

청각장애 부모를 위한 딥러닝 기반 다중모달 육아보조시스템의 설계 및 구현

김정태 · 김경도[†]

홍익대학교 산업데이터공학과

Design and Implementation of a Deep Learning - Based Multimodal Childcare Support System for Deaf Parents

Jeong-Tae Kim · Kyungdoh Kim

Department of Industrial & Data Engineering, Hongik University

Recent advances in machine learning and Internet of Things (IoT) technologies have led to the development of childcare support services that monitor infants' conditions in real time and provide immediate feedback. However, most existing services are designed for non-disabled users and rely heavily on auditory cues, which creates significant limitations for deaf parents who have difficulty perceiving sound-based situations. Although several prior studies have proposed childcare support systems for deaf parents, technical limitations that fail to adequately reflect real-life environments remain. To address these issues, this study designs and implements a deep learning - based multimodal childcare support system. The proposed system consists of three core functions: infant cry detection and cause classification, hazardous sound detection, and infant suffocation risk detection. An event-driven, hardware-based notification framework is employed to deliver alerts to a smartwatch. System demonstrations confirm stable real-time operation of all functions and successful alert transmission when hazardous situations occur.

Keywords: Hearing Impairment, Deep Learning, Childcare, Cry Detection, Sound Detection, Suffocation, Smartwatch, Infant, IoT

1. 서 론

최근 머신러닝과 사물인터넷의 발전으로 영아의 상태를 실시간으로 모니터링하고 즉각적인 피드백을 제공하는 다양한 육아보조서비스가 등장하고 있다. 해당 서비스들은 영상, 음향, 센서 데이터 등을 통해 영아의 움직임, 울음, 수면 상태 등을 분석하여 육아에 도움이 될 수 있는 여러 정보들을 보호자에게 제공한다. 특히 실시간 모니터링과 자동화된 상태 판단 기능은 영아의 안전을 확보하는 데 중요한 역할을 수행하여 보호자의 육아 부담을 완화하는 데 크게 기여하고 있다. 그러나 현재 상용화된 육아보조서비스의 대부분은 주요 타겟 사용자

를 비장애인 보호자로 설정하고 있으며, 청각 정보를 기반으로 한 알림과 상황 인식 방식에 크게 의존하고 있다.

2023년 보건복지부가 조사한 장애인 실태 조사에 의하면 청각장애인 중 69.3%가 자신과 동일한 장애를 가진 배우자와 혼인했고, 이 중 85.2%는 배우자의 장애 정도가 심한 것으로 나타나 상호보완적 역할 수행에 어려움이 있음을 시사했다(Ministry of Health and Welfare, 2023). 이러한 상황을 바탕으로 청각장애인 부부의 육아 시 발생 문제를 청각장애인 부부 5쌍의 생활을 정리한 연구(Lee, 2015)와 4쌍의 청각장애인 부부의 인터뷰 영상에서 도출했다(SohaengseongChackbang, 2020; BogeonbokjibuTV, 2022; Yoosonsaeng, 2021; HankyorehTV,

[†] 연락저자 : 김경도 교수, 04066 서울특별시 마포구 와우산로 94 홍익대학교 산업·데이터공학과, Tel : 02-320-1627, Fax : 02-336-1130,
E-mail : kyungdoh.kim@hongik.ac.kr

2026년 1월 13일 접수; 2026년 2월 21일 수정본 접수; 2026년 2월 13일 게재 확정.

2019). ‘본인의 목소리로 태교를 못해서 아쉬움’, ‘듣지 못해서 밤에 아이가 우는지 계속 확인해야 함’, ‘자녀가 우는 것을 느끼기 위해 몸을 밀착시켜 안고 자거나 팔에 끈을 묶어서 자야 함’, ‘아기가 침대에서 떨어졌으나 몰랐음’, ‘아이가 왜 우는지 판단이 안됐음’, ‘추후 수어 교육에 대한 걱정이 있음’ 등과 같은 우려가 확인됐다. 비장애인이 육아 시 가지고 있는 어려움에 더해 청각장애인으로서 가지고 있는 장애로 인한 추가적인 어려움을 가지는 것이다.

그럼에도 불구하고, 현재 청각장애인 맞춤형 육아보조서비스는 전무한 실정이다. 현존하는 육아보조서비스의 경우 사용자의 청력 소실을 고려해서 설계되지 않았기 때문에 사용이 어려운 경우가 대부분이고, 이용을 하더라도 단편적인 정보 수집 등의 일부 기능만 사용 가능한 상황이다. 결과적으로 청각장애 부모가 육아과정에서 도움이 되는 다양한 정보를 수집할 수 있는 청각장애 부모 맞춤형 육아보조서비스가 필요하다.

본 논문에서는 청각장애인 부모를 위한 딥러닝 기반의 다중모달 육아보조시스템을 정의 및 설계하고, 이를 직접 구현하는 방법을 제시한다. 청각장애인 부모의 육아 고충(Pain Point)을 출산 전, 영아기, 영아기 이후 세 시기로 나누어 분석했다. 모든 시기를 아우르는 시스템 설계의 한계를 고려해, 전체 고충의 64.7%로 고충이 가장 많이 집중됐던 영아 시기에 초점을 맞춰 연구 범위를 좁혔다. 영아기 고충의 주요 원인은 ‘아이에게서 눈을 뗄 수 없다’와 ‘아이의 소리를 듣지 못한다’는 두 가지였다. 따라서 본 연구는 이러한 영아기 육아 고충의 핵심 원인들을 기술적으로 해결하기 위한 시스템을 정의 및 설계하는 것으로 연구의 목적을 설정했다. 이를 통해 청각장애인 부모가 아이의 안전을 능동적으로 확보하고, 육아과정에서 심리적 안정감을 느끼며 독립적으로 육아를 할 수 있도록 도울 수 있는 육아보조서비스의 토대를 마련할 수 있다.

이러한 연구 목적을 구체화하고 타당성을 확보하기 위해, 본 논문에서는 먼저 현존하는 육아보조서비스의 기술적 동향을 살펴보고자 한다. 이어서 청각장애 부모를 대상으로 제안된 육아보조시스템 관련 선행연구들을 분석함으로써, 기존 연구들이 공통적으로 제공하는 기능과 그 한계점을 체계적으로 정리한다. 이를 통해 본 연구가 해결하고자 하는 연구 공백을 명확히 하고, 제안 시스템의 필요성을 논증한다.

2. 연구배경 및 관련 연구

2.1 육아보조서비스 동향

지난 20년간 여성의 고등 교육 및 유급 노동 참여 증가로 인한 부모 역할의 변화와 과학적 육아 자료의 보편화 등의 환경의 변화에 따라 부모의 육아보조서비스 사용량이 증가했다. 연구자 및 앱 개발자들은 전반적인 영유아 관리, 수유/영양, 성장 및 예방접종 추적, 영아 돌연사 증후군(SIDS) 예방을 주요 관리 요구 사항으로 두고 연구 및 서비스를 제안했다(Virani et

al., 2021). 특히 영아 모니터링 기술의 경우 2015년 이후 클라우드 컴퓨팅, 인공지능(AI), 사물인터넷(IoT)의 통합으로 부모에게 더욱 유연한 육아 환경을 제공하고 있다(Chandnani et al., 2025). IoT를 융합한 육아보조서비스는 영아의 돌봄 영역과 기능적 관점을 고려하여 크게 다음의 6가지 유형으로 구분할 수 있으며, 센서 기반의 모니터링이 다양한 형태로 제안되고 있는 특징을 가진다(Lee et al., 2017).

이는 영아 돌봄의 목적과 제공 기능을 기준으로 한 분류이다.

- 수면지원: 아이의 수면상태를 확인 후 부모의 스마트폰으로 전송하면 부모가 적절한 수면환경을 조성함.
- 베이비모니터: 카메라 및 웨어러블 기기를 활용하여 아이 상태, 체온, 생리적 변화 등을 전송함.
- 일상적 돌봄(배변과 수유) 지원: 스마트 기저귀가 대표적인 예시이며, 수유의 양, 대소변 양 등을 기록하여 영아의 건강상태를 확인함.
- 발달(의사소통) 지원: 양육자가 영아에게 얼마나 많은 양의 말을 하는지 측정하여 영아의 언어발달을 도움.
- 건강관리(체온측정): 무선 접착 패치를 영아의 피부에 부착하여 지속적인 체온 측정을 함.
- 종합형: 수면지원, 베이비모니터, 일상적 돌봄 등의 종합적인 육아보조서비스를 하나의 앱에서 관리함.

최근에는 아기의 생체 징후를 측정하거나 위험한 상황을 감지하여 생명을 위협하는 사고를 예방하는 아기건강 모니터링 시스템(BHMS)(Baby Health Monitoring System)에 머신러닝 등의 기술적 통합으로 더 복잡한 문제를 해결하려는 노력이 이어지고 있다. 기존의 IoT기반의 육아보조서비스는 카메라, 마이크, 맥박, 습도 센서등을 활용하여 수집한 데이터를 지표로 영아 상태를 확인하는 방식에 집중되어 왔다. 그러나 다양한 센서들을 통해 수집한 데이터와 합성곱 신경망(CNN), Support Vector Machine(SVM)등의 머신러닝을 접목한 결과, SIDS의 높은 위험 요인으로 간주되는 수면자세, 호흡률 및 체온을 추적하는 것이 가능해졌다(Khan et al., 2024). Pavel et al.(2020)는 정확한 진단이 어려운 신생아 발작 분야에서 머신러닝 알고리즘인 ANSeR을 활용하여 신생아 발작 발생 시간대 파악을 개선했다.

현재 국내에서도 영아의 부모들 사이에서 다양한 육아보조서비스가 활용되고 있다. 산후조리원 신생아 라이브 영상 서비스인 ‘아이보리 베베캠’의 경우 해당 서비스가 입점한 산후조리원의 영아를 정해진 시간에 상태 확인을 할 수 있도록 서비스를 제공하고 있다. 또한 영아 울음 분석 서비스인 ‘크라이베베’의 경우 아이가 우는 경우에 그 원인이 무엇인지 분석하여 주는 스마트 울음분석 서비스를 제공하고 있다. 해당 서비스들은 아이의 상태 확인을 돕는 도구로서 강점을 가졌다.

그러나 기존 서비스들은 대부분 비장애인을 주 사용자로 상정하고 있어, 지속적인 시각적 모니터링을 요구하거나 청각 정보 위주의 작동 기반을 채택하고 있다. 이러한 설계 방식은

청각장애 부모에게 명확한 사용상의 제약을 주며, 결과적으로 이들의 특수성을 고려한 실질적인 육아보조솔루션이 부재함을 시사한다.

2.2 청각장애 부모를 위한 육아보조시스템 연구

청각장애 부모들이 육아과정에서 가지는 어려움을 해결하기 위해 선행된 육아보조기술 연구들이 제안되어 왔다(Park *et al.*, 2015; Park *et al.*, 2024; Jeon *et al.*, 2021). Park *et al.*(2015)은 아두이노를 아이가 있는 방에 설치하여 아이가 울었을 때 울음의 원인을 데시벨의 크기에 따라 안드로이드 기반의 애플리케이션을 통해 청각장애 부모에게 알림을 주는 울음감지 시스템을 구현했다. Park *et al.*(2024)은 아이의 울음소리와 얼굴을 데시벨 센서와 카메라를 통해 인식하여 청각 장애를 가진 부모가 아이의 울음, 추락, 질식에 대응할 수 있도록 웨어러블 기기를 통해 LED와 진동으로 알리는 시스템을 구축했다. Jeon *et al.*(2021)은 SIDS의 가장 높은 비중을 차지하는 질식사에 대해 청각장애 부부가 대응할 수 있도록 침대에 카메라를 거치하여 낙상, 울음소리 및 원인, 질식사 위험을 감지 시 사용자에게 웨어러블 기기를 통해 알린다.

해당 연구들이 제안하는 공통적인 제공 기술과 기술한계점이 존재했다. 공통적으로 울음을 감지하는 기능, 낙상 및 질식사 위험 감지 기능을 제공하고 있다. 그러나 각각의 기능을 구현하는 방식에서 시스템을 실시간 환경에서 사용하기 어려운 5가지의 기술한계점들이 존재했다.

이는 선행연구들의 기술한계점이다.

- 데시벨 기반의 울음 원인 분류: 울음의 원인을 데시벨 기반으로 분류하는 것은 하드웨어의 설치 위치, 신호 대 잡음비(SNR)가 낮은 환경, 비정상 소음에 따라 성능 저하가 발생하기 때문에 큰 문제가 된다(Ramirez *et al.*, 2007).
- 배경 소음이 섞이지 않은 울음데이터 사용: 실제 생활 환경에서는 다양한 배경 소음이 존재하기 때문에 배경 소음이 없는 데이터로 학습된 모델은 실제 환경에서 성능이 급격히 저하된다(Yao *et al.*, 2022).
- OpenCV의 Haar Cascade를 활용하여 영아 질식사 위험 감지: Viola - Jones 알고리즘 기반의 Haar Cascade 얼굴 검출 기법은 조명변화에 취약하고, 정면 얼굴(frontal face)에 최적화되어 있어 얼굴 자세 변화에 취약하다는 한계성을 가지고 있다(Abdulhussien *et al.*, 2022). 이는 야간 상황에서의 호흡기 기반의 질식 위험 감지나 자세 기반의 질식 위험 감지에 매우 취약한 부분이다.
- 주변 환경 판단 기술 부재: 청각장애 부모는 청력 소실로 인해 소음, 경보 등의 주변 환경 상황을 파악하기 어렵다. 실제로 2020년 6월 청각장애인이 화재경보기 소리를 듣지 못해서 숨지는 사건이 있었다(ChungcheongToday, 2020). 또한 시각표시 기능이 있는 주택용 화재경보기가 보급됐으나 보지 못할 가능성이 있기 때문에 스마트폰과

연동 가능한 화재경보기, 가스경보기 소리 감지 기능이 추가적으로 필요하다(LG HelloVision Chungnam, 2022).

- 별도의 웨어러블 기기를 사용: 육아보조서비스를 통해 알림을 받기 위해서 별도로 만든 과도한 크기의 웨어러블 기기를 착용해야 한다. 그러나 부가적인 기능이 없이 알림의 기능만 있기 때문에 사용자의 편의성 및 실용성을 위해서 스마트워치로 알림을 보내는 방식을 적용하는 것이 더 나은 선택지다.

2.3 연구 방향

결과적으로 청각장애 부모를 위한 육아보조시스템은 단일 센서나 단일 신호에 의존하는 방식이 아니라, 서로 다른 특성을 가진 다양한 신호를 동시에 고려할 수 있는 설계가 요구된다. 특히 소리의 크기뿐 아니라 주파수 기반의 음향 특징으로 더 높은 성능의 분류 정확도를 가질 필요가 있으며(Hammoud *et al.*, 2024), 얼굴 전체가 아닌 호흡기를 통한 시각적 특징, 그리고 영아 외부의 주변 환경 위험 신호를 함께 분석할 수 있어야 한다. 이러한 요구사항을 충족하기 위해서는 복잡하고 이질적인 데이터를 효과적으로 통합·해석할 수 있는 고차원적 패턴 인식 기술이 필요하다. 딥러닝 기반 다중모달 시스템은 음향, 이미지 등 서로 다른 형태의 입력 데이터를 각각의 특성에 맞게 학습하고 이를 통합적으로 분석할 수 있는 구조를 제공한다. 이러한 접근은 기존의 규칙 기반 또는 단일 신호 기반 시스템이 가지는 한계를 극복하고, 실제 가정 환경과 같은 복합적인 상황에서도 보다 정밀한 위험 감지와 상황 판단을 가능하게 한다. 이에 본 연구에서는 선행연구에서 도출된 기술적 한계들을 해결하기 위해, 청각장애 부모를 위한 딥러닝 기반 다중모달 육아보조시스템을 설계하고 구현하고자 한다.

3. 제안 방법론

3.1 청각장애 부모를 위한 육아보조시스템의 핵심 기능 정의

먼저, 제2장의 내용에 기반하여 본 연구에서의 육아보조시스템의 설계를 위한 3가지 핵심 기능을 설정하였다. 각 기능들은 머신러닝 모델 및 딥러닝 모델로 구현했으며, 시스템은 영아의 상태를 모니터링하는 하드웨어와 알림을 전달받는 스마트워치로 구성된다.

- 영아 울음 감지 및 원인 분류 기능 정의

청각장애인 부모 다수가 인터뷰에서 아기 울음을 인지하지 못해 양육 과정에서 불편을 겪는다고 보고했다(SohaengseongChaekbang, 2020; Bogeonbokjibu TV, 2022; Yoosonsaeng, 2021; HankyorehTV, 2019). 때문에 아이가 울게 되면 울음소리를 감지하여 울음의 원인을 Dunstan Baby

Language에 기반하는 배고픔, 트림, 복통, 불편함, 피곤함의 5 가지 기준으로 분류하여 알람을 스마트워치로 전송한다. 단, 아이가 울 때마다 바로 반응하는 것은 아이가 스스로 잠드는 것을 오히려 어렵게 하기 때문에(Gradisar *et al.*, 2016) 아이가 우는 원인이 ‘피곤함’인 경우에는 점진적 소거법(울음에 바로 반응하지 않고, 기다리는 시간을 점차 늘림)을 적용하여 일정 시간 후 알람을 준다.

• 위험 소음 감지 기능 정의

주변 환경 판단 기술이 부재했던 선행연구의 한계를 해결하기 위해 가정 내에 화재나 가스 누출과 같은 긴급 상황이 발생하게 됐을 때 화재경보기, 가스누출기음을 감지하여 사용자에게 알린다.

• 영아 질식사 위험 감지 기능 정의

영아가 명확한 원인 없이 잠을 자다가 갑자기 사망하는 현상인 SIDS의 경우 정확한 원인이 밝혀지지는 않았지만 옆으로 누운 자세, 아이의 호흡기 덮음 등의 호흡 관련 요인으로 인해 발생하는 경우가 대부분이다(Duncan *et al.*, 2018). 따라서 아이가 수면 중 질식사하는 것을 방지하기 위해 영아의 호흡기인 코와 입을 일정 간격으로 탐지하여 호흡기가 감지되지 않으면 알람을 주도록 한다.

본 연구에서 사용한 핵심 기능들의 흐름은 <Figure 1>과 같

다. 각 기능을 구현한 모델들이 하드웨어에서 특정 사건을 감지하게 되면 Firebase로 사건을 저장하게 되며, 어떤 알람인가에 따라 알맞은 푸시알람이 스마트워치로 사용자에게 전달된다. 추가적으로 영아 울음 감지 및 원인 분류 모델의 경우 점진적 소거법을 적용시켜 작동한다. 울음의 원인이 ‘피곤함’으로 분류된다면 우선 수면 교육 알람을 스마트워치로 발송하고, 기준 시간인 3분, 5분, 10분, 15분마다 울음이 여전히 발생하고 있는지를 확인하여 울음이 여전히 감지된다면 해당 사실을 알람을 통해 전송한다. 그리고 울음이 더이상 감지되지 않는다면 다시 수면 교육 형식을 벗어나 울음을 감지하고, 울음의 원인을 분류하는 방식으로 작동한다.

본 연구의 시스템의 구조는 <Figure 2>와 같다. 하드웨어는 음향 감지 목적의 마이크와 객체 감지 목적의 적외선 카메라를 탑재한 라즈베리 파이5에 시스템의 기능을 담당할 3가지 모델이 로드되는 구조다. 그리고 하드웨어에서 전송한 감지 사건을 NoSQL클라우드 데이터베이스인 Firebase에 저장하게 되며, FCM(Fire Cloud Messaging)을 통해 스마트폰앱에 푸시알람을 보내는 방법이 사용된다. 스마트폰앱에 전송된 푸시알람은 연동된 스마트워치로 보내지게 되는 것으로 시스템이 마무리된다. 스마트폰앱의 경우 안드로이드 기반으로 개발했으며, 알람을 전달하는 것 외의 별도의 기능은 없다.

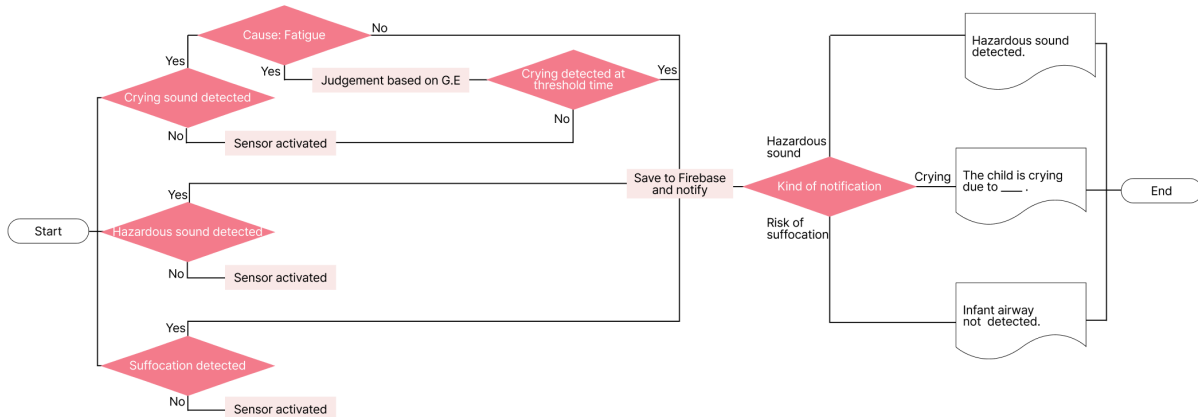


Figure 1. Service Flowchart Illustrating the Functional Structure of the Childcare Support System.



Figure 2. System Flowchart Visualizing the Overall System Architecture.

3.2 핵심 기능 모델 구현

(1) 데이터 수집 및 전처리

• 영아 울음 감지 및 원인 분류 모델

donate-a-cry-corpus-features-dataset(Valani, 2023)가 사용됐고, 해당 데이터는 울음데이터를 5가지 울음 원인으로 라벨링한 데이터로 총 457개(배고픔(382), 피곤함(24), 복통(16), 불편함(27), 트림(8))의 데이터가 포함되어 있다. 그리고 구글이 제공하는 대화, 기침, 음악 등의 여러 도시 소음을 모은 AudioSet 데이터(Gemmeke *et al.*, 2017)를 비울음 데이터로 사용하기 위해 사용했다. 데이터의 불균형 해소와 모델의 일반화 성능 향상을 목적으로 데이터 증강을 진행했다. 데이터 증강 방법은 총 5가지를 사용했고 데이터의 다양성을 위해 2-4개의 방법을 무작위 적용했다(Hammoud *et al.*, 2024). 첫 번째는 모델이 울음소리의 시작 위치에 덜 민감하도록 오디오 신호를 무작위로 좌우로 이동시키는 시간 이동(Time Shift)이다. 두 번째는 울음 지속 시간이나 재생 속도 변화에 대한 모델의 강건성을 높이기 위해 오디오 재생 속도를 무작위로 변경하는 시간 늘리기(Time Stretch)를 적용했다. 세 번째는 다양한 음색에 모델이 잘 적응하도록 오디오의 높이를 높이거나 낮추는 피치 변경(Pitch Shift) 방법이다. 네 번째는 모델의 노이즈 강건성을 향상시켜 배경 소음 속에서도 울음소리를 정확하게 구별할 수 있게 아주 낮은 수준의 백색 잡음(Gaussian Noise)을 추가하는 노이즈 추가(Add Noise)를 적용했다. 마지막으로 Audioset 데이터에서 19000여개의 생활소음 데이터 중 일부를 무작위의 배경음으로 섞는 배경 노이즈 추가(Add Background Noise) 기법을 적용하여 총 4612개로 데이터를 증강했다.

최종적으로는 스펙트로그램을 통한 시각화로 원본 데이터를 훼손하지 않는 선에서 데이터가 적절하게 증강됐음을 검증했다. <Figure 3>에서 원본 데이터 스펙트로그램인 (a)에 시간 이동, 피치 변경, 노이즈 추가 방법을 적용한 증강 오디오 스펙트로그램 (b)를 확인해보더라도 원본 데이터의 특성을 그대로 잘 반영하고 있음을 확인할 수 있다.

• 위험 소음 감지 모델

총 두 가지 데이터가 사용됐고, 첫 번째는 직접 녹음을 진행한 한국의 화재경보기, 가스누출기 경보음 데이터다. 한국의 경보음이 외국과 달라 수집을 진행했으며, 19개의 화재경보기

데이터와 5개의 가스누출기 데이터로 구성된다. 다른 하나는 영아 울음 감지 및 원인 분류 모델과 동일하게 비경보음 데이터로 사용하기 위해 AudioSet 데이터를 사용했다. 모델 학습에 있어 데이터 수가 매우 적고, 모델의 일반화 성능 향상을 위해 데이터 증강을 진행했다(Ramirez *et al.*, 2022). 우선적으로 학습모델로 선정한 YAMNet 모델에 사용할 수 있도록 .wav 포맷으로 형식을 변경하고, 16kHz로 주파수 변경 후 0.96초 클립으로 전처리하여 증강을 진행했다. 첫 번째는 모델이 경보음의 재생 속도에 덜 민감하도록 오디오 신호의 속도를 무작위로 변경시키는 시간 늘리기(Time Stretch)다. 두 번째는 다양한 음색에 모델이 잘 적응하도록 오디오의 높이를 높이거나 낮추는 피치 변경(Pitch Shift) 방법이다. 세 번째는 모델의 노이즈 강건성을 향상시켜 배경 소음 속에서도 울음소리를 정확하게 구별할 수 있게 아주 낮은 수준의 백색 잡음(Gaussian Noise)을 추가하는 노이즈 추가(Add Noise)를 적용했다. 네 번째는 경보음 소리의 크기에 대한 모델의 강건성을 높이기 위해 음의 크기를 무작위로 변경하는 음량 조절(Volume Adjustment)를 적용했다. 그리고 마지막으로 배경 소음이 있거나 없는 다양한 환경에서 모델이 잘 작동하도록 원본 오디오에 배경 소음(예: 대화, 음악)을 무작위로 섞어주는 배경 노이즈 추가(Add Background Noise) 기법을 증강 데이터의 약 50%에 적용했다. 약 50%로 섞은 이유는 배경 소음이 없는 경우에도 경보음이 울릴 수 있기 때문이다. 총 2400개의 데이터로 증강했고, 이 중 48.62%의 데이터가 배경데이터를 포함한다. 마찬가지로 스펙트로그램을 통한 시각화로 원본 데이터를 훼손하지 않는 선에서 데이터가 적절하게 증강됐음을 검증했다.

• 영아 질식사 위험 감지 모델

Simultaneously-collected Multimodal Mannequin Lying Pose (SMaL) 가 사용됐고, 해당 데이터는 담요 등으로 가려진 상태에서의 자세 추정을 목적으로 생성한 데이터셋이다(Kyrollos *et al.*, 2022). 자세를 바꿀 수 있는 마네킹을 이용하여 총 300개의 고유 자세를 포함하고, 각 자세마다 담요를 덮지 않은 상태, 얇은 담요를 덮은 상태, 두꺼운 담요를 덮은 상태의 3가지 케이스가 존재하며, 14개의 관절 위치 좌표가 함께 제공되는 것이 특징이다. 우선적으로 영아 질식사 모델의 경우 하드웨어에 로드된 이후부터는 어두운 환경에서 촬영된 이미지 데이터를 활용해야 하기 때문에 SMaL 학습 데이터를 하드웨어 수집

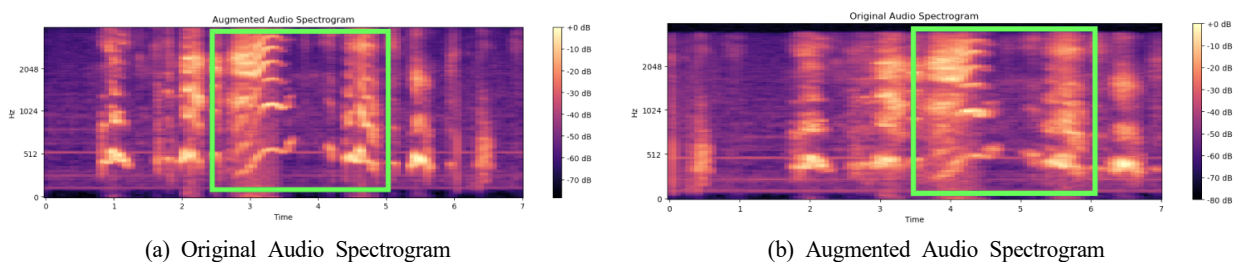
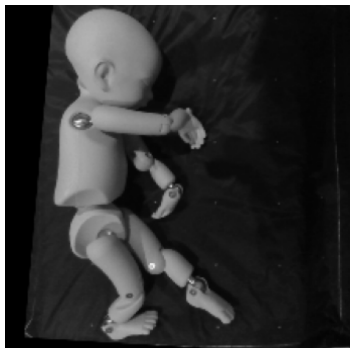


Figure 3. Comparison of Spectrograms between Original Infant Crying Data and Augmented Data Using Augmentation Methods Time Shift, Pitch Shift, and Add Noise.



(a) Original Input Image



(b) Grayscale-Converted Image



(c) Real-Time Acquisition Image

Figure 4. Examples of Grayscale Application to the SMaL Dataset.

데이터와 유사하게 만들 필요가 있었다.

따라서 <Figure 4>와 같이 원본 데이터의 RGB값에 (0.2989, 0.5870, 0.1140)의 가중치를 가진 Grayscale을 적용하여 하드웨어로 수집한 데이터인 (c)와 비슷하게 (a)를 (b)와 같이 만들었다. 또한 학습 모델로 선정한 YOLOv8n모델의 경우 객체 탐지 모델이기 때문에 바운딩 박스 라벨링(Bounding box annotation)이 필요하여, 학습 데이터의 ‘코’, ‘입’ 부분을 Roboflow를 활용하여 라벨링을 했다. 마지막으로 데이터 부족과 모델의 강건성을 향상시키기 위해서 총 4가지 방법을 사용하여 데이터 증강을 했다. 먼저 영아가 수면 시 자세가 변할 수 있기 때문에 자세에 대한 강건성을 높이기 위해 수평으로 데이터를 뒤집는 Flip과 -30도에서 30도에 해당하는 범위만큼 데이터를 회전시키는 Rotation을 무작위로 적용했다. 그리고 실시간으로 수집한 데이터의 화질저하에 대한 강건성을 높이기 위해서 0.6px까지 이미지의 선명도를 저하시키는 Blur와 -25%에서 25%에 해당하는 범위만큼 밝기 조절을 하는 Brightness를 무작위로 적용했고, 최종적으로 900개의 데이터에서 2160개의 데이터로 증강했다.

(2) 모델링(Modeling)

• 영아 울음 감지 및 원인 분류 모델

영아 울음 감지 및 원인 분류 모델의 경우 울음/비울음을 분류하는 Task와 울음의 원인을 분류하는 Task를 모두 수행해야 한다. 그래서 한 모델이 공통 피처를 학습하여 동시에 여러 태스크를 수행하는 Multi-Task Learning방법과 각 task별 독립된 피처를 학습하여 두 모델이 각각의 Task를 독립적으로 수행하는 Single-Task Learning을 비교하여 성능이 더 나은 모델 구조를 채택했다. 비교 결과 Multi-Task Learning의 경우 Accuracy 값이 각각 0.88, 0.39로 나왔다. 이는 한 모델이 공통 피처를 학습하다 보니 두 Task 성능을 모두 향상시키는 파라미터 학습이 어렵기 때문이다(Sener *et al.*, 2018). 또한 메모리 사용량이 많아 실시간 처리 속도 저하 가능성이 있었다. 결과적으로 해당 모델의 구조는 Single-Task Learning을 기반으로 구축됐다.

먼저 울음/비울음을 분류하는 영아 울음/비울음 분류 모델

의 모델링이다. 본 연구에서 수집된 아기 울음소리 및 소음 데이터의 양은 딥러닝 모델을 처음부터 학습(Training from scratch)시키기에는 충분하지 않다는 한계가 있다. 일반적으로 데이터가 부족한 환경에서 학습을 진행할 경우 과적합(Overfitting) 문제가 발생할 수 있다. 이를 극복하기 위해 본 연구에서는 전이 학습(Transfer Learning) 기법을 적용했다. 대규모 데이터셋으로 사전 학습된 모델의 가중치(Weight)를 활용하는 것이 소규모 데이터셋에서의 분류 성능과 일반화 능력을 효과적으로 향상시킨다(Kong *et al.*, 2020). 이에 따라, 약 200만 개의 방대한 오디오 데이터를 포함하는 AudioSet(Gemmeke *et al.*, 2017)으로 사전 학습된 YAMNet 모델을 특징 추출기(Feature Extractor)로 채택하여 데이터 부족 문제를 해결하고 모델의 신뢰성을 확보하였다. 또한 AudioSet의 경우 울음 클래스를 포함하고 있기 때문에 가장 적절하다고 판단하여 학습 모델로 선정했다. 먼저 pretrained 모델을 특징 추출기로 활용하고, 기존의 분류기 레이어를 제거 후, 증강 데이터를 이용하여 분류기를 새로 학습시켰다. 학습은 증강한 데이터와 비율을 데이터를 언더샘플링하여 비율을 맞춘 상태로 진행했다.

다음으로 영아 울음 원인 분류 모델의 모델링이다. 같은 데이터로 2024년 기준 SOTA모델의 정확도를 넘어선 모델 구조인 RandomForest + MFCC를 기준 모델로 정했다(Hammoud *et al.*, 2024). 해당 연구에서는 단일 특징 추출기를 사용하여 모델링을 진행했고, Tree계열 모델들의 울음 원인 분류 성능이 높다는 것을 확인할 수 있었다. 단일 특징 추출기를 사용한 것은 연구의 한계로 지적되었기 때문에 MFCC, ZCR, Spectral Centroid의 3가지 특성 추출기 조합을 가진 다중 특성 Tree계열 모델들을 3가지로 구현하여 비교했다. 각각 LightGBM, XGBoost, RandomForest모델을 사용했다. 단일 특징 추출기 모델로 40개의 계수를 가진 MFCC 2개를 이용하여 총 80개의 특성을 가지던 기존 연구 모델과 달리, 해당 모델들은 똑같이 MFCC 2개와 ZCR과 Spectral-Centroid의 평균과 표준편차 값, 즉 4개의 특성을 추가하여 84개의 특성을 가진 모델들이다.

• 위험 소음 감지 모델

영아 울음/비울음 감지 모델과 같은 이유로 데이터가 적기

Table 1. Model Performance Evaluation Table (* indicates the best performance)

Model	Accuracy	Recall	Precision	AUC	mAP50
Infant Cry / Non-Cry Detection Model	0.87	0.87	0.87	0.9671*	-
Infant Cry Cause Classification Model (LightGBM)	0.8329*	0.83	0.83	-	-
Infant Cry Cause Classification Model (XGBoost)	0.8289	0.83	0.83	-	-
Infant Cry Cause Classification Model (RandomForest)	0.7656	0.77	0.77	-	-
Hazardous Sound Detection Model	0.9697	0.97	0.97	0.9951*	-
Infant Suffocation Risk Detection Model	-	0.9692	0.9044	-	0.959*

때문에 AudioSet data로 학습한 pretrained model인 YAMNet 모델을 활용하여 전이 학습(Transfer Learning)을 했다(Kong *et al.*, 2020). pretrained 모델을 특징 추출기로 활용하고, 기존의 분류기 레이어를 제거 후, 직접 수집한 데이터를 이용하여 분류기를 새로 학습했다. 학습은 증강한 데이터와 비정보음 데이터를 언더샘플링하여 비율을 맞춘 상태로 진행했다.

• 영아 질식사 위험 감지 모델

호흡기 객체 탐지를 할 수 있으면서도 라즈베리 파이에서 안정적인 성능으로 실시간 추론이 가능해야 하기 때문에 YOLOv8n 모델을 선정했다(Okano *et al.*, 2025). 또한 딥러닝 기반 객체 탐지 알고리즘은 기존의 하드코딩된 특징 기반 방법보다 선행연구의 Haar Cascade가 취약했던 조명 변화와 얼굴 자세 변화등의 비정형적 환경에서도 우수한 성능을 나타낸다(Amjoud *et al.*, 2023). 학습에 사용할 이미지의 크기는 640x640으로 설정했으며, epochs를 100으로 설정했다. 또한 Valid 데이터셋(8%)과 Test 데이터셋(4%)을 별도로 두어 모델이 최적의 일반화 성능을 낼 수 있도록 반복 학습을 진행했다.

(3) 모델 평가

모델 평가는 각각의 모델에 적합한 성능지표를 사용했으며, <Table 1>에서 확인할 수 있다.

• 영아 울음 감지 및 원인 분류 모델

먼저 영아 울음/비울음 분류 모델의 성능평가다. 울음 데이터와 비울음 데이터를 편향 없이 분류해냈고, 정확도는 0.87로 준수한 성능을 보여 실사용 환경에서 안정적인 예측 성능을 보일 것으로 사료된다. 또한 AUC의 경우도 0.9671로 매우 높은 양/음성 판별 능력을 보였다.

다음으로 영아 울음 원인 분류 모델의 성능평가다. LightGBM, XGBoost, RandomForest 모델을 비교한 결과 모든 클래스를 편향 없이 잘 분류한 LightGBM과 XGBoost와 달리, RandomForest의 경우 hungry와 같은 특정 클래스를 오분류하는 경우가 많아 제외했다. 그리하여 정확도가 0.8329로 가장 높았던 LightGBM을 최종 영아 울음 원인 분류 모델로 선정했다.

• 위험 소음 감지 모델

경보음 데이터와 비경보음 데이터를 오분류가 거의 없이 분류해냈고, 정확도도 0.9697로 높은 수치를 보였다. AUC도

0.9951로 매우 높은 수치를 보였다.

• 영아 질식사 위험 감지 모델

객체 탐지 모델의 성능지표인 mAP(mean Average Precision)은 예측 박스가 라벨링 데이터와 겹치는 정도로 판단하는 지표이다. 해당 모델의 경우 mAP50(탐지 기준: 50% 이상)이 0.959로 매우 높게 나왔기 때문에 최종 모델로 선정했다.

4. 결 과

라즈베리 파이는 제한된 연산 성능을 가지고 있기 때문에 모델의 실시간 수행을 위해서는 딥러닝 모델 추론 성능 향상을 위해 TFLite, ONNX 등의 경량화 모델 포맷으로 변환이 필요하다(Ahn *et al.*, 2023). 최종 선정 모델들에 대해서 우선적으로 포맷변환을 했다. 영아 울음 원인 분류 모델에 대해서만 ONNX형식으로 변환하고, 나머지 모델들은 TFLite형식으로 변환했다. 또한 학습 모델을 라즈베리 파이에 로드한 이후부터는 적외선 카메라 모듈 및 마이크 모듈을 통해 수집한 실시간 데이터를 기반으로 이벤트를 감지하기 때문에 기존보다 모델 성능이 저하되는 경향이 있었다. 그래서 각 모델 별로 이벤트 감지 기준을 별도로 설정하여 성능을 높이는 방법을 사용했다. 또한 이에 대한 결과로 모델의 실시간 반응 속도를 확인하기 위해 이벤트가 발생한 시점부터 스마트워치에 알림이 오는 시점까지의 중단 간 지연시간을 측정하였다. 시스템 내의 핵심 기능 3가지에 대해서 각각 30개의 데이터를 무작위로 추출하여 기능별로 30번의 지연시간을 측정하였으며, 그 결과는 <Table 2>에서 확인할 수 있다.

• 영아 울음 감지 및 원인 분류 모델

우선 오디오 청크를 0.25s로 짧게 하고, deque를 사용하여 작은 단위로 자주 읽어 지연을 줄이고, 최신 데이터만 유지하여 메모리/연산을 최소화했다. 또한 RMS 임계값과 울음 예측 확률 임계값을 설정하여 무음이나 약한 신호에 대한 불필요한 추론과 알림을 줄였다. 마지막으로 연속 감지 횟수 조건을 부여하여 일시적 잡음이나 순간적인 신호에 의한 오탐지를 줄여 안정성을 향상시키는 방법을 사용했다.

중단 간 지연시간을 확인한 결과 평균 8.427초의 지연시간

을 가지는 것으로 확인됐다. 영아의 울음의 경우에는 즉각적인 생명 위협이 아니기 때문에 사용자가 아이의 울음에 대해서 알림을 받고 대응이 가능한 수준의 지연시간을 가진다고 판단했다.

- 위험 소음 감지 모델

데이터 샘플 길이를 고정하여 수집 데이터의 길이가 짧으면 0으로 패딩(zero-padding)하고, 길면 절단하여 모델 입력크기로 맞추도록 전처리 코드를 부여했다. 또한 영아 울음 감지 및 원인 분류 모델과 동일한 방식으로 RMS 임계값, 경보음 예측 확률 임계값, 연속 감지 횟수 조건을 부여하여 같은 효과를 얻었다. 마지막으로 주파수 도메인 필터를 활용하여 경보음이 주로 가지는 400~1000Hz의 주파수 대역을 가지는가를 검증하여 오탐지를 감소시킬 수 있도록 했다.

종단 간 지연시간을 분석한 결과, 해당 기능의 평균 지연시간은 13.48초로 다른 기능에 비해 상대적으로 길게 나타났다. 이는 모델의 분류 정확도를 향상시키기 위해 설정한 연속 감지 횟수 조건에 기인한 것으로, 사전에 설정한 임계값을 3회 연속 충족해야 다음 추론 단계로 진행하도록 설계된 프로세스의 영향으로 확인되었다. 임계값을 충족한 이후 스마트워치 알림 단계까지는 평균 4.513초가 소요되는 것으로 나타났다. 이러한 추가 시간은 오탐지를 최소화하고 보다 신뢰성 있는 탐지를 수행하기 위한 필수적인 검증 과정에 해당한다. 또한 전체 평균 지연시간 13.48초는 Occupational Safety and Health Administration (1980)에서 제시한 미국 연방 안전 규정에 따른 화재 감지기 작동 후 경고 발령까지의 허용 지연 기준인 30초에 비해 절반 이하 수준에 해당한다. 따라서 본 시스템의 지연시간은 안전 유효성 측면에서 허용 가능한 범위 내에 있다고 판단된다.

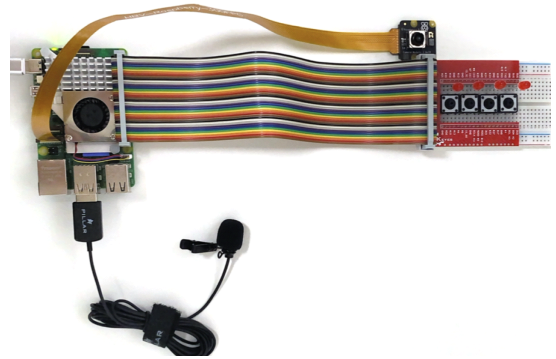
- 영아 질식사 위험 감지 모델

라즈베리 파이의 CPU 부담을 줄이기 위해 8초 간격으로 호흡기를 탐지하도록 했다. 질식사 위험은 옆으로 누운 자세, 아이의 호흡기가 덮임 등과 같이 단시간에 급변하지 않는 위험 요인에 의해 발생하므로, 수 초 단위의 주기적 탐지는 실시간성 요구를 충족한다고 판단하였다(Duncan *et al.*, 2018). 또한 전처리 비용을 최소화하기 위해 카메라를 YOLO입력 크기와 같은 640x640으로 설정하여 프레임 크기를 고정했다. 마지막으로 오탐지를 방지하기 위해 영아 질식사 위험 예측 확률 임계값을 설정하여 확실한 예측에 한하여 알림을 보내도록 적용했다.

종단 간 지연시간을 분석한 결과, 평균 2.22초로 세 가지 기

능 중 가장 짧은 지연시간을 보였다. 질식 사고는 영아에게 치명적인 결과를 초래할 수 있는 고위험 상황이므로, 보호자가 즉각적으로 대응할 수 있도록 지연시간을 최소화하는 것이 매우 중요하다. 본 시스템은 평균 2.22초의 신속한 알림 전달 성능을 보였으며, 이는 긴급 상황 대응 측면에서 적절한 수준의 안전 유효성을 확보한 것으로 판단된다.

4.1 하드웨어 구현



(a) Hardware Configuration



(b) Final Hardware Prototype Configuration

Figure 5. Photographs of the Hardware & Final Hardware Prototype Configuration.

하드웨어는 라즈베리 파이5를 기반으로 만들어졌다. <Figure 5>는 하드웨어의 구성 요소 사진과 최종 프로토타입의 완성 모습이다. 적외선 카메라인 라즈베리 파이 카메라 모듈 3 NoIR와 마이크 모듈인 CM-101USB를 연결하여 실시간

Table 2. End-to-end latency on the Raspberry Pi (Each experiment was repeated 30 times)

Model	Metrics		
	Avg. Latency (s)	Max Latency (s)	Std (s)
Cry Detection Model + Cause Classification Model (LightGBM)	8.427	8.601	0.058
Hazardous Sound Detection Model	13.480	17.380	1.423
Infant Suffocation Risk Detection Model	2.220	2.606	0.083

Scenario \ Function	Dangerous Noise Detection Model	Infant Suffocation Detection Model	Infant Crying Classification Model
Keeping an eye on the child The child is awake	ON	OFF	OFF
Keeping an eye on the child The child is asleep	ON	OFF	OFF
Not watching the child The child is awake	ON	OFF	ON
Not watching the child The child is asleep	ON	ON	ON

Figure 6. Scenario - Function Matrix Visualizing the User's Service Usage Flow.

으로 데이터를 수집할 수 있도록 설계하였다. 또한 사용자들에게 각각의 모델 제어 권한을 부여할 수 있도록 Bread Board와 물리스위치를 이용해 ON/OFF 기능을 부여했고, 이는 라즈베리 파이의 GPIO 입력 제어 시스템을 통해서 연결했다. 총 4개의 스위치가 있으며, 3개의 스위치는 개별 모델 제어 스위치이고, 나머지 하나는 전체 모델을 한 번에 제어할 수 있는 스위치이며, 사용자는 옆에 부착된 LED를 통해 현재 모델이 켜진 것인지, 꺼진 것인지 인지할 수 있다.

<Figure 6>은 사용자의 서비스 이용 상황을 사용자의 아이 확인 여부와 아이의 수면 여부를 기준으로 총 4가지로 도출하여 서비스 사용 목록을 구축한 것이다. 위험 소음 감지 모델의 경우에는 모든 경우에 켜져 있어야 하며, 아이 질식사 위험 감지 모델의 경우에는 아이가 잠들어 있고, 아이에게서 시선을 떼고 있는 경우에만 사용한다. 그리고 영아 울음 감지 및 원인 분류 모델의 경우에는 아이의 수면 여부와 별개로 아이를 보고 있지 않은 경우에 사용하는 것으로 정리됐다. 이를 통해 하드웨어에서 사용자가 개별 서비스를 ON/OFF할 수 있는 권한을 부여하는 것으로 결정됐다.

4.2 통신시스템 구현

본 연구의 육아보조시스템은 하드웨어를 통해 감지한 영아

의 이벤트를 사용자가 알 수 있도록 스마트워치의 푸시알림을 통해 알림을 주는 구조를 가진다. 이를 위해 하드웨어와 스마트워치를 연결할 통신시스템을 구현했다.

통신시스템 구축에는 이벤트(울음, 위험 소음, 질식 위험) 감지 사실을 저장할 Firebase Realtime Database와 푸시알림 시스템을 구현하기 위해 FCM(Firebase Cloud Messaging)을 사용했다. 서버에서 기기 고유 식별자인 Push Token을 생성하여, 개발한 안드로이드 스마트폰 앱에 알림을 보낼 수 있도록 했다. <Figure 2>에서 설명한 바와 같이 하드웨어에서 이벤트를 감지하게 되면 Firebase Realtime Database에 이벤트를 기록하게 되고, Database의 변화가 감지되면 스마트폰 앱에 FCM을 자동으로 전송한다. 최종적으로 사용자에게는 스마트폰과 연동된 스마트워치를 통해 푸시알림을 전송하게 된다.

4.3 시연 기반 검증

최종적으로는 성공적인 시스템 구현의 검증을 목적으로 시스템이 가지는 3가지 핵심 기능들의 시연을 진행했다. <Figure 7>이 핵심 기능들의 시연 모습이다. 또한 영아 울음 감지 및 원인 분류 기능의 경우는 점진적 소거법에 대한 시연을 추가적으로 진행했다. <Figure 8>이 점진적 소거법의 시연 모습이다.

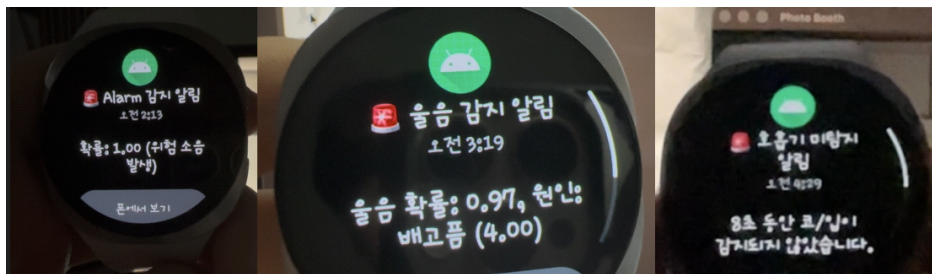


Figure 7. System Implementation Verification Images (from left to right: hazardous sound detection, infant crying detection, and infant suffocation risk detection).

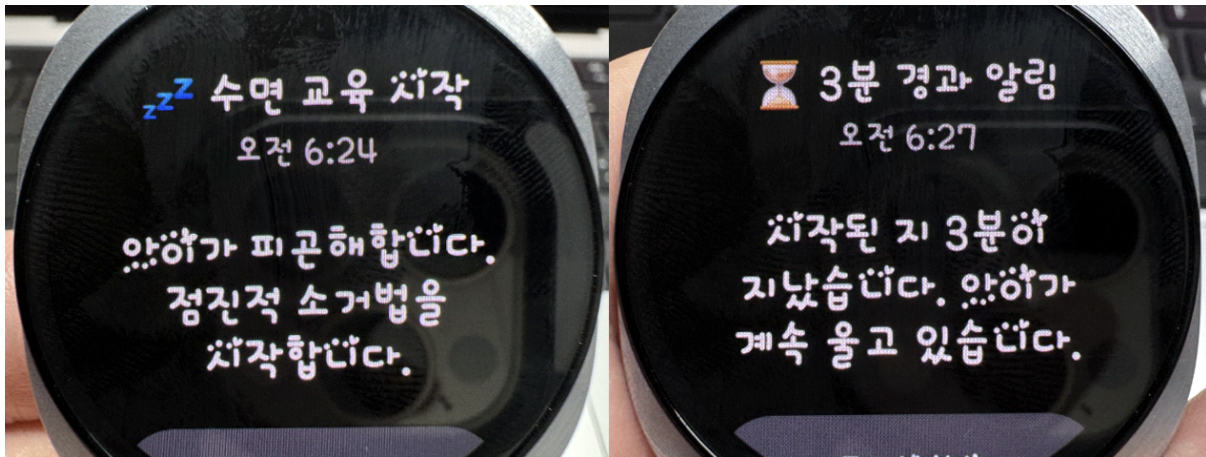


Figure 8. System Implementation of the Graduated Extinction for Infant Cry Detection (left: fatigue-induced crying; right: crying detected after 3 minutes).

• 영아 울음 감지 및 원인 분류 기능

영아 울음 감지 및 원인 분류 기능은 원인이 ‘배고픔’인 울음 음원과 무작위로 선정한 비울음 음원(아이들이 크리스마스 캐롤을 부르는 음원)을 사용하여 시연을 진행했다. 그 결과, 비울음 음원에 대해서는 모델이 소리에 대해서는 반응을 하지만 울음으로 감지하지 않는다는 것을 확인했다. 그리고 울음 음원에 대해서는 97%확률의 울음으로 확인하고, 울음의 원인이 ‘배고픔’이라고 분류함과 동시에 그 결과를 스마트워치의 푸시알림으로 성공적으로 전송하는 것을 확인했다.

점진적 소거법의 경우, <Figure 8>과 같이 원인이 ‘피곤함’으로 분류된 울음 음원을 3분의 임계 시간을 초과하도록 재생하여 시연을 진행했다. 먼저, ‘피곤함’이 원인인 울음이 감지되자 사용자에게 점진적 소거법이 시작되었음을 알리는 푸시알림이 정상적으로 전송되는 것을 확인했다. 이후 첫 번째 임계 시간인 3분이 경과한 뒤에도 울음이 지속적으로 감지되자, 아이가 여전히 울고 있음을 사용자에게 푸시알림을 통해 성공적으로 전달하는 것을 확인했다.

• 위험 소음 감지 기능

위험 소음 감지 기능은 화재경보음 음원을 사용하여 시연을 진행했다. 그 결과, 해당 음원을 100%확률의 위험 소음으로 분류함과 동시에 그 결과를 스마트워치의 푸시알림으로 성공적으로 전송하는 것을 확인했다.

• 영아 질식사 위험 감지 기능

영아 질식사 위험 감지 기능은 영아 더미를 활용하여 시연을 진행했다. 영아의 고개를 틀어 수면 중 자세가 변화한 상황과 영아의 코, 입을 휴지로 가려 영아 호흡기가 감지되지 않는 상황에 대해 검증했다. 그 결과, 영아 더미의 고개가 틀어진 상황에서도 호흡기를 제대로 탐지해내는 것을 확인했다. 그리고 영아 더미의 호흡기를 휴지로 가린 상황에서는 8초 동안 코/입을 감지하지 못함과 동시에 그 결과를 스마트워치의 푸시알림으로 성공적으로 전송하는 것을 확인했다.

5. 결 론

본 연구는 세 가지 핵심 기능 모델의 높은 성능 달성과 하드웨어와 스마트워치를 활용한 성공적인 시스템 구현을 통해 청각 장애인 부모의 육아 독립성 및 안전성 증진에 크게 기여할 잠재력을 보였다. 또한 기존 선행연구가 가지고 있던 주요 기술적 한계점을 효과적으로 보완하는 방향으로 딥러닝 기반 다중모달 육아보조시스템을 설계·구현했다.

먼저, 데시벨 기반의 울음 원인 분류 방식이 가지는 한계를 극복하기 위해 주파수 기반 음향 특징을 활용한 딥러닝 모델을 적용하였으며, AudioSet 데이터셋을 생활 소음 데이터로 활용함으로써 실제 환경에서의 일반화 성능을 향상시켰다. 또한 영아 질식사 위험 감지 기능에서는 조명 변화와 자세 변화에 취약한 Haar Cascade 기반 접근법을 최신 객체 탐지 모델인 YOLOv8n로 대체하여 탐지의 강건성을 개선하였다.

더불어, 기존 연구에서 다루지 않았던 위험 소음 감지 기능을 새롭게 추가하여 청각장애 부모가 화재나 가스 누출과 같은 위험 상황에 보다 효과적으로 대응할 수 있도록 하였다. 마지막으로, 별도의 전용 웨어러블 기기 대신 스마트워치를 활용한 알림 방식을 적용함으로써 사용자의 편의성과 실사용 가능성을 높였다.

5.1 연구의 기여점

본 연구의 기여점은 크게 이론적 기여점과 실용적 기여점으로 나눌 수 있다.

첫 번째 이론적 기여점은 한국형 음향 이벤트 분류 모델의 성능 최적화를 했다는 것이다. 범용 음향 분석YAMNet 모델을 한국 기반 데이터(화재/가스 경보기)로 재학습시켜 AUC 0.9951을 달성하며, 실제 환경 적용 가능성을 극대화하는 이론적 기반을 제시했다. 두 번째 이론적 기여점은 영아 울음 감지·분류 모델의 새로운 접근법을 제시했다는 점이다. 울음/비울

음 감지를 위한 YAMNet 모델, 울음 원인 분류를 위한 LightGBM 모델을 결합하여 AUC 0.9671, 정확도 0.8329라는 우수한 결과를 얻었다. 이는 연산 효율성과 성능을 동시에 확보할 수 있음을 보인 사례로 실시간 영아 울음 분류 기법의 새로운 가능성을 제시한다.

첫 번째 실용적 기여는 라즈베리 파이 기반의 실시간 육아보조시스템을 구현했다는 점에 있다. 육아 상황에서 발생하는 이벤트들을 라즈베리 파이에서 실시간으로 탐지·처리할 수 있는 경량 모델 구조를 제안 및 구현했다. 실용적 기여 두 번째는 청각장애인 부모의 육아 독립성 및 안전성을 증진할 수 있는 토대를 마련했다는 점에 있다. 영아 울음 원인, 질식사 위험, 화재/가스 경보 등의 필수 정보를 스마트워치로 실시간 전달하여 청각장애인 부모의 정보 단절 문제를 해소할 수 있도록 했다. 이를 통해 아이의 안전을 능동적으로 확보하고, 육아 독립성 및 심리적 안정감을 크게 향상시키는 데 기여할 수 있게 됐다.

5.2 추후 연구 방향

본 연구는 사용자 경험 및 사용성에 대한 정량적 검증이 충분히 이루어지지 못한 한계를 가진다. 실제 보호자를 대상으로 한 사용성 평가가 수행되지 않았으며, 시스템 사용 과정에서 발생할 수 있는 인지 부하에 대한 분석 또한 부족하다. 예를 들어, 스마트워치 알람 빈도가 보호자에게 스트레스 요인으로 작용할 가능성, 점진적 소거법 기반의 알람이 실제 양육 방식과 충돌할 가능성, 그리고 청각장애 부모를 대상으로 한 사용자 경험의 정량적 평가가 이루어지지 않았다는 점은 연구의 큰 한계점으로 볼 수 있다. 특히 SUS(System Usability Scale), NASA-TLX와 같은 표준화된 사용성 및 인지 부하 평가 지표가 적용되지 않았기 때문에, 시스템의 실질적인 사용성, 신뢰성, 알람 수용성에 대한 객관적 근거가 충분히 확보되지 못하였다. 이러한 요소들은 장기 사용 시 알람 빈도에 의한 피로, 사용자 신뢰 저하, 시스템 사용 중단으로 이어질 가능성이 있다.

또한 본 연구는 윤리적·안전적 제약으로 인해 실제 영아를 대상으로 한 반복적 위험 상황 재현이 불가능하였으므로, 영아 더미를 활용하여 기능 검증을 수행하였다. 이는 위험 상황을 통제된 환경에서 반복 재현할 수 있다는 장점이 있으나, 실제 영아의 움직임 다양성, 울음 패턴의 개인별 특성, 환경 노이즈 등 복합적 변인을 충분히 반영하지 못한다는 한계를 가진다. 이로 인해 실제 적용 시 모델의 일반화 성능 저하가 발생할 수 있으며, 결과적으로 오탐지(false positive) 및 미탐지(false negative)가 발생할 가능성이 존재한다. 특히 미탐지는 질식 위험 상황, 위험 소음 발생 상황, 장시간 울음 지속 상황에서 보호자의 개입을 지연시켜 잠재적 안전사고로 이어질 수 있다. 반대로 오탐지는 위험하지 않은 상황에서 반복적인 경고를 발생시켜 보호자에게 불필요한 심리적 부담을 주고, 경고 무시 현상으로 이어질 수 있다. 예를 들어, 실제 위험이 아닌 상황

서 화재 경보 알람이 반복적으로 발생할 경우 사용자는 시스템 알람을 신뢰하지 않게 될 수 있으며, 이는 실제 위험 발생 시 대응 지연으로 이어질 위험이 있다. 마찬가지로 울음 원인 분류의 오류가 반복될 경우 보호자는 시스템의 판단을 신뢰하지 않게 될 가능성이 있다. 따라서 오탐지와 미탐지는 단순한 성능 수치의 문제가 아니라, 사용자 신뢰도와 직결되는 핵심 요소이다.

이러한 한계점을 바탕으로 향후 연구에서는 다음과 같은 보완이 필요하다. 우선, 현재 시스템은 각 기능이 단일 모달 기반으로 구성되어 있다, 허나 향후 음향 및 영상 데이터를 통합한 멀티모달 융합 모델을 적용할 경우 실시간 환경에서의 탐지 정확도 및 강건성 향상을 기대할 수 있을 것이다. 또한 실시간으로 수집되는 데이터를 기반으로 한 지속적 모델 재학습 체계를 구축하여 실제 환경 변화에 대응할 필요가 있다. 더불어 실제 보호자를 대상으로 한 사용성 평가 연구를 수행하고, SUS, NASA-TLX 등의 표준화된 지표를 활용하여 인지 부하 및 사용자 경험을 정량적으로 분석할 필요가 있다. 마지막으로, 현재 하드웨어는 기능 검증을 목표로 한 초기 프로토타입 단계이므로 장시간 운용 시 발생할 수 있는 발열 문제 및 내구성 확보를 위한 설계 개선이 요구된다. 이러한 기술적·사용성 측면의 보완이 이루어진다면, 본 시스템은 실질적인 육아 안전 보조 서비스로서의 실체화에 한 걸음 더 다가설 수 있을 것이다.

참고문헌

- Abdulhussien, D. M. and Saud, L. J. (2022), An Evaluation Study of Face Detection by Viola-Jones Algorithm, *International Journal of Health Sciences*, **6**(S8), 4174-4182.
- Ahn, H., Chen, T., Alnaasan, N., Shafi, A., Abduljabbar, M., and Subramoni, H. (2023), Performance Characterization of Using Quantization for DNN Inference on Edge Devices: Extended Version, *arXiv preprint arXiv:2303.05016*.
- Amjoud, A. B. and Amrouh, M. (2023), Object Detection Using Deep Learning, CNNs and Vision Transformers: A Review, *IEEE Access*, **11**, 35479-35516.
- BogeonbokjibuTV (2022), CODA (Child of Deaf Adults) Born to Deaf Parents: The Story of Seven-Year-Old Nahi, YouTube.
- Chandnani, K., Tripathy, S., Parbhakar, A. K., Takiar, K., Singhal, U., Sasikumar, P., and Maheswari, S. (2025), A Novel Smart Baby Cradle System Utilizing IoT Sensors and Machine Learning for Optimized Parental Care, *Scientific Reports*, **15**(1), 19080.
- ChungcheongToday (2020), "Bul-ya" Shouted but Hearing-Impaired Individuals Cannot Escape, ChungcheongToday, Daejeon, Korea.
- Duncan, J. R. and Byard, R. W. (2018), SIDS Sudden Infant and Early Childhood Death: The Past, the Present and the Future, University of Adelaide Press, Adelaide, Australia.
- Gemmeke, J. F., Ellis, D. P. W., Freedman, D., Jansen, A., Lawrence, W., Moore, R. C., Plakal, M., and Ritter, M. (2017), AudioSet: An Ontology and Human-Labeled Dataset for Audio Events, *Proceeding IEEE International Conference on Acoustics, Speech and Signal*

- Processing (ICASSP)*, 776-780.
- Gradisar, M., Jackson, K., Spurrier, N. J., Gibson, J., Whitham, J., Williams, A. S., Dolby, R., and Kennaway, D. J. (2016), Behavioral Interventions for Infant Sleep Problems: A Randomized Controlled Trial, *Pediatrics*, **137**(6), e20151486.
- Hammoud, M., Getahun, M. N., Baldycheva, A., and Somov, A. (2024), Machine Learning-Based Infant Crying Interpretation, *Frontiers in Artificial Intelligence*, **7**, 1337356.
- HankyorehTV (2019), [B-Tag] How do Deaf Parents and Hearing Children Communicate?, YouTube.
- Jeon, S., Kim, J., Park, J., Woo, J., and Kim, J. (2021), A Sudden Infant Death Prevention System for Deaf Parents, *Proceeding Korea Information Processing Society Conference*, **28**(2), 1282-1285.
- Khan, N., Al-Hamid, D. Z., Chong, P. H., Mirza, F., and Kumar, A. (2024), Systematic Review of Sleep Monitoring Systems for Babies: Research Issues, Current Status, and Future Challenges, *Health & Social Care in the Community*, **2024**(1), 5510164.
- Kong, Q., Cao, Y., Iqbal, T., Wang, Y., Wang, W., and Plumbley, M. D. (2020), Panns: Large-Scale Pretrained Audio Neural Networks for Audio Pattern Recognition, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, **28**, 2880-2894.
- Kyrollos, D. and Green, J. (2022), Simultaneously-Collected Multimodal Mannequin Lying Pose (SMaL), *IEEE Dataport*.
- Lee, M. S. (2015), Marital Life Experiences of Deaf Couples, *Korean Journal of Social Welfare*, **67**(3), 125-150.
- Lee, Y. H. and Hwang, S. H. (2017), A Study on the Application of IoT in Infant Care, *The Journal of the Convergence on Culture Technology*, **3**(4), 59-66.
- LG HelloVision Chungnam (2022), Deaf People Vulnerable to Fire Hazards: Installing Visual Alarm Detectors, YouTube.
- Ministry of Health and Welfare (2023), Survey on the Actual Conditions of Persons with Disabilities, Ministry of Health and Welfare, Sejong, Korea.
- Occupational Safety and Health Administration (1980), Employee Alarm Systems (Standard No. 1910.165), United States Department of Labor, Washington, DC.
- Okano, M. T., Lopes, W. A. C., Ruggero, S. M., Vendrametto, O., and Fernandes, J. C. L. (2025), Edge AI for Industrial Visual Inspection: YOLOv8-Based Visual Conformity Detection Using Raspberry Pi, *Algorithms*, **18**(8), 510.
- Park, G. and Kim, I. (2024), A Wearable Embedded System for Monitoring Infants of Deaf Parents, *Proceeding Korean Institute of Information Scientists and Engineers Conference*, 286-288.
- Park, S., Jeon, P., and Park, S. (2015), Baby Crying Detection System for Deaf Parents, *Proceeding Korean Institute of Information and Communication Sciences Conference*, 152-154.
- Pavel, A. M., Rennie, J. M., de Vries, L. S., Blennow, M., Foran, A., Shah, D. K., Pressler, R. M., and Boylan, G. B. (2020), A Machine-Learning Algorithm for Neonatal Seizure Recognition: A Multicentre, Randomised, Controlled Trial, *The Lancet Child & Adolescent Health*, **4**(10), 740-749.
- Ramirez, A. E., Donati, E., and Chousidis, C. (2022), A Siren Identification System Using Deep Learning to Aid Hearing-Impaired People, *Engineering Applications of Artificial Intelligence*, **114**, 105000.
- Ramirez, J., Górriz, J. M., and Segura, J. C. (2007), Voice Activity Detection. Fundamentals and Speech Recognition System Robustness, *Robust Speech Recognition and Understanding*, **6**(9), 1-22.
- Sener, O. and Koltun, V. (2018), Multi-Task Learning as Multi-Objective Optimization, *Advances in Neural Information Processing Systems*, **31**.
- SohaengseongChaekbang (2020), What Happens When a Deaf Couple Has a Baby, YouTube.
- Valani, B. (2023), DonateACry Corpus Features Dataset, Kaggle.
- Virani, A., Duffett-Leger, L., and Letourneau, N. (2021), Parents' Use of Mobile Applications in the First Year of Parenthood: A Narrative Review of the Literature, *Health Technology*, **5**, 14.
- Yao, X., Micheletti, M., Johnson, M., Thomaz, E., and de Barbaro, K. (2022), Infant Crying Detection in Real-World Environments, *Proceeding IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 131-135.
- Yoonsaeng (2021), What Happens When Deaf Parents Raise a Child, YouTube.

저자소개

김정태: 홍익대학교 산업 · 데이터공학과에서 학부생으로 재학 중이다.

김경도: 서울대학교 산업공학과에서 1999년 학사, 2003년 석사 학위를 취득하고 퍼듀대학교에서 산업공학 박사학위를 취득하였다. sk텔레콤 매니저, 삼성전자 책임연구원을 역임하고 2012년부터 홍익대학교 산업 · 데이터공학과 교수로 재직하고 있다. 연구분야는 인간공학, HCI, 사용성 공학이다.