락 |

영상에서의 비식별조치는 식별자가 되는 얼굴을 검출한 후 개인 식별 영역을 변형하는 것이다. <표 4>에서 언급하는 바와 같이 대표적인 비식별조치 기법으로는 이미지 필터링, 이미지 암호화, 얼굴 합성, 인페인팅이 있다.

이미지 필터링은 블러링, 모자이크와 같이 영상에서 식별자 부분을 흐리게 하거나 블록 단위로 밝기를 조정하는 방식으로 간단하게 사용할 수 있는 기법이지만 이미지를 다시 복원할 수 있어 개인정보 유출 위험이 있다. 이미지 암호화는 영상 자체를 암호화하여 허가된 대상에게만 공개하는 기법으로, 복호키가 있어야만 복호화할 수 있어 안전하지만, 연산량이 많아서 속도가 느리며 키가 없으면 영상 데이터를 활용할 수 없다. 일 예로, 이경률, 임강빈, 이선영은 형태보존암호를 활용한 영상 내 마스킹 기술을 제안하였다[23]. 이 기술은 평문과 암호문이 동일한 형태를 가지는 형태보존 암호 기술을 활용하여 암호화를 할 때 이미지 크기가 증가하는 기존의 한계점을 보완하면서 영상에 포함된 개인정보를 보호한다. 얼굴 합성은 영상 데이터에 k-익명성 모델을 적합하게 확장하여 얼굴 객체 내에서만 비슷한 k개의 얼굴을 합성하는 방법이다. 이 방법은 합성한 얼굴을 대체하여 익명화된 영상 데이터의 활용도가 높다. 하지만 얼굴 합성은 저장된 얼굴 이미지를 사용하므로 실시간 영상 처리가 어렵다. 그리고 익명화 단계에서 유용성을 보장할 수 없으며 합성한 얼굴에서 원본 얼굴로 복원 불가능하다.

정형 데이터와 비정형 데이터의 비식별 기법은 기술적으로 큰 차이를 보인다. 그러나 현재 영상, 음성과 같은 비정형 데이터 비식별 기술의 연구는 정형 데이터보다 미흡한 상황이다. 비정형 데이터를 비식별화할 때 현재 비식별조치에서 제안하고 있는 암호화 및 블러링 기법을 활용하면 데이터의 재식별 위험이 존재한다. 그러나 산업계에서는 통상적으로 영상이나 이미지를 비식별할 때 여전히 마스킹, 암호화 및 블러링 기법을 사용하고 있다. 이 방식은 제거 기술이나 변형 기술보다 미적 요소가 보존되고, 다양한 분야에 간편하게 적용할 수 있지만, 데이터의 원본 복원이 비교적 쉽게 가능하고, 재식별 위험이 있다는 한계가 존재한다. 실제로 Meden, Emersic, Struc & Peer는 k-익명성 모델을 Generative Neural Network (GNN)과 결합하여 합성 이미지를 생성할 때 얼굴 표정, 성별, 나이와 같은 특징을 유추해내는 방법을 제안하기도 하였다[24]. 한편, 제거 및 변형 기술을 택하는 경우, 데이터의 활용이 어렵다. 비정형 데이터에서의 제거, 변형 기술은 개인정보에 해당하는 부분의 데이터만을 일부 삭제하거나 합성하여 사용되고 있다. 데이터를 제거, 변형하는 기술은 원본 영상 복원이 불가능해 상대적으로 다른 비식별 기법에 비해 재식별 위험에서 안전하므로 실무에서도 다양하게 활용되고 있다. 그러나, 확인할 수 있는 대부분의 식별 정보가 사라지고, 객체 식별이나 추적 가능성이 사라져 원본 데이터의 가치가 훼손되므로 실질적인 데이터 활용이 어렵다는 단점이 존재한다.

의료 산업에서는 종래 비식별 기술의 한계점을 보완하기 위해, 비정형 데이터를 다양한 기술로 비식별화하여 보안성을 강화하고, 데이터 내 객체를 인식 및 추적하여 활용성을 높이는 기술이 최근 활발히 연구되고 있다. Arsalan Shahid 등[25]은 Digital Imaging and Communication On Medicine (DICOM) 파일 형식으로 제공되는 CT 스캔 이미지에 대해 2단계의 비식별 프로세스를 제안한다. 1단계에서는 자체 Picture Archiving and Communication System (PACS)을 통해 환자의 Personally Identifiable Information (PII)을

제거하고, 2단계에서는 DICOM 헤더에 대한 속성 조사를 통해 추가 비식별화가 필요한 PII를 비식별화하여 보안성을 향상한다. Andreea Bianca Popescu 등[26]은 Variational AutoEncoder (VAE)와 임의의 비편향 픽셀 강도 매핑을 통해 비정형 데이터의 콘텐츠를 보호하는 이미지 난독화 알고리즘을 제안했다. 관상동맥 조영술에서 활용되는 데이터에 대해 제안 모델을 평가한 결과, 공격자가 타겟 데이터에 대한 사전 지식을 가지고 있어도 데이터를 완전히 복구할 수 없음이 증명되었다. 이처럼 차별화된 의료 비정형 데이터의 비식별 기법은 데이터의 활용성을 해치지 않으면서 보안성을 높일 수 있다.

한편, 보안성과 더불어 데이터의 활용성을 더욱 높인 연구도 있다. N. Anh Tu의 연구[27]에서는 Kafka 및 Spark Streaming을 기반으로 동작하는 얼굴 인식 및 트래킹 메커니즘에 기반한 영상 데이터의 프라이버시 보호 솔루션을 제안한다. 연구에서는 영상 데이터에서 개인정보 유출을 줄이고 특정 개인의 식별을 방지하기 위해, 얼굴 정보를 사용하지 않은 채로 동작 및 신체 자세와 얼굴을 연관시켜 개인을 트래킹할 수 있도록 했다. 이러한 트래킹의 경우, Open Pose를 사용하고 Face and body association (FBA) 병렬화 및 Hadoop Distributed File System (HDFS)을 통해 객체 추적에서 높은 효율성과 속도 향상을 보였다. 또한, HDFS를 활용하여 기존 영상 비식별화의 속도 및 오버헤드의 한계점을 극복하여, 실시간으로 객체 비식별이 가능하게 했다.

이러한 선행연구로 미루어보았을 때, 높은 보안성이 보장되는 환경에서 비식별화된 데이터 내 객체를 트래킹하는 기술이 연구 개발되고 있음을 알 수 있다. 그러나 현재 국내 정책에서 제시하고 있는 비정형 데이터의 비식별 기술은 정형 데이터의 비식별 기술을 활용하고 있어 단편적인 비식별만 가능한 상황이다. 김순석 등[28]은 의료 분야에서 다루어져야 할 주요 요구사항을 정리하였다. 특히 논문에서는 보호되어야 할 개인의료정보를 명확히 정의해야 하고, 개인의료정보의 민감성을 고려하여 높은 수준의 프라이버시 모델이 필요함을 주장했다. 박

민영 등[29]도 의료 산업에서 개인정보의 정의와 범위를 명확히 규정해야 하고, 비식별 기법을 산업에 적용하는 연구개발의 필요성을 강조했다. 이처럼 비정형 데이터 비식별조치 기술을 실 산업에 적용하기 위해 비식별조치 정책에 기술적 고려가 필요하다.

4. 의료 데이터 산업 활성화를 위한 비정형 빅데이터 비식별조치 정책

본 장에서는 의료 데이터 산업에서 비정형 데이터의 활용성 및 보안성을 높이기 위해 비식별조치 정책에 대한 세 가지 개선안을 제안한다.

4.1 비식별조치 정책 내 비정형 데이터 비식별 기술 및 사례 도입

현재 비식별화 가이드라인과 가명정보 처리 가이드라인에서는 비식별조치 기법이 명시되어 있지만, 해당 기법을 적용할 수 있는 실 데이터의 예시는 정형 데이터 중심이다. 가이드라인 내 개인정보 비식별조치 방법에서 제시한 일반적 비식별화 기법의 대상은 모두 정형 데이터에 해당한다. 2020년에 발표된 가명정보 처리 가이드라인에서는 가명정보 결합 내용이 추가되었고, 이후 2021년 10월 가명정보 처리 가이드라인 개정을 통해 가명정보 활용을 확산하기 위한 결합 방식을 개선하였으나, 이 또한 여전히 비정형 데이터의 비식별 사례 언급은 없는 상태이다. <표 5>를 통해 정형 데이터와 비정형 데이터 각각에 대한 비식별조치 정책과 특징의 차이를 비교할 수 있다.

<표 5> 정형 및 비정형 데이터 비식별조치 비교

| 구분 | 정형 데이터 | 비정형 데이터 |
|---------|---------------------------------------|-----------------------|
| 비식별 기법 | 가명화, 집계, 데이터 삭제, 데이터 범주화, 데이터 마스킹[10] | 비정형 데이터에 특화된 비식별기법 없음 |
| 데이터 형식 | 이동성 없음 | 이동성 존재 |
| 비식별화 대상 | 비식별 대상 확인 가능 | 비식별 대상의 사전 정의 필요 |

<표 5>에 따르면, 정형 데이터의 비식별 기법은 비정형 데이터에 비해 세부적으로 표현되어 있으나,

비정형 데이터에 맞춰진 비식별조치 정책은 전무하다. 정형 데이터는 고정적인 텍스트를 비식별화하므로 비식별 대상이 명료하지만, 일부 비정형 데이터의 경우, 데이터 내 객체가 움직이거나, 객체의 크기가 일정치 않기 때문에 비식별 대상에 대해 사전 조치가 필요하다.

특히 의료 데이터 산업에서 활용되는 수술실 내 CCTV 영상, MRI, Positron Emission Tomography (PET)-CT와 같은 판독 사진, 원내에서 활용하는 의료정보 아카이브 등은 대부분 비정형 데이터로 구성되어 있다. 따라서 의료 데이터의 성질과 형식에 최적화된 비식별 가이드라인을 마련하고 데이터의 보안성을 높이면서 활용성을 개선한 비식별조치 기술을 국내 정책에 반영할 필요가 있다. 해외 비식별조치 사례를 기반으로 비정형 데이터에 특수화된 비식별 조치를 검토해야 하며, 그에 따라 국내 비정형 데이터 비식별조치 기법을 수정해야 한다. 이를 통해 CCTV 영상 내 객체 정보를 최대한 비식별화할 수 있으며[27], 이를 통해 데이터 활용 폭도 넓힐 수 있다.

4.2 비정형 데이터에 대한 개인정보 판단기준 수립

현재 법률 및 가이드라인에 따르면, 개인정보의 개념이 상당히 포괄적으로 정의되어 있어 사전 검토 단계에서는 정형 데이터 및 비정형 데이터의 개인정보를 구분하는 기준이 모호하다. 또한, 비정형 데이터에 대한 구체적인 내용은 거의 언급되지 않아 비식별 대상을 파악하기 어렵다.

<표 6>은 현재 데이터의 개인정보 판단 기준에 대해 재가공한 것이다[30].

<표 6> 데이터 개인정보 판단 기준

| 구분 | 개인정보 | 가명정보 | 익명정보 |
|------|--|---|---------------|
| 정의 | 특정 개인을 식별할 수 있는 정보 또는 다른 정보와 쉽게 결합하여 식별할 수 있는 정보 | 추 가 적 인 정보를 결합하지 않고는 특정 개인을 식별할 수 없는 정보 | 개인 식별 불가능한 정보 |
| 적용 범 | 정보 주체의 | 특정 목적 | 제한 없이 자 |

| | | | |
|---|------------------|--------------------|--------|
| 위 | 사전 동의 범위 내에서만 가능 | (통계, 연구 등)으로 활용 가능 | 유롭게 사용 |
|---|------------------|--------------------|--------|

<표 6>에 따르면, 개인정보 판단의 범위는 개인 정보, 가명정보, 익명정보로 구분할 수 있다. 정형 데이터의 경우 주로 데이터베이스나, 텍스트 형태의 데이터로 이루어져 있어 비정형 데이터보다 가명처리 및 익명처리를 적용하기 수월하며, 개인정보 여부에 관한 판단을 내리기 쉽다. 그러나 비정형 데이터의 경우, 가이드라인 내부에서 식별정보 판단이 포괄적으로 명시되어 있어 비식별화를 거친 비정형 데이터가 개인정보인지 판단하기 어렵다. 그리고 실무적으로 적용할 수 있는 비정형 데이터의 비식별화 사례가 적어 실질적으로 비식별화를 수행하기 어려운 상황이다.

따라서, 비정형 데이터의 개인정보 포함 단계를 나누어 활용성에 제한을 두어야 하며, 추후 기술의 발전과 같은 환경적인 요소의 변화에 따라 적정성 평가 후 단계별로 적절하게 변경하여 활용할 수 있도록 해야 한다. 가령 의료 산업에서의 영상 정보 활용 단계를 3단계로 나눈다면, 최저 단계인 3단계에서는 의사, 환자 등과 같은 식별정보를 포함하는 데이터로 구성할 수 있으며, 이 경우 별도의 비식별조치가 필요 없다. 최고 단계 1단계는 의사의 행위, 환자의 이동 방향 등의 다양한 식별정보를 포함한 데이터로, 강화된 비식별화 조치를 거쳐야만 활용할 수 있다. 또한, 국가적으로 비정형 데이터 비식별 사례를 조사하고 이를 정형 데이터와 함께 가이드라인에 명확히 명시하여 실무에서 비정형 데이터의 비식별조치를 원활하게 적용할 수 있도록 해야 한다.

4.3 객체 특징 기반의 식별 ID 활용

최근 딥러닝 기술 기반의 CCTV 영상 처리 장치를 활용한 비정형 데이터 수집 및 객체 인식 기술이 발달하였다. 이에 따라 마케팅을 목적으로 보행자를 추적하는 리테일 공간의 혼잡도와 관심도를 분석하는 연구가 활성화되고 있다[31]. 국내에서는 이러한 기술을 이용해 이동 객체를 인식할 수 있지만, 객체의 이동 경로를 파악하기 위해서는 담당자의 수작업이 필

요하다[32]. 하지만 딥러닝 기술 기반의 CCTV 영상 처리 기술 및 비정형 데이터와 객체 인식 기술을 이용한 객체 추적 애플리케이션은 프라이버시 유출 문제로 인해 분석이 제한되거나 불가능한 상황이다.

따라서 영상 데이터에서 동작 및 신체 자세와 얼굴을 연관시켜 개인을 트래킹한 N. Anh Tu의 연구[27]와 같이, 의료 비정형 데이터에서 객체를 탐지한 후 객체에 식별할 수 있는 ID를 부여할 필요가 있다. 즉 얼굴 유출이 문제가 되는 CCTV 영상의 경우, 얼굴 정보를 사용하지 않고 동작 및 신체 자세와 얼굴을 연관지어 트래킹한다. 이후 지능형 사물에서 네트워크를 통해 단일 영상 및 음성 수집 기기에서 수집한 영상을 활용하는 것보다 더욱 정확하고, 방대한 양의 데이터를 확보할 수 있다.

<표 7>은 정형 데이터 비식별 기법인 데이터 범주화, 비정형 데이터 비식별 기법 중 하나인 인페인팅 기법, 식별 ID 기반의 비식별 기법의 특징을 비교한 것이다.

<표 7> 데이터별 비식별 기법 특징 비교

| 구분 | 데이터 범주화 [10] | 인페인팅 [22] | 식별 ID 기반의 비식별 기법 |
|----------|--------------|-------------------|------------------|
| 주요 적용 대상 | 정형 데이터 | 비정형 데이터 | 비정형 데이터 |
| 보안성 | 재식별 위험 존재 | 재식별 위험 적음 | 재식별 위험 적음 |
| 데이터 활용성 | 사용 가능 | 사용 어려움 | 사용 가능 |
| 사용 예시 | 나이, 지출 등 | CCTV에서의 객체 비식별화 등 | 객체 추적(이동 방향 등) |

<표 7>에 따르면 대표적인 정형 데이터 비식별 기법인 데이터 범주화는 텍스트 형식의 데이터에 맞춰진 비식별 기법으로 나이, 지출 등의 데이터에 활용된다. 이 기법을 사용하면 데이터의 활용성을 높일 수 있으나 재식별 위험이 존재한다. 반면 인페인팅 기법과 같은 비정형 데이터의 비식별 기술은 재식별 위험이 적어 개인정보를 차단하는데는 최적화되어 있으나, 비식별 후 데이터를 이용하는데 어려움이 있다. 이러한 데이터의 활용 한계는 비식별 정책 내 식별 ID 부여 프로세스를 도입함으로써 해결될 수 있다. 식별 ID 기반의 비식별 기법은 비정형 데이터를 타겟으로 하

며, 재식별 위험은 적다. 그러나 데이터를 다양하게 활용할 수 있다는 장점을 가져서 객체 추적 등에 사용 가능하다. 즉, 국내 비식별조치 정책 내에 식별 ID 부여 과정을 포함하여, 비정형 데이터 이용에 소극적이었던 의료계에서 비정형 데이터를 적극적으로 활용할 수 있도록 해야 한다.

또한, 비식별조치 과정을 더 상세히 나누어 실제 산업 현장에서의 활용도를 높여야 한다. 하나의 예로, 활용 목적에 따라 비식별 강도를 조절할 수 있으며, 그룹을 추적할 때 여러 식별 ID를 묶어 비식별 객체를 다음의 3단계 과정으로 관리할 수 있다. 객체 구분에 의미가 없거나 추적하는 객체에 대한 정보 제공을 원천 차단해야 하는 1단계에서는 객체를 동일하게 표시하고, 대상 그룹이 정해져 있는 2단계에서는 추적하는 표본 대상을 지정하여 표시할 수 있다. 병원 내 환자의 이동 경로를 추적할 때 활용될 수 있는 3단계에서는 환자별 특징, 성별, 나이 등의 세부 기준에 따라 컬러링하여 비식별 객체의 식별 ID 기술을 활용할 수 있다. 이를 통해 비정형 데이터의 데이터 특성을 훼손하지 않고 개인정보를 보호한 상태에서 데이터 이용이 가능할 것이다. CCTV 외에도 다양한 비정형 데이터에 제안하는 기술을 적용할 수 있다. 의료 판독 데이터도 3단계로 구분하여 관리할 수 있다. 많은 환자가 촬영하는 MRI와 같은 판독 데이터의 경우, 상대적으로 약한 비식별 기술을 활용하고, 소수의 중병 환자들이 촬영하여 환자의 신상이 재식별될 수 있는 판독 데이터의 경우, 강한 비식별 기술을 활용할 수 있다. 의료 판독 영상에서 질병 부위와 관련 없는 부위를 비식별할 수 있고, 지속적으로 촬영하는 의료 판독 데이터의 경우, 변경 사항 추적 시 객체를 컬러링함으로써 본 기술을 활용할 수 있다.

객체 특징 추출 기반의 식별 ID 부여 프로세스를 비식별조치 가이드라인에 명시하고, 정책에 반영하여 식별 ID 부여 이후 객체 비식별 방식에 세부 절차를 기재한다면, 의료 산업에서 비정형 데이터를 적극적으로 도입할 수 있는 실무적 기반이 마련될 것이다. 또한, 비식별화 단계에서 객체 특징 추출 기반의 식별 ID 부여 프로세스를 통해, 데이터 속성을 훼손하지 않고 기존 정형 데이터와 같이 비정형 데이터에서도 활용성과 개인정보 보호 측면을 복합적으로 고려할 수

있을 것이다.

즉, 의료 데이터 산업 활성화를 위해서는 비식별조치 정책에 비정형 데이터 내 개인정보에 대한 명확한 판단 기준을 세워서 정확한 비식별 대상을 파악하는 것이 필요하다. 그리고 비식별 가이드라인에 비정형 데이터에 최적화된 비식별 조치를 추가적으로 도입할 필요가 있다. 또한 식별 ID를 적용하여 데이터의 프라이버시가 고려된 트래킹 기술을 적극 활용해야 한다.

5. 결론

최근 의료 데이터 플랫폼의 등장과 네트워크 통신 기술의 발달로 인해 다양한 형태의 데이터가 기하급수적으로 증가하였다. 하지만 생성된 개인 보건의료 데이터는 개인정보 유출과 같은 부정적인 측면도 가지고 있다. 이를 해결하기 위해 가이드라인에서는 비식별화 기술을 제공하여 개인을 식별할 수 없도록 안내하고 있으나, 원본 복구 가능성이 있거나 데이터 가치를 훼손시킬 수 있다는 한계점을 갖고 있으며, 제공된 비식별화 기술은 정형 데이터에 초점을 맞추고 있는 경우가 많아 실무적인 비정형 데이터 비식별화에 어려움을 겪고 있다.

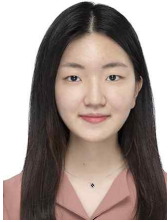
이에 본 논문에서는 현재 비식별조치 정책과 기술의 한계를 분석하고, 한계를 개선할 수 있는 선행연구를 조사하였다. 또한, 해당 기술을 활용하기 위한 정책을 함께 제안하였다. 제안하는 바와 같이 비식별조치 관련 정책이 개정된다면, 데이터의 익명성과 추적성, 데이터 활용성을 향상할 수 있으며, 의료 산업에서 지금까지 제한되었던 비정형 데이터의 활용 범위와 가치를 높일 수 있을 것이라 기대한다. 본 논문에서는 정책 자료와 기술 문헌을 분석하여 정책 개선의 필요성과 기술 도입의 타당성을 주장했으나 실험적 데이터나 통계 분석을 통한 실증을 수행하지 못했다. 후속 연구에서는 산업 현장 전문가 설문 조사를 통해 제안한 정책의 타당성을 검증하고 의료 데이터 비식별화 테스트 베드를 구축하여 비식별화 기능과 성능을 검증할 계획이다.

참고문헌

- [1] 김지은, 황정민, 홍영주, 김수경, "디지털 헬스 산업 분석 및 전망 연구", KHIDI, 2020. <https://www.khidi.or.kr/fileDownload?titleId=449375&fileId=1&fileDownType=C¶mMenuId=MENU02686> [Accessed: 2021-April-07].
- [2] 보건복지부, "보건 의료 데이터 · 인공지능 혁신전략" pp 1-47, Jun 2021. <https://eic.kdli.re.kr/policy/materialView.do?num=214544&topic=P&pp=20&datecount=&recommend=&pg=> [Accessed: 2021-Jun-03].
- [3] 신태섭, "보건의료 데이터 활용 가이드라인의 의미와 과제," 의료법학, 22(3), pp. 31-55, 2021. <https://www.kci.go.kr/kciportal/ci/sereArticleSearch/ciSereArtiView.kci?sereArticleSearchBean.artiId=ART002761333> [Accessed: 2022-Mar-03].
- [4] 보건복지부, "의료법," Dec. 2020. <https://www.law.go.kr/EB%B2%95%EB%A0%B9%EC%9D%98%EB%A3%8C%EB%B2%95> [Accessed: 27-Feb-2022].
- [5] K. Patel and G. B. Jethava, "Privacy Preserving Techniques for Big Data: A Survey," 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), pp. 194-199, Sep. 2018.
- [6] 하만석, 안현철, "정형 데이터와 비정형 데이터를 동시에 고려하는 기계학습 기반의 직업훈련 중도 탈락 예측 모형," 한국콘텐츠학회논문지, 19(1), pp. 1-15, Jan. 2019.
- [7] Grand View Research, "Preclinical Imaging Market Analysis By Product Type(Devices: CT, MRI, PET/SPECT, Multi-model, Optical, Ultrasound, Photoacoustic(PAT), Reagents and Services) And Segment Forecasts To 2024," pp. 1-87, Jun. 2016. <https://www.grandviewresearch.com/industry-analysis/preclinical-imaging-market> [Accessed: 27-Feb-2022].
- [8] 박민영, 최민경, "의료정보의 관리와 비식별화에 관한 법적 과제," 유럽헌법학회, (21), pp. 495-534, Aug. 2016.
- [9] 김상광, 김선경, "빅데이터 활용에 영향을 미치는 개인정보 규제요인과 데이터 결합요인의 탐색," 정보보호학회논문지, 30(2), pp. 287-304, Apr. 2020.
- [10] 한국인터넷진흥원, "개인정보 비식별 조치 가이드라인," pp. 1-80, Jun. 2016. https://www.privacy.go.kr/cmm/fms/FileDown.do?atchFileId=FILE_000000000827059&fileSn=0 [Accessed: 27-Feb-2022].
- [11] V. Bhagwan, T. Grandison, and C. Maltzahn, "Recommendation-Based De-identification: A Practical Systems Approach towards De-identification," 2012 IEEE Eighth World Congress on Services, pp. 155-162, Aug. 2012.
- [12] 전승재, 권현영 "개인정보, 가명정보, 익명정보에 관한4개국 법제 비교분석," 정보법학, 22(3), pp. 183-218, Dec. 2018.
- [13] 손형섭, "한국 개인정보보호법과 일본 개인정보보호법의 비교 분석 - ICT산업 생태계에 미치는 영향을 중심으로 -," 2019 NAVER Privacy White Paper. https://privacy.naver.com/download/2019_chapter2_JapanPersonalInformationProtectionAct.pdf [Accessed: 27-Feb-2022].
- [14] 개인정보 보호위원회, "개인정보의 범위에 관한 연구," pp. 1-335, Oct. 2014. http://pipc.go.kr/cmm/fms/FileDown.do?jsessionId=AFF7450287C3709AB913DD347E4355F5?atchFileId=FILE_00000000499259&fileSn=0 [Accessed: 27-Feb-2022].
- [15] ISO/IEC, "Privacy enhancing data de-identification terminology and classification of techniques," ISO/IEC 20889, pp. 1-46, Nov. 2018. <https://www.iso.org/standard/69373.html> [Accessed: 27-Feb-2022].
- [16] 개인정보 보호위원회, "가명정보 처리 가이드라인," pp. 1-56, Sep. 2020. <https://www.pipc.go.kr/np/cop/bbs/selectBoardArticle.do?bbsId=BS217&mCode=D010030000&nttId=8000> [Accessed: 02-May-2022].
- [17] 한국인터넷진흥원, "신규 ICT 환경의 해외 개인정보 활용사례 및 법제동향 분석연구," pp. 1-214, Oct. 2016. <https://www.kisa.or.kr/204/form?p>

- ostSeq=022686&lang_type=KO&page= [Accessed: 29-Dec-2020].
- [18] S. Garfinkel, "De-Identification of Personal Information," NISTIR 8053, pp. 1 - 46, Oct. 2015.
- [19] 이인경, 엄수현, 아로샤, 자파르, 이우기, "경로정보 개인 비식별화를 위한 K-익명성의 K값에 따른 데이터 활용과 익명성의 상충관계," 한국정보과학회 학술발표논문집, pp. 1241-1242, Jun. 2018.
- [20] 이대회, "개인정보 보호 및 활용 방안으로서의 가명·비식별정보 개념의 연구," 정보법학, 21(3), pp. 217-251, Dec. 2017.
- [21] Department of Health & Human Services, "Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule," Feb. 2012. <https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html> [Accessed:2-Feb-2022].
- [22] 한국정보화진흥원, "영상 데이터 익명화 기술 및 평가방안", pp. 1-17, Feb. 2019. https://kbig.kr/board/fileMngr?cmd=down&boardId=files_bigdata_report&bltnNo=11551661790597&fileSeq=1&subId=sub06 [Accessed: 27-Feb-2022].
- [23] 이경률, 임강빈, 이선영, "영상보안시스템에서의 형태보존암호를 활용한 효율적인 프라이버시 마스킹 기술 연구," 한국통신학회 학술대회논문집, pp. 813-814, Feb. 2018.
- [24] B. Meden, Z. Emersic, V. Struc, and P. Peer, "k-Same-Net: Neural-Network-Based Face Deidentification," 2017 International Conference and Workshop on Bioinspired Intelligence (IWOBI), pp. 1-7, Jul. 2017.
- [25] Shahid, Arsalan, et al. "A Two-Stage De-Identification Process for Privacy-Preserving Medical Image Analysis." Healthcare. Vol. 10. No. 5. Multidisciplinary Digital Publishing Institute, 2022.
- [26] Popescu, Andreea Bianca, et al. "Obfuscation Algorithm for Privacy-Preserving Deep Learning-Based Medical Image Analysis." Applied Sciences 12,8 (2022): 3997.
- [27] Tu, N.A., Huynh-The, T., Wong, KS, et al., "Toward efficient and intelligent video analytics with visual privacy protection for large-scale surveillance," J Supercomput 77, pp. 14374 - 14404, May. 2021.
- [28] 김순석, 황호성, 윤상진, 권오승, 인한진, 김범식. (2018).개인정보보호를 위한 비식별 조치 현안 및 향후 과제에 관한 연구-보건의료분야를 중심으로. 예술인문사회 융합 멀티미디어 논문지,8(5),599-607.
- [29] 박민영, 최민경.(2016). 의료정보의 관리와 비식별화에 관한 법적 과제.유럽헌법연구,(21),495-534.
- [30] Personal Information Protection Commission Act No. 16930, "Personal Information Protection Act," Personal Information Protection Policy Division, Feb. 2020.
- [31] 김종율, 김혁중, "리테일 마케팅 고도화를 위한 C CTV 영상 데이터 기반의 AI 융합 응용 서비스 활용 모델 연구," 디지털융복합연구, 19(5), pp. 197-205, 2021.
- [32] 박예승, 김도영, 최태림, 허수웅, 송혜원, 이상훈, "위험 상황 자동 인지 지능형 CCTV 시스템 연구," 한국통신학회 학술대회논문집, pp. 1174-1175, Feb. 2020.

— [저자 소개] —



이 선 진 (Sun-Jin Lee)
2021년 8월 성신여자대학교 융합보안
공학 학사
2021년 9월~ 현재 성신여자대학교 미
래융합기술공학 석사과정
email : 220214013@sungshin.ac.kr



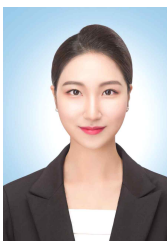
오 영 은 (Young-Eun Oh)
2020년 8월 성신여자대학교 융합보안
학 학사
2021년 1월~ 2021년 5월 딜로이트 안
진회계법인 리스크자문본부 정보보안
컨설턴트
2021년 5월~ 현재 라인비즈플러스
viz security
email : lauren5kr@gmail.com



박 태 립 (Tae-Rim Park)
2021년 8월 성신여자대학교 융합보안
공학 학사
2021년 9월~ 현재 성신여자대학교 미
래융합기술공학 석사과정
email : parktaerim2@gmail.com



이 일 구 (Il-Gu Lee)
2003년 2월 서강대학교 전자공학 학
사
2005년 2월 KAIST 정보통신 석사
2016년 2월 KAIST 전산학부 박사
2005년 2월~ 2017년 2월 한국전자통
신연구원 선임연구원
2017년 3월~ 현재 성신여자대학교 융
합보안공학과/미래융합기술공학과 조
교수
email : iglee@sungshin.ac.kr



김 소 희 (So-Hui Kim)
2020년 2월 성신여자대학교 융합보안
학 학사
2020년 4월~ 현재 신세계아이앤씨 정
보보안팀
email : sophiekim96@gmail.com