

생체신호를 활용한 학습기반 영유아 스트레스 상태 식별 모델 연구

A Machine Learning Approach for Stress Status Identification of Early Childhood by Using Bio-Signals

전유미(Yu-Mi Jeon)*, 한태성(Tae Seong Han)**, 김관호(Kwanho Kim)***

초 록

오늘날 감정 표현이 서툰 영유아가 처한 극도의 스트레스 상태를 자동적으로 파악하는 것은 영유아의 안전을 위협하며 지속적으로 발생하는 위험 상황의 실시간적인 인지를 위해 반드시 필요한 기술이다. 따라서 본 논문에서는 생체신호를 활용하여 영유아의 스트레스 상태를 분류하기 위한 기계학습 기반의 모델과 생체신호 수집용 스마트 밴드 및 모니터링용 모바일 어플리케이션을 제안한다. 구체적으로 본 연구에서는 영유아의 감정을 나타내는 주요한 요인이 되는 음성 및 심박 데이터의 조합을 활용하여 기존에 널리 알려진 데이터 마이닝 기법을 통해 영유아의 스트레스 상태 패턴을 학습하고 예측한다. 본 연구를 통해 생체신호를 활용하여 영유아의 스트레스 상태 식별을 자동화할 수 있는 가능성을 확인하였으며 나아가서 궁극적으로 영유아의 위험 상황 예방에 활용될 수 있을 것으로 기대된다.

ABSTRACT

Recently, identification of the extremely stressed condition of children is an essential skill for real-time recognition of a dangerous situation because incidents of children have been dramatically increased. In this paper, therefore, we present a model based on machine learning techniques for stress status identification of a child by using bio-signals such as voice and heart rate that are major factors for presenting a child's emotion. In addition, a smart band for collecting such bio-signals and a mobile application for monitoring child's stress status are also suggested. Specifically, the proposed method utilizes stress patterns of children that are obtained in advance for the purpose of training stress status identification model. Then, the model is used to predict the current stress status for a child and is designed based on conventional machine learning algorithms. The experiment results conducted by

이 논문은 교육부의 재원으로 한국과학창의재단(2016년도 학부생 연구프로그램) 및 미래창조과학부의 재원으로 한국연구재단(No. NRF-2014R1A1A1006458)의 지원을 받아 수행된 기초연구사업임.

* First Author, Department of Industrial and Management Engineering, Incheon National University (jym9425@gmail.com)

** Co-Author, Department of Industrial and Management Engineering, Incheon National University (hantaeseong4@naver.com)

*** Corresponding Author, Department of Industrial and Management Engineering, Incheon National University (khokim@inu.ac.kr)

Received: 2017-01-09, Review completed: 2017-02-06, Accepted: 2017-03-13

using a real-world dataset showed that the possibility of automated detection of a child's stress status with a satisfactory level of accuracy. Furthermore, the research results are expected to be used for preventing child's dangerous situations.

키워드 : 생체신호, 기계학습, 스트레스 상태 식별, 데이터마이닝, 데이터분석
Bio-signals, Machine Learning, Stress Status Identification, Data Mining, Data Analysis

1. 서 론

오늘날 빈번한 영유아 사건사고의 발생으로 영유아 안전에 대한 사회적 관심이 높아지면서 영유아의 상태를 실시간으로 판단할 수 있는 기술의 필요성이 증대되고 있다[15, 16]. 영유아의 경우 위험 상황에 처했을 때 보호자에게 스스로 도움을 요청하는 것에 어려움이 존재한다. 이와 관련된 문제 상황들이 사후에 밝혀지면서 보호자들의 불안이 가중되고 있다. 최근 발생한 어린이집에서의 학대사건 및 통학버스 방치로 인한 사망사고 등은 문제 상황의 뒤늦은 인지로 영유아의 안전을 위협한 예로서 유사한 사건사고들이 지속적으로 발생하고 있다[22, 32].

영유아가 처한 위험 상황의 즉각적인 인지는 보호자가 상황에 빠르게 대처하기 위해서 필요하다. 영유아의 위험 상황 인지를 위하여 CCTV가 널리 활용되고 있으나 보호자가 CCTV를 실시간으로 관측하여 문제 상황을 즉각적으로 발견하는 것은 현실적으로 어렵다. 또한, CCTV는 개개인이 아닌 다수를 대상으로 한 영상신호에 국한되고 사각지대가 존재하여 잠재적인 폭력, 사고 등을 인지할 수 없다는 한계를 지니고 있다. 따라서 보다 근본적인 위험 상황 인지를 위하여 영유아가 극도로 스트레스를 받는 상태를 실시간으로 식별할 수 있는 해결방안이 필요하다.

기존의 연구에서는 개인의 스트레스 상태를 식별하기 위한 방법으로 생체신호를 활용하였다. Bakker et al.[4], Healey and Picard[10], Jung and Yoon[11]은 각각 근로자, 운전자, 노인을 대상으로 생체신호를 수집한 후, 스트레스 수준을 평가하였다. 또한 Setz et al.[29], Melillo et al.[21], Kurniawan et al.[14]은 학습능력을 요구하는 시험을 실시하여 실험환경을 조성한 후, 스트레스 상태를 식별하였다. 그러나 기존 연구들은 사후적인 상황 식별을 고려하고 있으며 언어 및 인지 능력이 충분히 발달하지 않은 영유아에게 적합한 요인들을 복합적으로 고려하지 않았다는 한계를 지닌다.

따라서 본 논문에서는 음성 데이터와 심박 데이터의 조합을 활용하여 영유아 스트레스 상태 식별 모델을 제안한다. 영유아의 웃음 및 울음소리와 같은 음성 데이터는 감정 표현이 서툰 영유아의 스트레스 상태를 인지할 수 있는 중요한 요인이 된다[9, 23, 28]. 또한 심박 데이터는 스트레스 식별에 가장 많이 활용되는 생체 신호로써 영유아의 고조된 긴장 상태를 파악할 수 있는 근거가 된다[21, 24]. 또한, 두 데이터의 조합은 언어 및 인지능력과 관계없이 스트레스 상태 식별에 활용할 수 있기 때문에 영유아에게 적합하다.

영유아 스트레스 상태 식별을 위해서는 기계학습의 의사결정나무(DT, Decision Tree), 서포트 벡터머신(SVM, Support Vector Machine),

나이브베이즈(NB, Naive Bayes) 모델을 활용한다. 이 모델들은 대표적인 분류 알고리즘으로서 멀티모달 데이터를 학습한 사용자 예측, 문서의 자동 분류 등에서 널리 활용되고 있다[6, 12, 26]. 구체적으로 분류 모델을 이용하여 음성 데이터와 심박 데이터를 토대로 식별 모델을 생성한 후, 새로운 데이터를 입력받으면 영유아의 스트레스 상태 여부를 분류한다.

나아가서 영유아의 생체신호를 수집 한 후 스트레스 상태를 식별하고 그 결과를 모니터링할 수 있는 시스템을 제안한다. 제안된 시스템은 다음과 같이 세 가지 구성 요소를 포함한다. 첫 번째로 영유아의 생체신호를 수집하기 위하여 음성 및 심박 센서를 탑재한 스마트 밴드를 개발한다. 두 번째로 수집된 생체신호는 본 논문에서 제안하는 영유아 스트레스 상태 식별 모델에 의해 계산되어 스트레스 상태를 식별한다. 마지막으로 식별된 결과를 실시간으로 모니터링할 수 있는 모바일 어플리케이션을 개발한다. 제안된 시스템을 통하여 보호자는 영유아의 문제 상황을 실시간으로 인지하고 대처할 수 있다.

또한 음성 데이터만을 사용한 데이터 집합과 음성 데이터와 심박 데이터를 함께 사용한 데이터 집합의 비교 실험을 통하여 두 생체 데이터 조합의 실효성을 검증한다. 실험 결과, 음성 데이터와 심박 데이터를 함께 사용한 데이터 집합에서 SVM 모델이 약 89%의 정확도로 가장 높은 모델 성능을 보였으며, 이를 통해 영유아의 스트레스 상태식별을 자동화할 수 있는 가능성을 확인하였다.

본 논문의 구성은 다음과 같다. 제2장에서는 생체 데이터를 활용한 스트레스 상태식별과 관련한 기존 연구들을 서술하며, 제3장에서는

영유아 스트레스 상태식별 모델 및 프레임워크를 제안한다. 제4장에서는 시스템 구현과 함께 모델을 이용한 실험 및 결과를 보여주며 마지막으로 제5장에서는 추후의 연구방향 제시와 함께 결론을 맺는다.

2. 관련 연구

본 장에서는 생체신호를 활용한 스트레스상태식별과 관련된 기존 연구를 목적, 사용된 생체신호, 분석 기법에 따라 <Table 1>과 같이 요약할 수 있다.

기존 연구들은 생체신호로써 심박(HR, Heart Rate), 피부 전도도(GSR, Galvanic Skin Response), 심전도(ECG, Electrocardiogram), 근전도(EMG, Electromyography), 뇌파(EEG, Electroencephalography) 등의 다양한 조합을 활용하여 스트레스 상태를 식별하였다. 이 중 심박 데이터는 스트레스 식별에 가장 많이 이용되고 있으나 심박의 변이는 스트레스 외의 다양한 환경에서 발생할 수 있기에 다른 데이터와 함께 활용되고 있다[30]. Sun et al.[31]은 심박 데이터와 함께 피부전도도 및 가속도계(Accelerometer) 데이터를 사용하여 물리적 활동에서의 스트레스를 식별하였다. 그러나 피부전도도는 다른 생체신호보다 움직임에 민감하기 때문에 움직임이 많은 상황에서 정확한 데이터를 얻기 힘들다.

한편, 영유아의 울음소리는 영유아의 감정을 나타내는 주요한 요인이 된다[28]. Abou-Abbas et al.[1], Rosales-Pérez et al.[27], Ruvolo and Movellan[28]은 울음소리를 포함한 음성 데이터를 감지하는 모델을 제안하였다. 이 중,

〈Table 1〉 Summary of Related Studies

Objects(Topics)	Signals Used	Analysis Methodologies	References
Activity-aware mental stress detection (sitting, standing and walking)	HR, GSR, Accelerometer	DT, SVM, Bayes network	Sun et al.[31]
Stress elicitation by examination	Voice, GSR	DT, SVM, K-Means,	Kurniawan[14]
	HR	Latent dirichlet allocation	Melillo et al.[21]
Multi-level assessment model for monitoring elder's health condition	HR, EEG, ECG	SVM, DT, Expectation maximization	Jung and Yoon [11]
Personal health system for detecting stress	GSR	Latent dirichlet allocation, SVM	Setz et al.[29]
Automatic cry detection in early childhood	Voice	Gentle-boost	Ruvolo and Movellan[29]
Automatic classification of infant crying for early disease detection	Voice	Genetic selection of a fuzzy model	Rosales-Pérez et al.[27]
Automatic detection of the expiratory and inspiratory phases in newborn cry signals	Voice	Hidden markov model	Abou-Abbas et al.[1]
Automatic identification of stress causes of employees	GSR	Adaptive windowing	Bakker et al.[4]
Detecting real-world driving stress	HR, EMG, Respiration	Continuous correlations	Healey and Picard[10]

Ruvolo and Movellan[28]는 유치원의 일상소리에서 울음소리를 감지한다는 점에서 본 연구와 유사하다.

또한 Kurniawan et al.은 음성 데이터를 활용하여 피실험자가 내뱉는 단어를 분석하여 스트레스를 식별하는 실험을 진행하였다[14]. 이 외로 Setz et al.[29]과 Melillo et al.[21]은 피실험자에게 학습능력을 요구하는 시험을 실시하여 실험 환경을 조성하였다.

이처럼 기존의 연구들은 언어 및 인지 능력이 충분히 발달하지 않은 영유아에게 적용하기에 어려움이 존재한다. 또한, 음성 데이터를 활용한 연구의 경우 연구의 목적이 영유아의 위험 상황 인지의 측면이 아니라 질병 여부를 예측하는 의학적인 측면 혹은 로봇 개발과 같

은 교육적인 측면이라는 점에 차이가 있다.

본 논문에서는 기존 연구들이 가지는 한계점을 고려하여 영유아에게 적합한 음성 및 심박 데이터의 조합을 사용한다. 또한, 두 가지의 생체신호를 활용한 영유아의 스트레스 상태식별 모델을 통해 영유아의 위험 상황을 감지할 수 있는 연구를 제안한다.

3. 영유아 스트레스 상태 식별 모델

본 장에서는 영유아 스트레스 상태식별 모델을 제안한다. 제3.1절에서는 본 연구에서 다루고자 하는 영유아의 스트레스 상태에 대하여 정의하며, 제3.2절에서는 영유아 스트레스

상태식별 모델의 전체적인 프레임워크를 보여 준다. 제3.3절에서는 데이터 표현 및 전처리 과정을 서술하고 마지막으로 제3.4절은 기계학습을 활용하여 제안된 영유아 스트레스 상태식별 모델을 서술한다.

3.1 스트레스 상태 정의

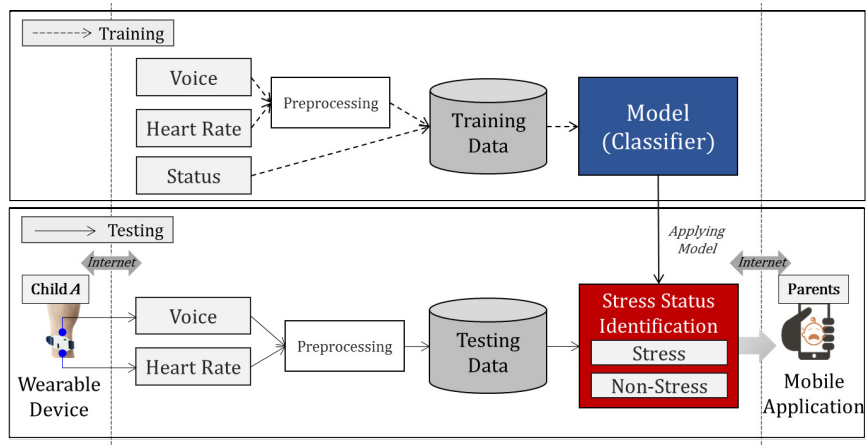
<Figure 1>은 음성 데이터와 심박 데이터의 조합에 따른 영유아 스트레스 상태의 정의를 도식화하고 있다. *L1*과 *L3*는 영유아에게 일반적으로 발생할 수 있는 상황을 나타내며 *L2*와 *L4*는 그 외로 발생할 수 있는 예외적인 상황을 나타낸다.

*L1*은 영유아가 울고 있을 때 심박 수가 높은 상태로서 ‘Stress’로 정의한다. 반면에 *L3*는 영유아가 울지 않을 때 심박 수가 낮은 상태로서 ‘Non-Stress’로 정의한다. *L1*과 *L3*는 각각 영유아가 스트레스를 받으면서 울고 있을 때와 영유아가 휴식을 취할 때의 상황으로 예를 들 수 있다.

		Voice			
		No cry		Cry	
Heart Rate	High	(L4) Non-Stress Audio: No Cry Heart Rate: High Example: Running	(L1) Stress Audio: Cry Heart Rate: High Example: Child Abuse		
	Low	(L3) Non-Stress Audio: No Cry Heart Rate: Low Example: Rest	(L2) Non-Stress Audio: Cry Heart Rate: Low Example: Other child were crying		

<Figure 1> Definition of Stress Status According to the Combination of Voice and Heart Rate

한편 *L2*는 영유아의 심박 수가 낮으나 영유아의 울음소리가 인식되는 상태이다. 예를 들어 관찰 대상의 영유아는 울고 있지 않으나 옆에 있는 다른 영유아가 우는 상황이 될 수 있다. *L4*는 심박 수가 높으나 영유아가 울지 않는 상태이다. 예를 들어 영유아가 달리기와 같은 체육 활동으로 인하여 심박 수가 올라간 상황이



<Figure 2> Proposed Framework for Stress Status Identification of a Child

될 수 있다. $L2$ 와 $L4$ 는 $L3$ 과 함께 'Non-Stress'로 정의한다.

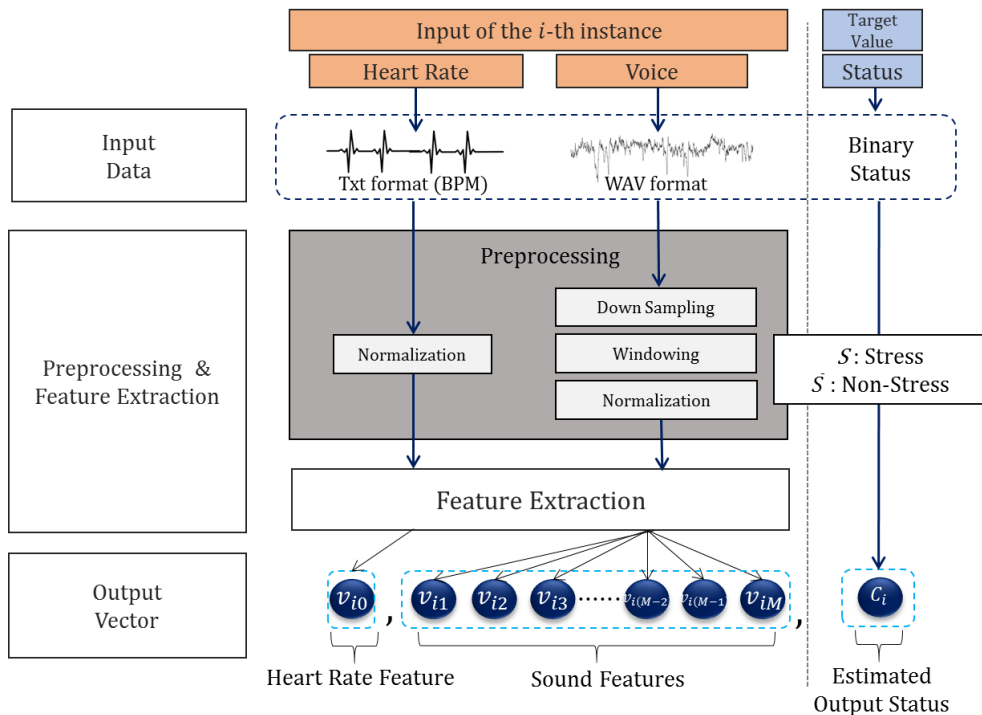
즉, $L1$ 만이 영유아가 스트레스를 받는 상태를 나타내며 그 외의 $L2, L3, L4$ 는 모두 영유아가 스트레스를 받지 않는 상태를 나타낸다.

3.2 프레임워크

<Figure 2>는 본 논문에서 제안하는 영유아 스트레스 상태식별 모델의 프레임워크를 보여준다. 먼저 음성 및 심박 데이터로 이루어진 입력 데이터는 전처리 과정을 거쳐 학습 데이터(Training Data)로 이용되며, 학습 데이터는 영유아의 스트레스 상태식별 모델을 생성한다.

생성된 모델은 학습 데이터가 아닌 새로운 실험 데이터(Testing Data)에서 영유아의 스트레스 상태식별을 위해 활용된다. 실험 데이터가 모델에 적용되면 영유아의 상태는 'Stress' 또는 'Non-Stress'로 식별된다. 또한 실험 데이터는 영유아 A(Child A)가 착용한 웨어러블 디바이스인 스마트 밴드로부터 음성과 심박 신호의 입력 데이터를 수집한 후 전처리 과정을 거쳐 이용된다.

식별 결과는 영유아 보호자(Parents)의 모바일 어플리케이션에 실시간으로 전송된다. 스마트 밴드로부터 영유아의 스트레스 상태가 감지되면 모바일 어플리케이션은 보호자에게 영유아의 상태를 알림으로 전송하여 위험 상황을 즉각적으로 인지할 수 있도록 한다.



<Figure 3> Flow of Input Data Representation and Feature Extraction

3.3 데이터 표현 및 전처리

본 연구에서는 총 N 개의 인스턴스(Instance), $T = \{V_1, V_2, V_3, \dots, V_N\}$,를 사용하며 T 의 원소는 독립 동일 분포(Independent Identical Distribution)를 가진다. 각 인스턴스는 한 개의 심박 데이터와 M 개의 음성 데이터로 구성되며 그 중 i 번째 인스턴스는 입력벡터 $V_i = \{v_{i0}, v_{i1}, \dots, v_{iM}\}$, $i = 1, 2, \dots, N$ 로 표현된다. 또한, T 에서 입력벡터의 j 번째 속성들의 집합은 $U_j = \{v_{1j}, v_{2j}, \dots, v_{Nj}\}$, $j = 0, 1, \dots, M$ 로 표현되며 U_j 와 $U_{j'} = \{v_{1j'}, v_{2j'}, \dots, v_{Nj'}\}$, $j \neq j', j' = 0, 1, \dots, M$ 는 서로 독립임을 가정한다.

<Figure 3>은 입력벡터를 생성하는 과정을 도식화하고 있다. 구체적으로, 심박 데이터로는 분당 심박수 값(BPM, Beats Per Minute)을 사용하였으며 이는 연속된 두 번의 심장 박동 간의 시간 간격(R-R Interval)에 의해 계산된다. 이 때, 첫 번째 심장 박동의 피크 값을 R_1 이

라 하고 두 번째 심장 박동의 피크 값을 R_2 라 한다. R_1 과 R_2 간의 시간 간격이 60초일 경우 심장 박동 발생 시, i 번째 입력벡터의 분당 심박수 h_i , $i = 1, 2, \dots, N$ 는 식 (1)과 같이 계산한다[2]. 최종적으로 h_i 는 식 (2)와 같이 정규화를 통해 i 번째 입력벡터의 첫 번째 원소인 v_{i0} , $i = 1, 2, \dots, N$ 로 설정된다.

$$h_i = \frac{60}{R_2 - R_1}, \quad (1)$$

$$v_{i0} = \frac{h_i - \min(h_{i'})}{\max(h_{i'}) - \min(h_{i'})}, \quad (2)$$

여기에서 $\min(h_{i'})$, $i' = 1, 2, \dots, N$ 는 h_i 에서의 최솟값을 의미하며 $\max(h_{i'})$, $i' = 1, 2, \dots, N$ 는 h_i 에서의 최댓값을 의미한다.

음성 데이터는 소리특징추출기를 통해 일련의 과정을 거쳐 추출된 소리의 특징 값이다. 이 때 소리의 특징 값이란 소리가 가지는 특징들

<Table 2> Representative Sound Features

No	Features	No	Features	No	Features
1	Power Spectrum	10	Spectral Flux	19	Fraction of Low-Energy Frames
2	Magnitude Spectrum	11	Partial-Based Spectral Flux	20	Linear Predictive Coding
3	Magnitude Spectrum Peaks	12	Method of Moments	21	Beat Histogram
4	Spectral Variability	13	Area Method of Moments	22	Strongest Beat
5	Spectral Centroid	14	Mel Frequency Cepstral Coefficient(MFCC)	23	Beat Sum
6	Partial-Based Spectral Centroid	15	Area Method of Moments of MFCCs	24	Strength of Strongest Beat
7	Partial-Based Spectral Smoothness	16	Zero Crossings	25	Strongest Frequency via Zero Crossings
8	Compactness	17	Root Mean Square	26	Strongest Frequency via Spectral Centroid
9	Spectral Roll-off Point	18	Relative Difference Function	27	Strongest Frequency via FFT Maximum

을 푸리에 변환(Fourier Transform)을 통해 수치화시킨 값이며, 소리가 입력되면 다운 샘플링(Down Sampling)과 윈도우(Windowing)의 전처리 과정을 거친다[19]. 다운 샘플링은 출력 데이터를 추출하는 시간을 줄이기 위해 사용되며 윈도우는 소리의 특징이 불연속적으로 급격히 변하는 것을 막기 위해 사용된다[18, 7]. 마지막으로 소리 특징 값의 정규화를 거쳐 최종적인 음성특징 값을 추출한다.

본 논문에서는 소리 특징 값의 추출을 위하여 공개된 소리특징추출기 라이브러리를 사용하였다[20]. <Table 2>는 McKay의 연구에서 사용되는 소리특징추출기 라이브러리를 통해서 추출된 대표적인 값들의 특징을 나타낸다[19].

또한 입력벡터에서 제안 모델을 통해 예측되는 영유아의 상태식별 결과 C_i , $i = 1, 2, \dots, N$ 는 식 (3)과 같이 'Stress' 상태인 S 와 'Non-Stress' 상태인 \bar{S} 로 분류된다.

$$C_i = \begin{cases} S, & \text{if } V_i \text{ is in the 'Stress' status} \\ \bar{S}, & \text{otherwise} \end{cases} \quad (3)$$

3.4 영유아 스트레스 상태식별 모델

본 연구에서는 대표적인 기계학습 분류모델인 NB, DT, SVM을 적용하여 영유아의 스트레스 상태를 식별한다.

첫 번째 분류모델 NB는 V_i , $i = 1, 2, \dots, N$ 가 상태 S 에 속할 확률 $P(C_i = S | V_i)$, $i = 1, 2, \dots, N$ 를 식 (4)와 같이 계산하며 상태 \bar{S} 에 속할 확률 $P(C_i = \bar{S} | V_i)$, $i = 1, 2, \dots, N$ 는 $1 - P(C_i = S | V_i)$ 와 같이 계산한다.

$$\begin{aligned} & P(C_i = S | V_i) \\ &= \frac{P(C_i) \cdot P(v_{i0}, v_{i1}, \dots, v_{iM} | C_i)}{P(V_i)} \quad (4) \\ &\propto P(C_i) \prod_{j=0}^M P(v_{ij} | C_i) \end{aligned}$$

연속적인 값을 지니는 심박 및 음성 데이터는 감독 이산화(Supervised Discretization)를 거쳐 활용하였다[8]. 또한, 집합 U_j 의 원소가 독립을 가정하기 때문에 $P(v_{i0}, v_{i1}, \dots, v_{iM} | C_i)$ 는 $\prod_{j=0}^M P(v_{ij} | C_i)$, $i = 1, 2, \dots, N$, $j = 0, 1, \dots, M$ 으로 다시 나타낼 수 있다. 계산된 확률 값을 바탕으로 NB는 V_i , $i = 1, 2, \dots, N$ 에서 $P(C_i = S | V_i) > P(C_i = \bar{S} | V_i)$ 인 경우 상태 S 로 분류하며, 반대의 경우 \bar{S} 로 분류하여 의사결정을 한다.

두 번째 분류모델 DT는 가장 큰 정보이득량(Information Gain)을 갖는 입력벡터 속성을 결정 노드(Decision Node)로 선택하여 최종 의사결정 값을 분류할 때까지 가치를 전개한다[24]. 정보이득량은 엔트로피를 기반으로 계산되며 엔트로피는 T 에서 S 또는 \bar{S} 의 혼잡성을 계산하며 영유아가 주로 스트레스를 받거나 반대로 전혀 스트레스를 받지 않는 경우와 같이 S 혹은 \bar{S} 의 한 가지의 상태만을 가질 때 최소가 된다.

결정노드가 되는 입력벡터의 속성은 $U_j = \{v_{1j}, v_{2j}, \dots, v_{Nj}\}$ 라 표기된다. 연속형 원소를 갖는 U_j 에서 분할점(Split Point)을 설정하기 위해 U_j 에서 연속된 원소 두 개의 평균을 계산한다. 총 $N-1$ 개 평균에서의 엔트로피를 계산 후 엔트로피가 가장 낮을 때를 분할점 $Split(v^*)$

라 할 때, 의사결정나무는 $v_{ij} < Split(v^*)$ 와 $v_{ij} > Split(v^*)$ 으로 의사결정 값을 분류한다[8].

마지막 분류모델 SVM은 벡터 공간에서 S 에 속한 입력벡터와 \bar{S} 에 속한 입력벡터 사이의 마진(Margin)을 최대화하는 판별 경계를 찾아 영유아의 스트레스 상태를 분류하며 식 (5)는 목적함수를 나타낸다.

$$\begin{aligned} \min \quad & \frac{1}{2} \| \mathbf{w} \|^2 + \lambda \sum_{i=1}^N \xi_i \\ \text{s.t.} \quad & y(\mathbf{w}^T \cdot \mathbf{V}_i + b) - 1 + \xi_i \geq 0, \\ & \xi_i \geq 0, \\ & \sum_{i=1}^N \xi_i \leq \lambda, \\ & i = 1, 2, \dots, N \end{aligned} \quad (5)$$

여기에서 y 는 영유아의 상태로서 S 에서 1의 값을 가지며 \bar{S} 에서 -1의 값을 가진다. w 는 판별 경계에 수직인 법선 벡터를 의미한다. 본 논

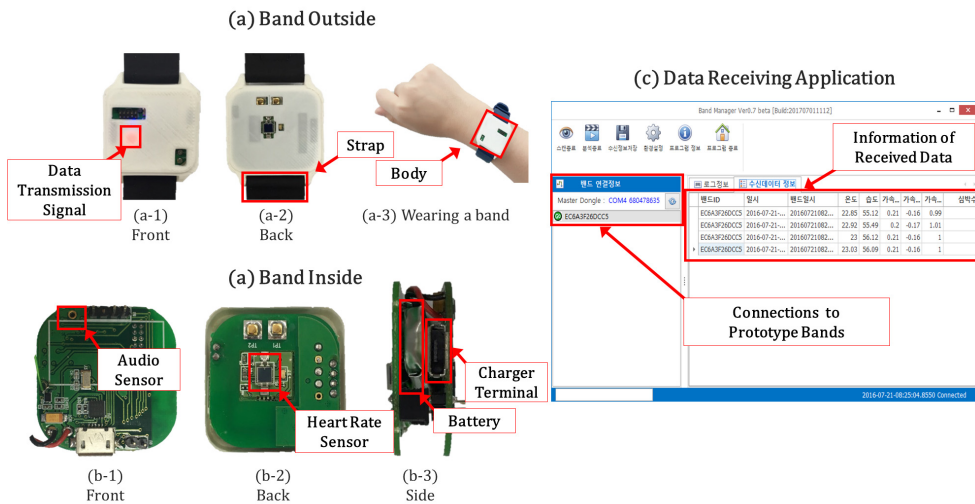
문에서는 S 와 \bar{S} 만으로 분류할 수 없는 데이터의 노이즈를 감안하여 판별 경계에 여유를 둔 소프트 마진 방식을 사용한다. ξ_i 는 모델이 감안하는 오류의 크기이며 λ 는 판별 경계선의 영향을 조절하는 변수이다.

SVM은 학습 데이터를 통하여 판별 식 $(\hat{\mathbf{w}}^T \cdot \mathbf{V}_i + \hat{b}) - 1 + \hat{\xi} = 0$ 이 생성되면 입력벡터가 $(\hat{\mathbf{w}}^T \cdot \mathbf{V}_i + \hat{b}) - 1 + \hat{\xi} > 0$ 에 속하는지 $(\hat{\mathbf{w}}^T \cdot \mathbf{V}_i + \hat{b}) - 1 + \hat{\xi} < 0$ 에 속하는지에 따라 영유아의 스트레스 상태를 분류한다.

4. 시스템 구현 및 실험결과

4.1 스마트 밴드 구현

<Figure 4>는 개발된 스마트 밴드의 프로토타입과 데이터 수신 프로그램을 보여준다.



<Figure 4> Outside(a) and Inside(b) of the Developed Band, and a Screenshot of Data Receiving Application(c)

스마트 밴드는 영유아 맞춤형 밴드로서 음성 센서와 심박 센서가 내장되어 있다. 생체신호는 두 가지의 센서를 통하여 수집된다.

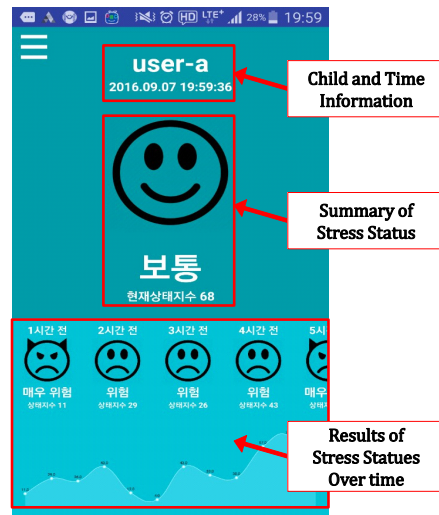
<Figure 4>의 (a)는 스마트 밴드의 외형을 나타낸다. (a-1), (a-2), (a-3)은 각각 밴드의 앞면, 뒷면, 착용 상태의 모습을 보여준다. (b)는 스마트 밴드의 내부로 (b-1), (b-2), (b-3)에서 음성 센서, 심박센서, 배터리와 충전 단자를 포함한다.

<Figure 4>의 (c)는 스마트 밴드로부터 수집되는 데이터를 확인할 수 있는 어플리케이션의 화면이다. 개발된 데이터 수신 어플리케이션을 통하여 밴드의 연결 상태 및 데이터 수집 상태를 확인할 수 있다. 수집된 데이터는 전처리 과정을 거친 후 입력벡터를 형성하며, 입력 벡터는 서버로 전송되어 데이터베이스에 저장된다.

4.2 모바일 어플리케이션 구현

<Figure 5>는 보호자가 관심 있는 영유아의 스트레스 상태를 시간에 따라 모니터링 할 수 있는 모바일 어플리케이션 화면을 나타낸다. 모바일 어플리케이션은 실시간 및 과거 시간대별 영유아의 스트레스 상태 정보를 보여주며 사용자가 습관적이고 지속적으로 사용함으로써 간편하고 실시간적인 모니터링의 매개체 역할을 수행할 수 있다[13, 25].

구체적으로 모바일 어플리케이션은 스마트 밴드와 일대일 연동되어 동작되어, 스마트 밴드에서 실시간으로 전송하는 정보와 데이터베이스에 저장된 지나간 정보들을 한 화면에서 보여준다. 만약 영유아의 상태가 스트레스로 식별된다면 모바일 어플리케이션은 알림을 통하여 보호자에게 즉각적인 상황을 전달할 수 있다.



<Figure 5> Screenshot of Mobile Application for Parents

4.3 실험 환경

본 논문에서 제안하는 영유아 스트레스 상태식별 모델 검증에 위한 실험 데이터로는 <Table 3>과 같이 두 가지의 데이터 집합 (Dataset)을 구성하였다.

<Table 3> Dataset Descriptions

Datasets	Features Used
Dataset <i>EH</i>	Sound Features
Dataset <i>IH</i>	Sound Features and a Heart Rate Feature

먼저 Dataset *EH*는 입력 벡터의 속성이 음성 데이터에서 추출된 소리 특징 값으로만 구성되어 있다. 반면 Dataset *IH*는 음성 데이터와 심박 데이터의 조합을 사용하여 입력 벡터의 속성이 음성 특징 값과 함께 심박 특징 값을 포함하여 구성되어 있다.

구체적으로 음성 데이터는 10초 동안의 기록으로 다양한 소리 및 음향효과를 포함하는 공개된 데이터[3]를 이용하여 ‘Stress’ 상태를 표현하는 데이터로써 울음소리를, ‘Non-Stress’ 상태를 표현하는 데이터로써 웃음소리, 대화소리, 비명소리, 박수 소리 및 생활 잡음을 구성하였다. 또한 음성 데이터 추출 시 설정은 기존 연구들을 참고하여 표본 추출 비율(Sampling Rate) 16kHz에서 윈도우의 사이즈는 32ms로 설정하며 데이터의 손실을 최소화하기 위하여 윈도우를 겹치는 오버래핑(Overlapping)을 66%로 설정하여 진행했다[17].

심박 데이터는 일반 피실험자 6명을 대상으로 일주일동안 일상 생활에서의 데이터를 수집한 후 설문조사를 통해 피실험자가 스트레스 상태에 있을 때와 비스트레스 상태에 있을 때의 심박 수를 분류하였다. 그 후, 수집한 음성 데이터를 듣고 각 음성과 적합한 상태일 때의 심박수 값을 피실험자들의 심박수 값과 비교하여 전문가의 견해에 따라 형성하였다.

데이터 집합 입력벡터의 수는 Dataset *EH*와 Dataset *IH* 모두 총 496개이며 상태 *S*로 구분되는 입력벡터 수는 124개, \bar{S} 로 구분되는 입력벡터 수는 372개로 구성되어 있다.

분류 모델의 파라미터 설정은 다음과 같다. DT 모델은 불필요한 트리의 전개를 막기 위해 가지치기 방식을 선택하여 가지치기를 조절하기 위한 임계값으로 0.5로 설정하였다. SVM 모델은 선형 커널을 사용하였으며, λ 의 값을 1.0으로 설정하였다.

제안된 모델의 성능을 평가하기 위한 지표로써는 데이터의 불균형을 고려하여 일반 정확도(Accuracy)와 균형된 정확도(Balanced Accuracy)를 함께 사용하였다. 균형된 정확도

는 불균형한 데이터 집합에서의 편향된 분류를 감지할 수 있으며 균형적인 데이터 집합을 가질 때 대수적으로 일반 정확도와 같은 값을 갖는다[5].

본 논문에서 일반 정확도는 식 (6)과 같이 전체 데이터의 개수에서 영유아의 상태를 올바르게 분류한 데이터 개수의 비율을 의미한다.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (6)$$

여기에서 *TP*(True Positive)와 *TN*(True Negative)는 각각 영유아가 스트레스를 받을 때 *S*로 옳게 분류한 결과와 영유아가 스트레스를 받지 않을 때 \bar{S} 로 옳게 분류한 결과를 의미한다. 반면 *FP*(False Positive)와 *FN*(False Negative)은 각각 영유아가 스트레스를 받고 있지 않으나 *S*로 잘못 분류한 결과와 영유아가 스트레스를 받고 있으나 \bar{S} 로 잘못 분류한 결과를 의미한다.

균형된 정확도는 실제 영유아가 스트레스를 받을 때 *S*로 분류한 비율인 식 (7)의 민감도(Sensitivity)와 실제 영유아가 스트레스를 받고 있지 않을 때 \bar{S} 로 분류한 비율인 식 (8)의 특이도(Specificity)를 이용하여 식 (9)와 같이 계산한다.

$$Sensitivity = \frac{TP}{TP + FN} \quad (7)$$

$$Specificity = \frac{TN}{FP + TN} \quad (8)$$

$$Balanced Accuracy = (Sensitivity + Specificity) / 2 \quad (9)$$

학습 데이터와 실험 데이터의 비율은 9:1로 10겹 교차 검증(10-folds cross validation)을

실시하여 모든 데이터가 실험에 쓰일 수 있도록 하였다.

4.4 실험 결과 및 분석

<Table 4>는 정확도 계산에 사용된 *TP*, *TN*, *FP*, *FN* 각각의 개수를 나타낸다. Dataset *EH*에서는 영유아가 실제로 스트레스를 받는 상태를 바르게 예측한 *TP*가 매우 적음을 확인할 수 있다. 이는 본 논문에서 \bar{S} 의 예외적인 상황을 가정하기 위해 데이터의 상태 값을 편향적으로 구성하여 불균형이 생겼기 때문으로 해석된다. 특히 Dataset *EH*의 DT 모델 결과의 경우 유사한 변수 정보량을 가지는 변수들 중 큰 값을 갖는 하나의 변수만으로 가치를 전개하는 규칙 (Rule) 기반 모델의 한계인 것으로 판단된다.

반면 Dataset *IH*에서 *TP*는 DT 모델과 SVM 모델에서 각각 89개, 93개로 Dataset *EH*의 *TP*에 비하여 크게 증가하는 것을 확인할 수 있다. 이는 곧 심박 데이터가 DT 모델에서 상위 결정노드에 위치하여 성능을 향상시켰으며, SVM 모델 또한 더 정교한 판별경계를 가질 수 있음을 보여준다. 즉, Dataset *IH*에서 *TP*의 증가가 모델 정확도 향상에 영향을 끼쳤으며 이는 곧 음성 데이터와 심박 데이터의 조

합이 영유아의 스트레스 상태를 식별하기에 적합함을 의미한다.

<Table 5>는 각 데이터 집합에 따른 분류 모델의 민감도, 특이도, 일반 정확도, 균형된 정확도를 보여준다. 실험 결과, 세 가지 분류 모델의 일반 정확도와 균형된 정확도 평균은 Dataset *EH*에서 각각 65.52%와 52.47%를 나타냈으며 Dataset *IH*에서 80.64%와 78.23%를 나타냈다.

분류 모델은 균형된 정확도로 성능 측정 시, 일반 정확도에 비하여 낮은 성능을 보임을 확인할 수 있다. 균형된 정확도는 민감도와 특이도에 의해 계산되어 클래스 불균형에 의해 발생하는 오류를 구분할 수 있기 때문이다. Dataset *EH*에서 DT와 SVM의 낮은 균형된 정확도는 민감도가 0에 가까운 것으로 보아 대다수의 인스턴스를 \bar{S} 로 분류하였기 때문인 것으로 판단된다.

실험 결과를 통하여 음성 데이터만 사용하였을 때보다 음성 및 심박 데이터의 조합을 사용하였을 때 모델이 더 좋은 성능을 보이는 것을 검증하였다. 즉, 음성 데이터만으로 영유아의 스트레스 상태를 식별하기에는 한계가 존재하여 보조적인 수단이 필요하며 심박 데이터를 추가한 음성 및 심박 데이터의 조합이 그 한계를 절충하여 성능 개선을 보인 것으로 분석된다.

<Table 4> Numbers of Correct and Incorrect Classification Instances

	Dataset <i>EH</i>			Dataset <i>IH</i>		
	Naïve Bayes	Decision Tree	SVM	Naïve Bayes	Decision Tree	SVM
<i>TP</i> (True Positive)	91	4	3	91	89	93
<i>TN</i> (True Negative)	220	358	299	227	350	350
<i>FP</i> (False Positive)	152	14	73	145	22	22
<i>FN</i> (False Negative)	33	120	121	33	35	31

〈Table 5〉 Performances of Classifiers According to Dataset Used

Classifiers \ Datasets	Sensitivity		Specificity		Accuracy		Balanced Accuracy	
	<i>EH</i>	<i>IH</i>	<i>EH</i>	<i>IH</i>	<i>EH</i>	<i>IH</i>	<i>EH</i>	<i>IH</i>
NB	73.39	73.39	59.14	61.02	62.7	64.11	66.27	67.21
DT	3.23	71.77	96.24	94.09	72.98	88.51	49.74	82.93
SVM	2.42	80.38	75	94.09	60.89	89.31	41.4	84.55
Average	26.35	75.18	76.79	83.07	65.52	80.64	52.47	78.23

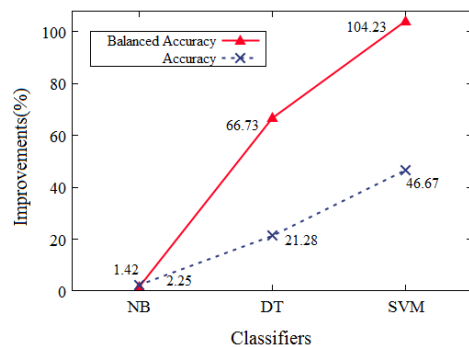
다음으로 세 가지 분류모델의 성능을 비교하기 위하여 Dataset *IH*에서 가장 좋은 성능을 보인 SVM 모델을 다른 모델과 비교하였을 때의 일반 정확도 및 균형된 정확도의 향상률(Improvements)을 확인하였다. 정확도 향상률은 식 (10)과 같이 계산된다. 여기서 $Accuracy_{SVM}$ 은 SVM 모델의 일반 정확도 및 균형된 정확도를 나타내며, $Accuracy_{comp}$ 은 비교 모델의 일반 정확도 및 균형된 정확도를 의미한다.

$$Improvements = \frac{Accuracy_{SVM} - Accuracy_{comp}}{Accuracy_{comp}} \quad (10)$$

Dataset *IH*에서 SVM 모델의 일반 정확도는 DT 모델과 NB 모델 대비 약 0.9%, 39.31%의 향상률을 보였으며, 균형된 정확도는 각각 1.96%, 25.8%의 향상률을 보였다. SVM 모델의 향상률은 NB 모델에 대비하여 눈에 띄는 향상률을 보여준다. 반면 DT 모델 대비 향상률은 미미함을 확인할 수 있다. 이는 DT 모델과 SVM 모델이 Dataset *IH*에서 모두 높은 확률로 영유아의 스트레스 상태를 분류하였기 때문이다. 그러나 Dataset *EH*에서 DT 모델이 SVM 모델보다 더 좋은 성능을 보였던 것을 감안하였을 때, Dataset *IH*에서 SVM 모델이 DT 모델보다 높은 결과를 가져온 것은 음성

및 심박 데이터 조합이 SVM 모델에서 보다 강력한 성능을 낼 수 있는 것으로 판단된다.

〈Figure 6〉은 두 데이터 집합 대비 일반 정확도와 균형잡힌 정확도의 향상률을 나타낸다. 실험 결과, 세 가지 분류모델은 Dataset *IH*를 사용하였을 경우 모두 Dataset *EH* 대비 더 나은 성능을 보여준다. SVM 모델의 경우 일반 정확도와 균형된 정확도에서 각각 약 46%, 104%로 가장 높은 정확도 향상률을 보였다. 특히 균형된 정확도를 사용하였을 때 크게 향상되는 것을 확인할 수 있으며 이는 Dataset *EH*에서 발생한 불균형 데이터 집합으로 인한 오류를 음성 및 심박 데이터의 조합을 통해 개선하였음을 의미한다.



〈Figure 6〉 Improvements According to Datasets in Terms of Accuracy and Balanced Accuracy

반면, NB 모델은 두 데이터 집합에서 모두 정확도 약 60% 대의 현저히 낮은 결과를 보였다. 이는 본 실험에서 사용된 데이터 집합이 입력벡터 속성의 개수에 비하여 적은 입력벡터의 수를 가지기 때문에 다음과 같은 한계점을 지닌 것으로 분석된다. 실험 결과, 인스턴스 각 속성의 확률 값이 대체로 0에 근사하였고 이에 연속적인 곱으로 입력벡터가 상태에 속할 확률이 0에 수렴하였다. 따라서 조건부 확률 식에 로그 변환 도입이 필요할 것으로 판단된다.

5. 결 론

본 논문에서는 생체신호를 활용하여 영유아의 스트레스 상태를 식별하는 모델을 제안하였다. 감정 표현이 서툰 영유아의 경우 위험 상황이 사후에 밝혀지며 문제가 가중되기 때문에, 실시간적으로 영유아의 스트레스 상태를 파악하여 위험 상황을 인지하기 위한 방법론을 고안하였다. 생체신호로서 영유아의 감정을 나타낼 수 있는 음성 데이터와 심박 데이터를 활용하였으며, 기계학습 기반의 모델을 생성하여 영유아의 스트레스 상태를 식별하였다.

제안 모델은 음성 및 심박 데이터의 조합을 사용하였을 때 가장 우수한 성능을 나타내었다. 또한 스마트 밴드를 통해 생체신호를 수집하고 제안 모델을 통해 식별된 영유아의 스트레스 상태를 실시간으로 모바일 어플리케이션에 전송하는 프레임워크를 제안하였다. 이를 통해 보호자는 영유아의 위험 상황을 즉각적으로 인지하고 빠르게 대처할 수 있을 것이다.

본 논문의 연구 결과는 두 생체신호의 조

합을 활용한 영유아 스트레스 상태식별 모델 및 제안 프레임워크가 실제 영유아의 위험 상황에 적용될 수 있는 가능성을 시사한다. 추후 어린이집 및 보호자의 협조를 통해 영유아를 대상으로 한 실험 환경이 조성된다면 실제 상황에서 모델을 검증할 수 있을 것으로 기대된다.

또한 가속도계, 자이로스코프 센서 등 영유아의 움직임을 감지할 수 있는 생체 센서를 추가하여 구체적인 영유아의 행동 패턴을 분석하고 보다 정교한 영유아의 스트레스 상태식별 모델을 수립하고자 한다.

References

- [1] Abou-Abbas, L., Alaie, H. F., and Tadj, C., "Automatic Detection of the Expiratory and Inspiratory Phases in Newborn Cry Signals," *Biomedical Signal Processing and Control*, Vol. 19, pp. 35-43, 2015.
- [2] Asyali, M. H., "Discrimination Power of Long-term Heart Rate Variability Measures," *Engineering in Medicine and Biology Society*, Vol. 1, pp. 200-203, 2003.
- [3] AudioMicro, [URL] <http://www.audiomicro.com>.
- [4] Bakker, J., Pechenizkiy, M., and Sidorova, N., "What's Your Current Stress Level? Detection of Stress Patterns from GSR Sensor Data," *Proceedings of the IEEE International Conference on Data Mining Workshop*, pp. 573-580, 2011.

- [5] Brodersen, K. H., Ong, C. S., Stephan, K. E., and Buhmann, J. M., "The balanced accuracy and its posterior distribution," Proceedings of the 20th International Conference on Pattern Recognition IEEE Computer Society, pp. 3121-3124, 2010.
- [6] Choi, Y., Park, K., Kim, S., and Park, J., "A Two-Phase On-Device Analysis for Gender Prediction of Mobile Users Using Discriminative and Popular Wordsets," The Journal of Society for e-Business Studies, Vol. 21, No. 1, pp. 65-77, 2016.
- [7] Eckstein, A. and Pavlos P. P., "Assessment of Advanced Windowing Techniques for Digital Particle Image Velocimetry," Measurement Science and Technology, Vol. 20, No. 7, 2009.
- [8] Fayyad, U. and Irani, K., "Multi-Interval Discretization of Continuous-Valued Attributes for Classification Learning," Proceedings of the International Joint Conference on Uncertainty in AI, pp. 1022-1027, 1993.
- [9] Frodi, A. M. and Lamb, M. E., "Child abusers' responses to infant smiles and cries," Child development, pp. 238-241, 1980.
- [10] Healey, J. A. and Picard, R. W., "Detecting Stress During Real-World Driving Tasks Using Physiological Sensors," IEEE Transactions on intelligent transportation systems, Vol. 6, No. 2, pp. 156-166, 2005.
- [11] Jung, Y. and Yoon, Y. I., "Monitoring Senior Wellness Status Using Multimodal Biosensors," Proceedings of the 2016 International Conference on Big Data and Smart Computing (BigComp), pp. 435-438, 2016.
- [12] Kim, Y., Choi, Y., Kim, S., Park, K., and Park, J., "A Study on Method for User Gender Prediction Using Multi-Modal Smart Device Log Data," The Journal of Society for e-Business Studies, Vol. 21, No. 1, pp. 147-163, 2016.
- [13] Koo, C., Kim, H.-W., and Joun, Y. H., "A Study on the Explorative or Exploitive Use of Smartphone: Focusing on IS Continuance and Habit Theory," The Journal of Society for e-Business Studies, Vol. 16, No. 3, pp. 67-91, 2011.
- [14] Kurniawan, H., Maslov, A. V., and Pechenizkiy, M., "Stress Detection from Speech and Galvanic Skin Response Signals," Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems, pp. 209-214, 2013.
- [15] Seo, J. S., "Through Monitoring Solution Child Abuse and Children's Rights at Child Care Center," Convergence Security Journal, Vol. 13, No. 2, pp. 55-66, 2013.
- [16] Lee, K.-S., Park, J.-A., and Shin, Y.-J., "An Investigation into Mothers' Experience and Perception of Child Abuse at Day Care Centers and Need for Counseling," The Korean Journal of Woman Psychology, Vol. 20, No. 3, pp. 301-327, 2015.
- [17] Liu, Z., Wang, Y., and Chen, T., "Audio

- feature extraction and analysis for scene segmentation and classification,” *Journal of VLSI signal processing systems for signal, image and video technology*, Vol. 20, No. 1-2, pp. 61-79, 1998.
- [18] Lyons, R., “Understanding Digital Signal Processing,” Prentice Hall, 2001.
- [19] MaKay, C., “Automatic Music Classification with jMIR,” PhD thesis, McGill University, Canada, 2010.
- [20] Mcennis, D., McKay, C., Fujinaga, I., and Depalle, P., “jAudio: A Feature Extraction Library,” *Proceedings of the International Conference on Music Information Retrieval*, pp. 600-603, 2005.
- [21] Melillo, P., Bracale, M., and Pecchia, L., “Nonlinear Heart Rate Variability features for Real-life Stress Detection. Case Study: Students Under Stress Due to University Examination,” *Biomedical engineering online*, Vol. 10, No. 1, p. 96, 2011.
- [22] News1, “통학버스 왜 이리나...아동 방치 사고 10여일만에 또 사망사고,” [URL] <http://news1.kr/articles/?2743803>.
- [23] O'brien, S. J., “For Parents Particularly: Childhood Stress: A Creeping Phenomenon. Childhood education,” Vol. 65, No. 2, pp. 105-106, 1988.
- [24] Quinlan, J. R., “Induction of Decision Trees,” *Machine Learning*, Vol. 1, No. 1, pp. 81-106, 1986.
- [25] Rho, M., Kim, J., and Lee, J., “An Exploratory Study on Smart-Phone and Service Convergence,” *The Journal of Society for e-Business Studies*, Vol. 15, No. 4, pp. 59-77, 2010.
- [26] Roh, J.-H., Kim, H.-J., and Chang, J.-Y., “Improving Hypertext Classification Systems through WordNet-based Feature Abstraction,” *The Journal of Society for e-Business Studies*, Vol. 18, No. 2, pp. 95-110, 2013.
- [27] Rosales-Pérez, A., Reyes-García, C. A., Gonzalez, J. A., Reyes-Galaviz, O. F., Escalante, H. J., and Orlandi, S., “Classifying Infant Cry Patterns by the Genetic Selection of a Fuzzy Model,” *Biomedical Signal Processing and Control*, Vol. 17, pp. 38-46, 2015.
- [28] Ruvolo, P. and Movellan, J., “Automatic Cry Detection in Early Childhood Education Settings,” *Proceedings of the IEEE International Conference on Development and Learning*, Vol. 7, pp. 204-208, 2008.
- [29] Setz, C., Arnrich, B., Schumm, J., La Marca, R., Tröster, G., and Ehlert, U., “Discriminating Stress from Cognitive Load Using a Wearable EDA Device,” *IEEE Transactions on information technology in biomedicine*, Vol. 14, No. 2, pp. 410-417, 2010.
- [30] Sharma, N. and Gedeon, T., “Objective measures, sensors and computational techniques for stress recognition and classification: A survey,” *Computer methods and programs in biomedicine*, Vol. 108, pp. 1287-1301, 2012.
- [31] Sun, F. T., Kuo, C., Cheng, H. T., Buthpitiya,

S., Collins, P., and Griss, M., "Activity-aware Mental Stress Detection Using Physiological sensors," Proceedings of the International Conference on Mobile Computing, Applications, and Services, pp.

211-230, 2010.

[32] Womennews, "인천서 또 어린이집 폭행... 한 달 새 3건 발생 '충격'," [URL] <http://www.womennews.co.kr/news/view.asp?num=79850>.

저 자 소개



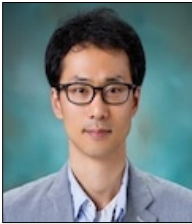
전유미
2016년
2016년~현재
관심분야

(E-mail: jym9425@gmail.com)
인천대학교 산업경영공학과 (학사)
인천대학교 산업경영공학과 (석사과정)
생체빅데이터 분석, 데이터 마이닝, 추천시스템



한태성
2017년
2017년~현재
관심분야

(E-mail: hantaeseong4@naver.com)
인천대학교 산업경영공학과 (학사)
인천대학교 산업경영공학과 (석사과정)
웹 데이터 분석, 데이터 마이닝



김관호
2006년
2012년
2013년
2014년~현재
관심분야

(E-mail: khokim@inu.ac.kr)
동국대학교 정보시스템전공 (학사)
서울대학교 산업공학과 (박사)
경희대학교 (연구박사)
인천대학교 산업경영공학과 교수
통계적 기계학습, 빅데이터