

A Study on Intelligent Skin Image Identification From Social media big data

Hyung-Hoon Kim*, Jeong-Ran Cho**

*Professor, Dept. of Cosmetic Science, Kwangju Womens University, Gwangju, Korea

**Professor, Dept. of Health Administration, Kwangju Womens University, Gwangju, Korea

[Abstract]

In this paper, we developed a system that intelligently identifies skin image data from big data collected from social media Instagram and extracts standardized skin sample data for skin condition diagnosis and management. The system proposed in this paper consists of big data collection and analysis stage, skin image analysis stage, training data preparation stage, artificial neural network training stage, and skin image identification stage. In the big data collection and analysis stage, big data is collected from Instagram and image information for skin condition diagnosis and management is stored as an analysis result. In the skin image analysis stage, the evaluation and analysis results of the skin image are obtained using a traditional image processing technique. In the training data preparation stage, the training data were prepared by extracting the skin sample data from the skin image analysis result. And in the artificial neural network training stage, an artificial neural network AnnSampleSkin that intelligently predicts the skin image type using this training data was built up, and the model was completed through training. In the skin image identification step, skin samples are extracted from images collected from social media, and the image type prediction results of the trained artificial neural network AnnSampleSkin are integrated to intelligently identify the final skin image type.

The skin image identification method proposed in this paper shows explain high skin image identification accuracy of about 92% or more, and can provide standardized skin sample image big data. The extracted skin sample set is expected to be used as standardized skin image data that is very efficient and useful for diagnosing and managing skin conditions.

▶ **Key words:** skin image identification, skin condition diagnosis, skin sample image, big data, artificial neural network

• First Author: Hyung-Hoon Kim, Corresponding Author: Jeong-Ran Cho
*Hyung-Hoon Kim (hhkim@kwu.ac.kr), Dept. of Cosmetic Science, Kwangju Womens University
**Jeong-Ran Cho (jrcho@kwu.ac.kr), Dept. of Health Administration, Kwangju Womens University
• Received: 2022. 08. 18, Revised: 2022. 09. 14, Accepted: 2022. 09. 19.

[요 약]

화장품 및 뷰티산업에서 고객 맞춤형 제품과 서비스를 제공하는 것은 주요 기술 트렌드이고, 피부상태 진단과 관리는 중요한 필수기능이다. 고객의 요구 수준은 더욱더 높아지고 있으며 이에 대한 다양하고 섬세한 고민과 요구 사항이 소셜미디어 커뮤니티에서 활발하게 다루어지고 있다. 소셜미디어 상의 이미지는 매우 다양하고 비정형적이므로 피부상태 진단 및 관리에 필요한 체계적인 피부 이미지 식별을 위한 시스템이 필요하다.

본 논문에서는 소셜미디어 인스타그램에서 수집한 빅데이터로부터 피부 이미지 데이터를 지능적으로 식별하고, 피부상태 진단 및 관리를 위한 정형화된 피부 샘플 데이터를 추출하는 시스템을 개발하였다. 본 논문에서 제안한 시스템은 빅데이터수집분석단계, 피부이미지분석단계, 훈련데이터준비단계, 인공지능망훈련단계, 피부이미지식별단계로 구성된다. 빅데이터수집분석단계에서는 인스타그램으로부터 빅데이터를 수집하고 피부 상태 진단 및 관리를 위한 이미지 정보를 분석결과로 저장한다. 피부이미지분석단계에서는 전통적인 이미지 처리 기법을 사용하여 피부 이미지의 평가 및 분석 결과를 획득한다. 훈련데이터준비단계에서는 피부이미지 분석결과로부터 피부 샘플 데이터를 추출하여 훈련데이터를 준비하였다. 그리고 인공지능망훈련단계에서는 이 훈련데이터를 사용하여 지능적으로 피부 이미지 유형을 예측하는 인공지능망 AnnSampleSkin을 단계별 고도화와 훈련을 통해 모델을 완성하였다. 피부이미지식별단계에서는 소셜미디어로부터 수집된 이미지에 대해 피부샘플을 추출하고, 훈련된 인공지능망 AnnSampleSkin의 이미지 유형 예측 결과들을 통합하여 최종 피부 이미지 유형을 지능적으로 식별한다.

본 논문에서 제안된 피부이미지식별 방법은 약 92% 이상의 높은 피부 이미지 식별 정확도를 나타내고 있고, 정형화된 피부 샘플 이미지 빅데이터를 제공할 수 있게 되었다. 추출된 피부샘플 세트는 피부 상태를 진단하고 관리하는데 매우 효율적이고 유용한 정형화된 피부 이미지 데이터로 사용될 것으로 기대된다.

▶ **주제어:** 피부이미지식별, 피부 상태 진단, 피부 샘플 이미지, 빅데이터, 인공지능망

I. Introduction

최근 화장품산업과 뷰티산업의 동향은 기능과 성분을 중시하며 기능성 화장품과 개인 맞춤형 뷰티산업 방향으로 발전되어 가고 있다. 개인 맞춤형 기능성 화장품과 피부미용 서비스를 제공하기 위해서는 정확한 피부 상태의 파악과 진단이 필요하다. 피부 상태 진단은 대부분 고객의 피부에 대한 육안 관찰과 피부 상태에 대한 고객 의견 분석으로 이루어진다. 고객의 요구 수준은 더욱더 높아지고 있으며 이에 대한 다양하고 섬세한 고민과 요구 사항은 직접 대면의 불편한 방식이 아닌 편리한 소셜미디어 커뮤니티와 같은 비대면 온라인 환경에서 활발하게 다루어지고 있다. 따라서 소셜미디어 커뮤니티와 같은 온라인 환경을 통한 고객의 피부 상태 진단 및 관리가 필요하다.

일반적으로 고객의 피부 상태에 대한 요구 사항은 매우 주관적이고 비전문적인 내용으로 표현된다. 따라서 온라인 비대면 환경에서 정확한 피부 상태 진단 및 관리를 위하여

고객의 객관적인 피부 이미지 분석이 필요하다. 고객의 피부 이미지는 피부 상태를 파악할 수 있는 가장 직접적인 정보이다. 이와 같은 이유로 기존의 뷰티산업 분야에서도 고객의 의견과 함께 피부의 육안 관찰을 통한 피부 상태 파악 및 개인 맞춤형 미용 서비스를 제공하고 있다.

화장품 및 피부미용과 같은 민감하고 개별적인 내용은 기존의 직접 대면 환경보다 소셜미디어 커뮤니티가 훨씬 자유롭고 다양한 정보가 교환되고 편리하게 응답을 받을 수 있는 장점이 존재하는 환경이다. 따라서 이와 같은 새로운 환경에서의 보다 개선된 화장품 및 피부미용 서비스 제공을 위한 소셜미디어 빅데이터 기반의 미래 지향적 시스템 개발이 요구된다. 즉, 소셜미디어 커뮤니티와 같은 다양한 온라인 환경에서 화장품 및 피부미용 관리에 필요한 빅데이터를 수집, 분석하여 사용자의 요구 사항을 파악하고 진단함으로써 보다 높은 수준의 피부미용 관리와 고

객과의 긴밀한 상호작용이 가능하다[7].

화장품 및 피부 미용에 대한 상호작용을 위해서는 피부 이미지와 고객의 의견을 효율적으로 다룰 수 있는 소셜미디어가 필요하다. 소셜미디어 인스타그램은 이미지를 중심으로 간단한 해시태그와 텍스트로 콘텐츠를 작성할 수 있는 대표적 커뮤니티이고, 화장품 및 피부 미용과 관련하여서도 많은 사용자와 콘텐츠를 확보하고 있다. 소셜미디어 인스타그램에서 사용되는 텍스트와 이미지는 매우 비정형적이며 다양한 내용과 형태의 콘텐츠로 작성되므로 수집된 빅데이터 정보는 화장품 및 피부미용 관리에 적합하게 분석처리 되어야 한다. 즉, 고객의 의견을 담고 있는 텍스트 콘텐츠에 대해서는 한국어 자연어처리와 분석을 통하여 화장품 및 피부미용과 관련한 중요한 주제와 단어를 분석하고 추출하는 것이 필요하다. 그리고 고객의 피부 상태를 포함하고 있는 이미지 콘텐츠에 대해서는 다양한 내용의 이미지로부터 피부 이미지를 식별하여 추출하고 이미지 내용에서 피부가 아닌 영역을 제거하여 피부 상태 관리의 사용 목적에 적합한 정형화된 피부 이미지로 분석 처리하는 것이 필요하다. 그러나 기존의 이미지 처리기술은 이미지 영상에서 피부 색상 정보만을 사용하여 인식함으로써 피부색과 유사한 색을 지닌 다른 이미지 영상이 포함되어 있는 경우 이를 식별하기 어려운 문제가 있다.

화장품 및 뷰티산업 분야에서 다양한 고객 니즈를 반영한 기능성 화장품과 맞춤형 서비스를 제공하기 위하여 소셜미디어에 게재되는 방대한 규모의 고객 피부 상태 이미지들을 지능적으로 식별하여 수집하고, 피부 상태 진단 등 과정에 사용 가능한 상태로 처리된 피부 이미지를 제공할 필요가 있다. 이에 본 논문에서는 소셜미디어 인스타그램으로부터 수집된 빅데이터로부터 지능적으로 화장품 및 피부 미용과 관련된 피부 이미지를 인식하여 수집하고, 정형화된 피부 샘플 이미지를 추출하는 인공지능 기반의 시스템을 연구하였다. 제안된 시스템으로 추출되는 피부 샘플 이미지는 개인 식별 정보가 포함되기 어려운 샘플 이미지 크기로 랜덤한 위치에서 생성되므로 개인정보보호법을 위배하지 않는 안전한 방식으로 사용될 수 있다. 제안된 인공지능 기반 피부 이미지 식별 시스템은 피부 이미지에 포함된 복합적인 특징을 학습하는 방법을 사용함으로써 기존의 피부 색상만을 사용한 피부 이미지 식별 방법보다 정확하게 피부 이미지를 인식할 수 있고 사용 목적에 따라 적응력있게 피부 이미지를 식별할 수 있는 지능화된 피부 이미지 식별 시스템으로 활용될 수 있다.

II. Related Research

2.1 Skin types in skin beauty care

최근 맞춤형화장품판매 제도가 시행됨에 따라 화장품과 뷰티산업의 제품과 서비스 패러다임이 고객 중심으로 변화 발전하고 있다. 뷰티산업분야에서 이와 같은 개인 맞춤형 화장품과 미용서비스를 제공하기 위해 고객의 요구와 피부 상태를 파악하는 것은 매우 기본적이며 중요한 관심사이다. 뷰티산업 분야에서는 고객의 피부 상태 관리, 피부 미용 목적으로 피부 상태를 진단하기 위한 많은 노력을 진행해 왔다. 고객의 피부 상태를 진단하기 위해서는 고객이 자각하는 피부 상태에 대한 요구 파악에 의한 주관적 평가와 다양한 피부 상태 측정 기기를 사용한 객관적 평가를 병행하는 방법이 필요하다. 그러나 피부 상태를 객관적으로 평가하는 정밀한 측정 기기는 고가의 비용문제와 특별한 환경을 요구하는 설치운영 환경 등의 이유로 대중화하는데 한계가 있다.

뷰티산업종사자들이 피부 미용 고객들을 대상으로 현재까지 일반적으로 많이 사용하고 있는 피부 상태 평가 방법은 육안으로 피부 각질층의 표면 상태를 관찰하는 방법과 문진에 의한 고객의 요구를 파악하는 방법을 병행하고 있다. 피부 각질층의 표면 상태는 피부 각질층 하부의 다양한 변화를 반영하고 있으므로 육안으로 피부상태를 관찰하여 피부 상태 유형을 판단하는 것은 타당성 있는 뷰티산업분야의 판별 방법이라 할 수 있다. 고객의 요구가 고급화되고 편리성이 강조되면서 인터넷 등을 활용한 보다 편리한 문진 방법과 스마트폰 등의 보편적이고 일반화된 기기를 사용한 피부 상태 이미지를 활용할 수 있는 정보처리 기술이 결합된 보다 개선된 피부 상태 진단 환경이 요구되고 있다.

피부 각질층에 대한 육안 판단으로 분류되는 피부 유형은 크게 피부 표면의 형태적 요소에 의한 분류와 피부표면의 색채적 요소에 의한 분류로 나누어진다. 피부표면의 형태적 요소에는 피부결(skin texture), 모공(pore), 각질(scale), 주름(winkle)이 있고, 색채적 요소에는 피부톤(skin tone), 홍반(홍조, redness), 여드름(흑색면포, 백색면포, black head, white head), 색소침착(pigmentation)이 있다[1, 2]. 피부 상태의 다른 분류 방법으로는 피부 각질층의 수분함량, 피지 분비량 등을 기준으로 건성, 지성, 중성, 민감성, 복합성의 5가지 종류로 분류하는 방법도 사용되고 있다[3].

2.2 Skin image processing technology

화장품 및 뷰티산업의 기술적 요구를 해결하기 위해 피부 이미지 인식과 처리에 인공지능 등 정보처리기술을 활용한 다양한 연구가 진행되고 있다.

고객의 피부 상태를 자동적으로 진단하기 위하여 기계 학습 인공지능 기술을 활용한 피부 이미지 진단 연구가 진행되어 왔다. 기존 연구에서는 인공지능기술을 사용한 고객의 피부 상태 진단을 위하여 고객의 피부 이미지로부터 ORB(Oriented FAST and Rotated BRIEF) 처리를 사용하여 피부 이미지에 대한 특징점들을 추출하고, 이 데이터를 인공지능 기계학습 분류 방법인 SVM을 사용하여 피부 상태를 건성과 지성으로 진단한다[3]. 피부 이미지 훈련 데이터를 선형적 분류 알고리즘인 SVM(Support Vector Machine)으로 기계학습을 진행하고, 피부 이미지를 지성과 건성으로 분류하는 피부 상태 진단 모델에 관한 연구를 진행하였다.

이 연구 결과를 실질적으로 사용하기 위해서는 기계 학습 훈련에 사용할 충분한 규모의 피부 이미지를 확보하는 문제를 해결해야 한다. 그리고 방대한 규모의 피부 이미지로부터 기하학적 특징점을 추출하는데 많은 시간적 비용 부담이 발생 될 수 있고, 연구에서 제시한 기하학적 특징점만으로 실제 사용되는 다양한 피부 상태 유형을 진단하는 모든 특징을 추출하기는 어렵다는 문제점이 있다. 또한 뷰티산업분야에서 사용하는 피부 상태 진단 유형은 선형적 분류 알고리즘과 분류 방법보다는 여러 가지 복합적인 특성을 고려하여 진단할 수 있는 비선형적 분류 알고리즘과 진단 방법을 요구한다.

피부 이미지를 식별하고 사용하기 위해서는 얼굴 이미지 탐지와 피부 영역 검출 방법이 필요하다. [4]에서는 특정한 오브젝트를 인식하기 위해 특정 객체의 특징을 기계 학습 기반으로 하여 검출하는 Harr Cascade 알고리즘을 연구하였다. 이 연구에서는 얼굴을 탐지할 수 있는 Haar Features를 획득하기 위해 얼굴이 포함된 이미지와 얼굴이 포함되지 않은 이미지를 사용하여 얼굴 인식을 위한 특징 분류기를 학습시키는 방법이다. 이 방법은 이미지를 픽셀 단위가 아닌 직사각형 영역 단위로 처리하기 때문에 픽셀 단위 이미지 처리 방법 보다 동작 속도가 매우 빠르다는 특징을 갖는다.

소셜미디어에 게시되는 이미지에는 얼굴이 포함되어 있지 않고 신체 일부만 존재하거나 피부 이외의 다양한 내용을 포함하고 있으므로 피부 이미지를 사용하기 위하여 피부 영역만 추출하는 이미지 처리기술이 필요하다. [5, 6]에서는 피부 이미지를 검출하기 위해 피부 이미지의 컬러색

상 모델인 RGB를 피부색 분류에 적합한 YCbCr 컬러색상 모델로 변형하여 피부에 대한 색상 범위를 적용한다. 기본적으로 피부 영역으로 선택하는 색상 범위는 $133 < Cr < 173$, $77 < Cb < 127$ 이다. 그러나 피부 색상 정보를 중심으로 피부 영역을 검출하는 방법은 피부색과 유사한 색상을 갖는 이미지에 대해서는 올바르게 피부 이미지를 인식하지 못하고, 온전한 피부 영역을 검출하지 못하는 한계점이 있다. 그러므로 피부 이미지 인식과 피부 영역 검출에 대한 보다 지능적이고 효율적인 방법이 요구된다.

III. The Proposed System

3.1 Overview

화장품 및 뷰티산업분야에서 요구되는 소비자 중심의 맞춤형 화장품 및 미용 서비스를 제공하기 위해 고객의 피부 상태 진단이 반드시 필요하고, 생활수준 향상과 함께 점점 높아지는 소비자의 욕구를 충족시키기 위해 인공지능 기술을 접목한 피부 상태 진단 도구 및 환경이 필요하다. 그러나 피부 상태 진단에 인공지능 기계학습 기술을 사용하기 위해서는 충분한 규모의 훈련데이터, 즉 빅데이터로부터 훈련데이터를 확보할 수 있어야 한다.

현재 소셜미디어는 거의 일반인 모두에게 일상생활에서 다양한 분야의 정보교환 및 상호작용 매체로 사용되고 있다. 특정한 소셜미디어 커뮤니티에서는 화장품 및 피부미용에 관련하여서도 다양한 사용자들의 의견과 정보가 생산되고 교류되어 자연스럽게 화장품 및 피부 미용에 대한 빅데이터 정보생산지 역할을 하고 있다. 그러나 소셜미디어 커뮤니티에서 교류되는 정보는 매우 비정형적이고 데이터 사용 목적에 부합되지 않는 다양한 내용과 형태를 포함하고 있으므로 특정한 목적으로 사용하기 위해서는 적절한 분석 처리가 필요하다. 즉, 소셜미디어 커뮤니티의 피부 이미지 정보를 화장품 및 피부미용 목적의 피부 상태 진단 및 관리에 사용하기 위해서는 그 용도에 적합하도록 빅데이터에 대한 전처리 작업이 필요하다[7].

본 논문에서는 소셜미디어 커뮤니티로부터 화장품 사용 및 피부 상태 관리에 필요한 피부 이미지를 자동적으로 식별하여 수집하고, 피부 상태 진단에 적합한 정형화된 피부 샘플을 추출하는 인공지능망을 기반으로 한 시스템을 개발하였다. 제안된 시스템에서는 먼저 소셜미디어에서 수집된 이미지에 대하여 기존에 연구된 얼굴인식과 피부색 기반의 피부 이미지 인식 방법을 활용하여 기초적인 피부 이미지들을 수집하였다. 그리고 피부 이미지 분석 정보와 정

형화된 피부 샘플 이미지를 훈련데이터로 사용하여 피부 이미지를 지능적으로 식별할 수 있도록 하였다. 개발된 인공신경망은 피부 이미지에 포함된 복합적인 특징을 학습함으로써 기존의 피부 색상만을 사용한 피부 이미지 식별 방법보다 다양한 형태의 피부 이미지를 정확하게 인식할 수 있고, 사용 목적에 따라 적응력 있게 피부 이미지를 식별할 수 있는 지능화된 피부 이미지 식별 시스템으로 활용될 수 있다. 개발된 시스템의 결과물인 정형화된 피부 샘플 이미지는 고객의 피부 상태를 진단하기 위한 향후 추가적 연구 분야에서 다양하게 활용될 수 있다.

본 논문에서 제안한 시스템은 Fig. 1과 같이 빅데이터수집분석단계, 피부이미지분석단계, 훈련데이터준비단계, 인공신경망훈련단계, 피부이미지식별단계로 구성된다. 첫 번째 단계인 빅데이터수집분석단계에서는 소셜미디어 커뮤니티로부터 빅데이터를 수집분석하여 텍스트와 이미지를 저장한다. 피부이미지분석단계에서는 전통적인 이미지 처리 기법을 사용하여 피부 이미지의 평가 및 분석 결과를 획득한다. 훈련데이터준비단계에서는 피부 샘플 데이터를 추출하여 훈련데이터를 준비하고, 인공신경망훈련단계에서는 인공신경망 AnnSampleSkin을 빌드업하고 훈련함으로써 지능적으로 피부 이미지를 식별할 수 있도록 한다. 마지막 단계인 피부이미지식별단계에서는 소셜미디어로부터 수집된 임의의 이미지에 대하여 피부샘플을 추출하고 피부이미지를 식별한다.

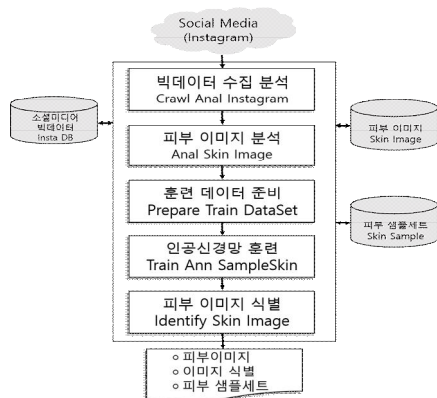


Fig. 1. Overview

3.2 Detailed Proposed System

본 논문에서 제안한 시스템의 각 단계에 대한 구체적인 내용은 다음과 같다.

3.2.1 Big data collection and analysis stage

빅데이터수집분석단계(crawlAnalInstagram)는 피부 상태 관리를 위한 빅데이터를 소셜미디어 인스타그램으로부터 수집하고 분석하여 텍스트와 이미지 결과를 저장하는 단계이다. 피부상태 관리를 위한 빅데이터를 수집하기 위해 피부표면의 형태적 요소와 색채적 요소에 해당하는 키워드를 사용하여 인스타그램에 게시된 텍스트와 이미지를 분석하고 결과를 파일로 저장한다[7]. 빅데이터수집분석은 빅데이터수집, 텍스트전처리, 텍스트단어분석 과정으로 진행된다. 빅데이터수집은 인스타그램을 대상으로 피부 표면의 형태적 평가요소와 색채적 평가요소를 검색 키워드로 검색하여 데이터를 수집하였다. 인스타그램에서 수집된 데이터는 게시글 본문, 해시태그, 게시된 날짜, 좋아요, 댓글로된 텍스트 데이터와 게시된 이미지 데이터로 구성된 빅데이터 정보이다.

텍스트전처리는 수집된 각 게시물의 데이터 가운데 텍스트 데이터인 본문 내용, 해시 태그, 댓글 내용, 댓글 해시 태그를 대상으로 Komoran 한국어 문장의 형태소 분석 과정을 진행하여 단어를 분리 추출하고 각 단어의 품사 유형을 구분하여 피부 상태 분석에 적합한 형태의 단어를 추출하여 저장한다[8, 9, 10, 11, 12]. 텍스트단어분석은 추출된 단어의 빈도수 및 상관성분석, 동시출현 분석을 진행하고, 피부 상태 평가 요소별 핵심단어 추출과 시기별, 지역별 상관성을 분석하였다[13, 14].

3.2.2 Skin image analysis stage

피부이미지분석단계(analSkinImage)에서는 인스타그램에서 수집분석된 이미지에 대하여 이미지 처리기술로 분석하여 얼굴피부(face skin), 바디피부(body skin) 영역으로 분류하여 탐지하고, 검출된 피부 이미지와 분석정보를 결과로 저장한다. 피부이미지분석단계의 목적은 빅데이터 이미지로부터 피부 이미지를 분별하여 추출함으로써 인공신경망의 피부이미지 훈련데이터에 대한 소스를 생성하는 것이다. 피부이미지분석단계는 Fig. 2와 같이 얼굴이미지탐지(detectFaceArea), 대표얼굴이미지추출(cropFaceImage), 피부영역마스킹(maskSkinArea), 피부이미지유형분석(analSkinImage) 과정으로 처리된다. 얼굴이미지탐지과정은 수집된 이미지 데이터로부터 얼굴이미지가 존재하는지 탐지하여 얼굴 영역의 위치정보를 얼굴을 포함하는 바운딩 박스의 좌표 정보를 근거로 해당영역의 이미지를 추출한다.



Fig. 2. Skin image analysis process

얼굴이미지탐지는 Haar Features[4]를 이용하여 처리한다. 대표얼굴이미지추출과정에서는 이미지 안에 1개 이상의 얼굴이 존재하는 경우 실험을 간소화하기 위해 탐지된 얼굴 영역 가운데 가장 큰 얼굴만을 분석할 얼굴 이미지로 추출한다. 이때 소스이미지 크기 대비 추출된 대표얼굴 이미지 크기의 비율에 대한 정보를 얼굴분석정보(facelmgProc) 값으로 산출하여 저장한다.

피부영역마스크과정은 이미지로부터 피부 이미지만 남기고 피부 외의 모든 요소는 제거하기 위해 마스크하는 과정이다. 즉, 탐지된 얼굴 이미지 또는 기타 신체에서 머리, 카락, 눈썹 또는 배경 등 피부 이외의 내용을 제거하는 마스크를 진행한다. 피부영역마스크과정에서는 피부 이미지의 컬러색상 모델인 RGB를 피부색 분류에 적합한 YCbCr 컬러색상 모델로 변형하여 처리한다. 본 논문에서 피부영역으로 선택하는 색상 범위는 $133 < Cr < 173, 77 < Cb < 127$ 값의 색상 범위[5, 6]를 사용하였다. 그 외 색상 범위는 피부가 아닌 영역으로 제거하는 마스크 처리를 하였다.

Table 1. Skin image types by image analysis

skinImgLabel	faceImgProc	skinImgProc
No skin(0)	≤ 0	$\leq \text{skinRatio}$
Face skin(1)	> 0	-
Body skin(2)	≤ 0	$> \text{skinRatio}$

피부이미지유형분석과정은 Table 1과 같은 피부이미지 유형 분류 조건을 사용하여 피부 이미지유형을 결정한다. 피부영역마스크과정의 결과 피부 이미지로 사용하기에 적합하지 않은 작은 조각 형태의 노이즈성 피부 영역을 제외하기 위해 연속적으로 연결된 피부 영역을 바운딩박스 좌표로 탐색 하고, 바운딩박스 내부의 피부 색상 요소의 비율(skinImgProc)이 임의로 지정된 값(skinImgProc>0.2) 이상의 조건을 만족하는 영역만을 피부유형분석 대상으로 사용한다. Table 1에서 skinImgLabel은 피부이미지유형의 종류인 “No skin”, “Face skin”, “Body skin”을 타내내고, facelmgProc은 얼굴이 탐지되지 않은 경우는 0값,

얼굴이 탐지된 경우는 이미지 대비 탐지된 얼굴영역 크기 비율 값(>0)을 나타낸다. 피부이미지유형분석과정의 분석 결과인 skinImgLabel은 Table 1에 나타난 것과 같은 분류 기준을 사용하여 결정된다. 얼굴과 피부를 포함하지 않는 경우(facelmgProc ≤ 0 and skinImgProc $\leq \text{skinRatio}$)에는 “No skin” 값으로, 얼굴을 포함하는 피부인 경우 (facelmgProc>0)에는 “Face skin” 값으로, 얼굴이 인식되지 않은 상태에서 특정 비율 이상의 피부 색상을 갖는 경우(facelmgProc ≤ 0 and skinImgProc>skinRatio)에는 “Body skin” 값으로 분류되어 저장된다.

3.2.3 Training data preparation stage

훈련데이터준비단계(prepareTrainDataset)는 피부 이미지를 지능적으로 식별할 수 있는 인공신경망을 훈련시키기 위해 사용될 훈련데이터셋을 준비한다. 이 단계는 정형화된 피부샘플을 추출하는 과정과 훈련데이터셋을 준비하는 과정으로 진행된다.

1) skin sampling process

피부샘플을 추출하는 과정은 훈련데이터의 정형화된 이미지 소스를 생성하는 과정이다. 성공적인 인공신경망을 빌드업하고 훈련하기 위해서는 충분한 분량의 훈련데이터 소스가 필수적이다. 피부샘플추출과정에서는 각 이미지에 대한 피부샘플들을 추출하여 Fig. 3과 같은 구조의 피부샘플세트이미지를 생성한다.

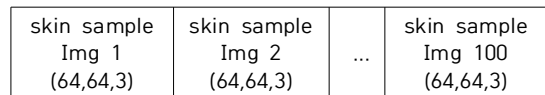


Fig. 3. Structure of skin sample set image

피부샘플세트이미지는 최대 피부샘플 개수(num_sampl e=100) 만큼의 피부샘플이미지(skin sample Img)들로 구성된다. 각 피부샘플이미지는 향후 연구 과정에서 이미지 패턴을 분석하고 사용하는 환경과 목적에 따라 변경가능 하며 본 논문에서는 현 실험 환경을 고려하여 임의로 설정된 크기(sample_shape = (64, 64, 3))로 이미지의 랜덤한 위치에서 샘플링을 시도하여 피부색상만으로 구성된 이미지만 피부샘플세트이미지에 포함한다.

Table 2는 수집된 각 계시물의 소스 정보 keyword, srcIdx 항목에 대하여 이전 단계인 피부이미지분석단계의 결과로 산출된 항목인 skinImgProc, facelmgProc, skinI mgLabel에 추가하여, 현 단계인 피부샘플추출과정의 결과로 생성된 피부샘플세트이미지를 저장하고 있는 imgFil e 항목을 나타내고 있다.

Table 2. Skin sample set image data

index	keyword	src Idx	skinImg Proc	faceImg Proc	skinImg Label	imgFile
0	피부결	0	0.554219	0.265327	1	5_0.jpg
1	여드름	2	0.320884	0	2	6_0.jpg
...

2) Training data set preparation process

훈련데이터세트를 준비하는 과정에서는 피부샘플세트 이미지 데이터로부터 인공신경망을 빌드업하고 훈련하는 과정에 사용할 훈련데이터세트를 생성한다. 훈련데이터세트의 각 개별 데이터는 입력 값인 피부샘플이미지 X (shape=(# of sample, 64, 64, 3))와 출력 값인 이미지 유형 라벨 y (shape=(# of sample, label), label="No skin" or "Yes skin") 요소로 구성된다. 훈련데이터세트의 데이터는 매 훈련과정마다 랜덤하게 훈련데이터(X_train, y_train)와 테스트데이터(X_test, y_test)로 구분되어 사용된다.

훈련데이터는 인공신경망을 빌드업하고 훈련하는 데 사용되고, 전체 데이터의 80%를 랜덤하게 선택하여 준비한다. 그리고 훈련데이터의 10%는 검증데이터(X_valid, y_valid)로 사용되고, 이는 인공신경망 훈련의 종료 여부를 판단하기 위해 사용된다. 테스트데이터는 훈련된 인공신경망을 평가하는데 사용되고, 전체 데이터의 20%를 랜덤하게 선택하여 준비한다. 이와 같이 훈련데이터와 구분된 테스트데이터를 인공신경망 평가에 사용하는 이유는 훈련 과정에서 사용되지 않은 새로운 이미지에 대하여 모델의 예측 정확성을 평가함으로써 실제 현장에서 사용될 때 예상되는 모델의 성능을 보다 객관적으로 평가하고 분석하기 위한 것이다.

3.2.4 Artificial neural network training stage

인공신경망훈련단계(trainAnnSampleSkin)는 모델 빌드업과정과 모델피팅과정으로 진행된다. 모델 빌드업과정에서는 사용할 인공신경망 모델 AnnSampleSkin을 설계하고 구현한다. 그리고 모델피팅과정에서는 준비된 훈련데이터세트를 사용하여 학습률 등의 하이퍼 파라미터를 적절하게 설정하고 훈련을 반복 진행함으로써 목표하는 성능을 갖는 훈련된 모델을 획득한다.

인공신경망 모델 빌드업과정은 텐서플로우(Tensorflow)의 딥러닝 프레임워크와 케라스(Keras) 라이브러리를 사용하여 인공신경망 AnnSampleSkin 모델의 구조를 설계하고 구현하였다. Fig. 4는 인공신경망 AnnSampleSkin 모델에 대한 개념적 구조를 나타내고 있다.

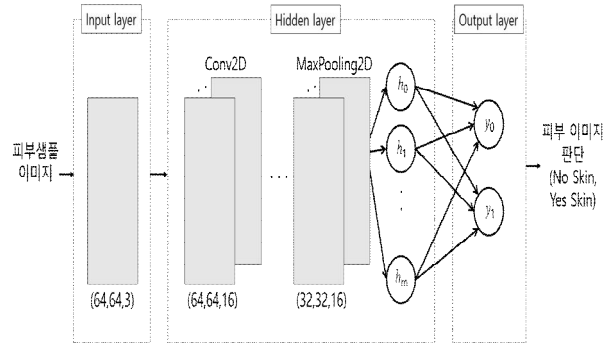


Fig. 4. Conceptual architecture of AnnSampleSkin

입력층(Input Layer)은 추출된 샘플 이미지를 입력 받을 수 있도록 이미지의 형식(shape=(64, 64, 3))을 반영한 구조의 뉴런으로 구성된다. 출력층(Output Layer)은 샘플 이미지의 종류인 "No skin", "Yes skin"의 예측 값을 출력할 수 있도록 2개의 뉴런으로 구성된다. 은닉층(Hidden Layer)은 반복적인 훈련 과정을 통하여 목표하는 성능을 달성하기에 적합한 은닉층 개수, 은닉층 유형, 뉴런의 개수 등이 결정되어 구성된다. 은닉층의 유형은 이미지 패턴 학습에 적합한 Conv2D, MaxPooling2D 등으로 구성되는 합성곱 계층 유형과 이미지 유형 인식 결과를 얻기 위해 이미지 정보를 1차원 벡터로 만드는 평탄화 계층과 전통적인 계층 유형으로 구성된다.

인공신경망 모델 AnnSampleSkin의 피팅과정은 준비된 훈련데이터를 사용하여 반복적인 훈련 과정에서 학습률, 활성화함수 등의 인공신경망에 대한 하이퍼파라미터를 수정, 보완하여 최적의 예측 능력을 나타내는 인공신경망 모델을 확정하는 과정이다. AnnSampleSkin의 예측 결과가 훈련데이터의 정답과 일치하는 경우에는 합성곱계층의 필터 특징들과 일반 계층의 가중치를 강화하고, 오답인 경우에는 정답이 나오도록 필터 특징들과 가중치를 수정하여 점진적으로 인공신경망 모델의 예측 정확도를 높여가는 과정이 반복되며, 이 과정에서 모델의 성능과 훈련의 효율성을 높이기 위해 적절한 모델의 하이퍼 파라미터를 설정하는 것이 필요하다.

```
# compile model
model.compile(loss='sparse_categorical_crossentropy',
              optimizer='adam', metrics=['accuracy'])
# fitting model
model.fit(X_train, y_train, batch_size=200, epochs=32,
         verbose=1, validation_data=(X_valid, y_valid))
```

Fig. 5. The codes for fitting model AnnSampleSkin

Fig. 5는 인공신경망 모델 AnnSampleSkin의 빌드업 과정에서 설정된 모델의 하이퍼 파라미터인 손실함수(loss

= 'categorical_crossentropy'), 옵티마이저(optimizer='adam'), 배치크기(batch_size=200), 에포크(epochs=32)를 포함한 모델 피팅 프로그램 코드의 일부이다. 손실함수 "sparse_categorical_crossentropy"는 모델이 예측한 이미지 유형 결과 값과 정답과의 차이 값을 산출하는 함수이다. 인공신경망 훈련과정은 손실함수에 의하여 산출된 차이 값을 최소화하여 모델의 성능을 최적화시키기 위해 훈련데이터를 사용하여 가중치를 반복적으로 수정한다. 옵티마이저 "Adam"은 손실함수를 미분하여 훈련 중인 모델의 가중치를 변경하는 방법으로 사용된다. "Adam" 옵티마이저는 최근의 그래디언트를 사용함으로써 전역적 최저점을 효과적으로 찾아가는 방법으로 현재 가장 많이 사용되는 옵티마이저이다.

3.2.5 Skin image identification stage

피부이미지식별단계(identifySkinImage)에서는 소셜미디어로부터 수집된 이미지에 대해 피부이미지분석단계와 피부샘플추출과정을 진행하고 훈련된 인공신경망 AnnSampleSkin의 피부 샘플들에 대한 예측 값들을 통하여 이미지 유형을 식별한다. 즉, 인공신경망 AnnSampleSkin은 피부이미지분석을 통해 획득한 100개의 피부 샘플에 대한 이미지 유형을 "No skin"과 "Yes skin" 가운데 하나로 분류한다. 피부이미지식별단계에서는 100개 피부 샘플에 대한 예측된 값에서 상대적으로 더 많은 비율을 차지하는 이미지 유형을 식별 결과로 결정한다.

Fig. 6은 피부 이미지 식별을 위한 프로그램 코드의 일부이다. Fig. 6에서 X_images는 소셜미디어의 특정 이미지로부터 피부이미지분석단계를 통하여 랜덤하게 추출된 샘플 이미지세트이다. prediction은 훈련된 인공신경망 AnnSampleSkin이 예측한 이미지 유형의 결과 값을 갖고 있다. 피부이미지식별단계의 이미지 식별 결과(prdIdx)는 피부 샘플 이미지 세트에 대한 예측 값 prediction에서 상대적으로 더 많은 비율을 차지하는 이미지 유형으로 결정된다.

```
# predict for sample image
prediction = model.predict(X_images)

# identify skin image
prdIdxs = [prd.argmax() for prd in prediction]
noSkin = prdIdxs.count(0)
yesSkin = prdIdxs.count(1)
prdIdx = 0 if noSkin > yesSkin else 1
```

Fig. 6. Code for skin image identification

IV. Analysis on Proposed System

본 논문에서는 소셜미디어 인스타그램으로부터 지능적으로 피부 이미지를 식별하고 피부 샘플 이미지를 추출하는 인공신경망기반 시스템을 개발하고 그 결과를 분석하였다. 제안된 시스템 개발을 위해 파이썬(Python) v3.8 프로그래밍 언어와 인공지능 기계학습 및 데이터분석에 적합한 아나콘다3(Anaconda3), 텐서플로우(TensorFlow), 케라스(Keras) 개발도구를 활용하였다.

4.1 Results of Skin Image Analysis

피부이미지분석단계는 소셜미디어로부터 수집된 빅데이터 이미지로부터 피부 이미지 분석 정보를 제공함으로써 피부 상태 진단에 필요한 빅데이터 피부 이미지를 사용할 수 있도록 한다. 피부이미지분석단계의 결과는 소셜미디어의 빅데이터로부터 피부 이미지를 지능적으로 판단할 수 있는 인공신경망의 훈련데이터에 대한 소스로 활용하였다.

		faceImgProc	0.85
		skinImgProc	0.25
		skinImgLabel	1
		faceImgProc	0.61
		skinImgProc	0
		skinImgLabel	2
		faceImgProc	0.66
		skinImgProc	0
		skinImgLabel	2

Fig. 7. Results of skin image analysis

Fig. 7은 피부이미지분석단계의 처리 결과를 나타내고 있다. Fig. 7에서 (a)는 소셜미디어로부터 수집된 원본 이미지, (b)는 피부이미지분석단계에서 추출된 피부이미지, (c)는 피부이미지분석단계에서 산출된 이미지분석정보와 이미지 유형 결과를 나타낸다. (a)에 나타난 원본 이미지는 피부 상태 진단 및 관리에 적합하지 않은 다양한 내용을 포함하고 있다. (b)에 나타난 피부이미지는 (a)의 원본 이미지로부터 얼굴이미지를 추출하고, 피부 영역만 남기는 피부영역마스킹처리를 진행한 결과이다. (c)에 나타난 faceImgProc와 skinImgProc은 얼굴영역에 대한 정보와 피부영역에 대한 정보이다. 얼굴영역에 대한 정보는 얼굴 영역의 원본 이미지 대비 얼굴영역 크기의 비율이고, 피부영역

역에 대한 정보는 피부 색상 영역 크기의 비율이다. (c)에서 skinImgLabel은 이미지 유형 결과를 “No skin”은 0, “Face skin”은 1, “Body skin”은 2로 나타내고 있다.

4.2 Results of training data preparation stage

훈련데이터준비단계에서는 피부이미지분석에 적합한 정형화된 피부샘플과 인공신경망 AnnSampleSkin을 훈련시키기 위한 훈련데이터셋을 생성하고 준비하였다. 훈련데이터준비단계의 첫 번째 과정은 피부이미지분석단계에서 추출된 피부이미지로부터 정형화된 피부샘플을 추출하였다.

Fig. 8은 비정형적인 이미지로부터 피부상태 진단 및 관리에 사용하기에 적합한 정형화된 피부샘플이미지를 추출한 결과를 나타내고 있다. Fig. 8에서 (a)는 피부이미지분석단계에서 추출된 피부이미지이고, (b)는 랜덤하게 샘플링된 피부샘플(shape=(64, 64, 3))들로 구성된 샘플이미지(10개만 표시)를 나타내고 있다. 피부샘플이미지는 각 피부이미지에 대해 최대 100개까지 랜덤하게 샘플링된 샘플이미지이다.

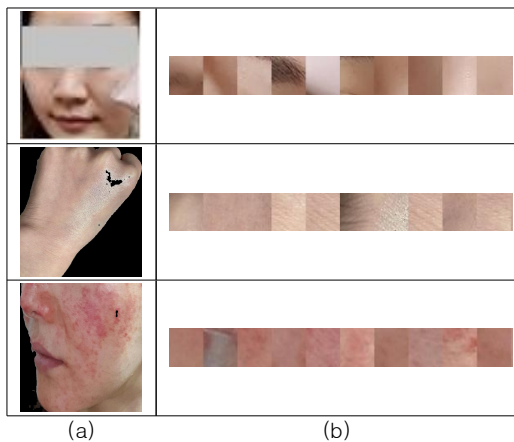


Fig. 8. Results of skin sampling

훈련데이터준비단계의 두 번째 과정은 피부 이미지를 지능적으로 식별하기 위한 인공신경망 AnnSampleSkin을 빌드업하고 훈련하는데 사용할 훈련데이터셋을 준비한다. 훈련데이터셋은 피부 이미지를 식별할 목적으로 사용하기 위하여 피부 이미지인 “Yes skin”과 피부가 아닌 이미지인 “No skin”으로 라벨링된 훈련데이터셋으로 준비한다. 즉, 피부이미지분석단계에서 결정된 이미지유형 (skinImgLabel)의 값을 활용하여 피부샘플데이터를 이미지유형이 0(No skin)인 경우에는 “No skin”으로, 이미지유형이 1, 2(Face skin, Body skin)인 경우에는 “Yes skin”으로 라벨링된 피부이미지 훈련데이터셋을 준비한다.

훈련 샘플 X_train 데이터 크기: (21600, 64, 64, 3)
훈련 샘플 y_train 레이블 크기: 21600
검증 샘플 X_valid 데이터 크기: (2400, 64, 64, 3)
검증 샘플 y_valid 레이블 크기: 2400
테스트 샘플 X_test 데이터 크기: (5999, 64, 64, 3)
테스트 샘플 y_test 레이블 크기: 5999

Fig. 9. Summary of prepared training dataset

Fig. 9는 인공신경망 AnnSampleSkin을 훈련하고 평가하기 위해 훈련데이터준비단계에서 준비된 훈련데이터셋 구성에 대한 요약이다. 전체 훈련데이터셋은 29,999개가 준비되었으며, 훈련샘플데이터(X_train, y_train)와 검증샘플데이터(X_valid, y_valid)는 인공신경망을 훈련하고 검증하기 위해 24,000개(80%)가 랜덤하게 선택되어 사용되고, 테스트샘플데이터(X_test, y_test)는 인공신경망을 평가하기 위해 5,999개(20%)가 랜덤하게 선택되어 사용된다. 인공신경망 모델 AnnSampleSkin의 반복 훈련 과정은 준비된 훈련샘플데이터와 검증샘플데이터로 목표하는 훈련 성능과 검증 성능을 달성하면 종료한다. 테스트샘플데이터는 훈련샘플데이터와 검증샘플데이터에 포함되어 있지 않는 데이터로 구성되어 있어 훈련이 종료된 인공신경망 AnnSampleSkin의 성능을 객관적으로 평가하는 데 사용된다.

4.3 Results of AnnSampleSkin training and evaluation

인공신경망훈련 및 평가단계에서는 준비된 훈련데이터셋을 사용하여 반복적으로 훈련과 평가를 진행한다. 이 과정에서 인공신경망 모델의 하이퍼 파라미터를 변경 적용하며 성능을 분석함으로써 목표하는 성능을 갖는 인공신경망 AnnSampleSkin의 구축을 완성한다.

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 64, 64, 16)	448
max_pooling2d (MaxPooling2D)	(None, 32, 32, 16)	0
conv2d_1 (Conv2D)	(None, 32, 32, 16)	2320
max_pooling2d_1 (MaxPooling2D)	(None, 16, 16, 16)	0
conv2d_2 (Conv2D)	(None, 16, 16, 16)	2320
flatten (Flatten)	(None, 4096)	0
dense (Dense)	(None, 1024)	4195328
dense_1 (Dense)	(None, 512)	524800
dense_2 (Dense)	(None, 256)	131328
dense_3 (Dense)	(None, 128)	32896
dense_4 (Dense)	(None, 32)	4128
dense_5 (Dense)	(None, 2)	66
Total params: 4,893,634		
Trainable params: 4,893,634		
Non-trainable params: 0		

Fig. 10. Summary of the built model AnnSampleSkin

Fig. 10은 구축된 인공신경망 모델 AnnSampleSkin에 대한 요약을 summary함수에 의해 출력된 결과를 나타내고 있다. 입력층(Input Layer)의 conv2d는 피부샘플데이터(input_shape=(64, 64, 3))를 입력받아 Conv2D 레이어와 16개의 필터로 구성된 합성곱 계층으로 구성되어 있다. 출력층(Output Layer)의 dense_5는 입력된 피부샘플에 대한 이미지 유형의 예측 결과 값을 나타낼 수 있도록 2개의 뉴런으로 구성되어 있다. 출력층의 2개의 뉴런은 이미지 유형의 종류인 “No skin”과 “Yes skin”에 대한 각 결과 값을 나타낸다. 은닉층(Hidden Layer)은 max_pooling2d로부터 conv2d_2까지의 합성곱단계와 flatten의 평탄화과정과 이후 다층 퍼셉트론구조로 구성하였다. 인공신경망 훈련 단계에서는 빌드업된 인공신경망 AnnSampleSkin의 훈련과정을 통하여 Fig. 10과 같이 합성곱단계와 평탄화단계의 각 계층에 대한 필터와 가중치의 파라미터 4,893,634개를 학습 결과로 취득하게 된다.

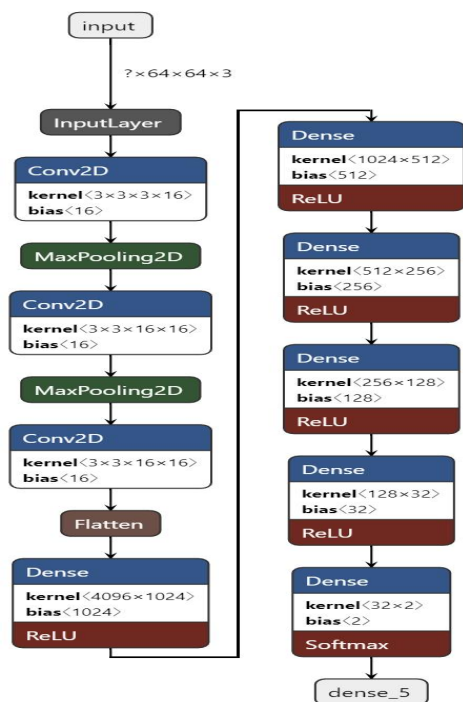


Fig. 11. Architecture of AnnSampleSkin

Fig. 11은 구축된 인공신경망 모델 AnnSampleSkin의 구조를 시각화하여 나타내고 있다. Fig. 11에서 입력층은 샘플링된 이미지(shape=(64, 64, 3))를 입력받아 인공신경망 AnnSampleSkin이 이미지 분류를 시작하는 계층의 구조를 보여주고 있다. Keras의 Conv2D와 Maxpooling2D 유형의 합성곱계층들과 Flatten 유형, Dense 유형의 평탄화된 계층은 은닉층의 구조를 나타내고 있다. 각 은닉층의

활성화함수는 ReLU함수를 적용하였다. 활성화함수 ReLU 함수는 AnnSampleSkin과 같은 규모의 다층신경망에서 간단한 계산방식으로 빠른 처리속도를 유지하고 그래디언트 감쇠가 일어나지 않는 적합한 성능을 나타내어 본 논문의 실험에서 채택되었다. 출력층은 본 논문에서 목적하는 학습된 피부이미지 식별 방법으로 소셜미디어의 이미지에 대하여 신속하고 정확하게 게시된 이미지로부터 추출된 샘플 이미지에 대하여 “No skin”, “Yes skin”의 샘플 이미지의 유형을 식별하기 위해 이에 적합한 2개의 뉴런을 갖는 Dense 유형과 Softmax 활성화 함수를 갖는 계층의 구조를 보여주고 있다.

Fig. 12는 인공신경망 AnnSampleSkin 모델의 빌드업 및 훈련과정에서 분석된 성능 결과를 보여주고 있다.

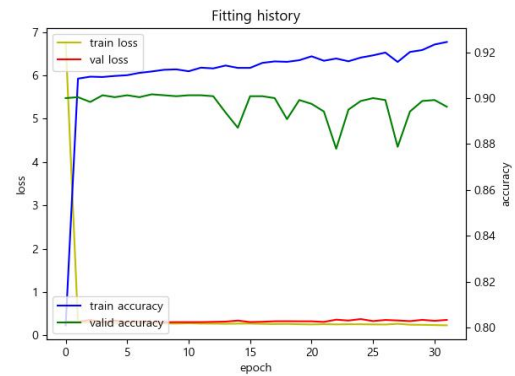


Fig. 12. Performance results during build-up and training of the AnnSampleSkin model

훈련샘플데이터에 대한 훈련 정확도(train accuracy)는 5 에포크에서 30 에포크(epoch)까지 정확도가 점진적으로 상승 되어 약 92% 이상의 정확도를 나타내고 있고, 훈련 손실 값(train loss)은 0에 가까운 낮은 값을 나타내고 있다. 검증샘플데이터에 대한 검증 정확도(valid accuracy)는 15에포크에서 30에포크까지 불안정한 정확도를 나타내고 있으나 평균적으로 약 90%의 정확성을 나타내고 있고, 검증 손실 값(valid loss)은 0에 가까운 낮은 값을 나타내어 인공신경망 검증 완료가 가능한 성능 분석 결과를 나타내고 있다.

4.4 Results of skin image identification

피부이미지식별단계에서는 소셜미디어에서 수집한 이미지에 대해 피부이미지분석과 피부샘플을 추출하고 훈련된 인공신경망 AnnSampleSkin의 이미지 유형 예측 결과를 사용하여 지능적으로 피부 이미지를 식별하였다. 인공신경망 AnnSampleSkin의 성능평가는 훈련샘플데이터에 포

합되지 않은 테스트샘플데이터를 사용함으로써 평가 분석 결과에 대한 객관성을 높일 수 있도록 하였다. 훈련된 인공신경망 AnnSampleSkin의 테스트샘플 데이터의 예측 성능에 대한 평가분석 결과는 Fig. 13과 같다.

```

Evaluate Results
테스트 샘플 X_test 데이터 크기: (5999, 64, 64, 3)
테스트 샘플 y_test 레이블 크기: 5999

batch_size = 200
score = [0.3519286513328552, 0.9101517200469971]
test loss = 0.3519286513328552
test accuracy = 0.9101517200469971
    
```

Fig. 13. Evaluate results of AnnSampleSkin

테스트 샘플 이미지 X_test와 라벨 데이터 y_test(5,999개)에 대해 이미지 유형 예측 값의 정확성은 약 91%의 성능을 보였고, 손실 값(test loss)은 약 0.35로 0에 가까운 낮은 오차를 나타내고 있다.

	source Image	skin Image	AnnSampleSkin Result	Traditional Result
(a)			Yes skin	Yes skin
(b)			Yes skin	Yes skin
(c)			Yes skin	No skin
			No skin	Yes skin
			No skin	Yes skin

Fig. 14. The final skin image identification result of the skin image identification stage

Fig. 14는 피부이미지식별단계의 최종적인 피부이미지 식별 결과를 나타내고 있다. 소스이미지(sourceImage)는 소셜미디어에서 수집된 원본 이미지를 나타내고 있으며 다양한 형태와 내용을 담고 있는 것으로 피부 상태 진단 및 관리에 바로 사용하기 어려운 상태임을 볼 수 있다. skinImage는 얼굴탐지, 피부영역 마스크 및 피부샘플 추출의 피부이미지분석과정에 의해 추출된 결과를 나타내고 있다. AnnSampleSkinResult는 인공신경망 AnnSampleSkin

의 이미지유형 예측 결과를 종합하여 결정된 피부이미지 식별 결과의 이미지 유형을 나타내고 있다. TraditionalResult는 피부이미지분석단계의 기존 이미지처리 기술을 사용하여 결정된 이미지 유형을 나타내고 있다. Fig. 14의 (c)에 보인 것과 같이 피부 색상을 중심으로 식별하는 전통적인 이미지 처리기술의 피부 이미지를 식별한 결과에서는 피부 이미지를 분별하지 못하는 오인식된 결과가 포함되어 있음을 알 수 있다. 그러나 본 논문에서 제안된 피부이미지식별 결과에서는 피부 색상과 유사한 이미지에 대해서도 피부 샘플에 대해 다양한 특징으로 훈련된 인공신경망의 예측 값을 종합하여 판단함으로써 기존의 피부 색상을 중심으로한 피부 이미지 식별 방법 보다 정확하게 식별하는 결과를 얻을 수 있었다.

V. Conclusions

일반적으로 화장품 및 뷰티산업 종사자들이 피부 미용을 목적으로 피부 상태 진단을 하는 경우에 사용하는 방법은 대부분 직접 대면을 통하여 피부 상태의 육안 관찰과 고객의 의견을 파악하는 것이다. 그러나 정보통신기술의 발전과 고객의 요구 수준이 높아짐에 따라 더욱더 효율적이고 편리한 온라인 비대면 환경으로 변화를 요구받고 있다. 따라서 피부상태 진단 및 관리에 있어서 소셜미디어 커뮤니티와 같은 다양한 온라인 환경으로부터 고객의 요구 사항을 파악하고 진단하는 보다 긴밀한 상호작용을 할 필요가 있다. 소셜미디어에서 다루어지는 이미지는 매우 비정형적이며 다양한 내용을 포함하고 있으므로 피부 상태 진단을 위한 피부 이미지 사용에 앞서 사용 목적에 적합하게 피부 이미지를 식별하고 정형화된 피부 이미지를 추출하는 빅데이터 전처리 기술이 필요하다. 그러나 기존의 피부 색상 정보만을 사용하는 피부 이미지 식별 방법에서는 피부색과 유사한 색을 지닌 다른 이미지 영상과 구별하기 어려운 문제가 있다.

본 논문에서는 소셜미디어 인스타그램에서 화장품 및 피부미용을 위한 피부 이미지를 식별하여 수집하고, 정형화된 피부 샘플을 추출하는 인공신경망 기반의 지능화된 피부 이미지 식별 및 피부 샘플 추출 시스템을 개발하였다. 본 논문에서 제안된 시스템은 소셜미디어로부터 체계적으로 충분한 분량의 피부 이미지를 식별하여 수집하고, 정형화된 피부 샘플 이미지를 제공함으로써 피부 유형 진단 등 다양한 향후 연구에 필요한 기초 데이터를 제공할 수 있을 것이다. 기존의 기계학습 기술을 활용한 피부 상태 진단 관

런 연구 결과를 실용화 하는데 있어서 선결되어야 할 과제는 정형화된 충분한 규모의 피부 이미지와 라벨링된 훈련 데이터가 제공될 수 있어야 하는 것이다. 그리고 화장품 및 피부 미용에 대한 고객의 요구 사항과 환경은 생활환경 및 수준 등에 따라 계속적으로 변화하므로 소셜미디어의 빅데이터와 같은 현재 시점의 고객 요구를 반영하는 피부 이미지를 수집하고 분석하여 제공할 필요가 있다.

본 논문에서 제안한 시스템은 빅데이터수집분석단계, 피부이미지분석단계, 훈련데이터준비단계, 인공지능망훈련단계, 피부이미지식별단계로 구성된다. 빅데이터수집분석단계에서는 소셜미디어 커뮤니티로부터 피부상태 관리에 대한 빅데이터를 수집하고 텍스트분석을 진행하였다. 피부이미지분석단계에서는 텍스트분석결과로부터 피부상태 관리와 관련된 이미지에 대하여 전통적인 이미지 처리 기법을 사용하여 피부 이미지를 탐지하고 분석 결과를 획득하였다. 훈련데이터준비단계와 인공지능망훈련단계에서는 피부이미지분석단계의 피부 이미지 분석 결과로부터 피부 샘플 데이터를 추출하여 훈련데이터를 준비하고 이를 사용하여 인공지능망을 빌드업하고 훈련함으로써 피부 샘플에 대하여 지능적으로 피부 이미지 유형을 예측할 수 있도록 하였다. 피부이미지식별단계에서는 소셜미디어로부터 수집된 이미지에 대해 피부 샘플을 랜덤하게 추출하고 훈련된 인공지능망의 피부샘플의 이미지 유형 예측 결과들을 통합하여 최종적인 피부 이미지 유형을 지능적으로 식별하였다.

피부샘플세트이미지는 이미지로부터 임의의 위치에서 랜덤하게 충분한 개수만큼 샘플링함으로써 이를 통하여 이미지 유형을 식별하기에 적합한 정형화된 피부샘플이미지로 평가된다. 뿐만 아니라 추출된 피부샘플세트이미지는 이후 피부 상태를 진단하는데 매우 효율적이고 유용한 피부 이미지 데이터로 사용될 것으로 기대된다. 피부이미지를 식별하기 위하여 빌드업되고 훈련된 인공지능망 AnnSampleSkin은 훈련샘플데이터에 대한 훈련 정확도는 약 92% 이상의 높은 정확도를 나타내어 피부 이미지를 지능적이고 효율적으로 식별할 수 있었다.

본 논문에서 제안한 시스템은 소셜미디어로부터 피부 이미지를 수집하고 정규화된 피부 샘플이미지세트를 생성하여 제공함으로써 소셜미디어 기반의 피부 상태 진단 및 관리에 반드시 필요한 피부이미지 빅데이터와 데이터 공급 체계를 마련하였다고 평가할 수 있다. 그리고 화장품 및 뷰티산업 분야에서 다양한 고객 니즈를 반영한 기능성 화장품과 맞춤형 서비스를 제공하는데 활용될 수 있을 것이다.

향후에는 추출된 피부 샘플 이미지를 사용하여 피부 상태 진단 및 관리 등 화장품 및 뷰티산업에서 요구되는 기

술 수요를 제공하기 위한 피부 상태 진단 및 관리를 위한 추가적인 피부샘플이미지처리 연구가 필요하다. 또한 향후 연구에서는 소셜미디어의 피부 이미지 빅데이터와 특수장비를 사용하여 취득한 피부 이미지를 병행하여 활용은 보다 개선된 피부 상태 진단 및 관리 기술을 개발하여 실험하고 분석할 필요가 있다. 향후 소셜미디어의 피부관리에 대한 빅데이터 정보를 활용하는 시스템은 다양한 방식으로 고객의 니즈를 파악하고 고객 맞춤형 화장품 및 뷰티산업을 발전시켜 가는데 있어서 매우 미래 지향적이며 중요한 기초 기술로 활용 가능할 것으로 기대된다.

REFERENCES

- [1] Eui-Hyang Kim, Hyun-joo Kim, "A Study on the Analysis Method of Skin Condition through Visual Confirmation of Skin Surface", *Journal of Convergence for Information Technology*, Vol. 11. No. 5, pp. 267-275, May. 2021 DOI : 10.22156/CS4SMB.2021.11.05.267
- [2] J. S. Koh, "Observation of Sebum, Moisture content and Microtopography of Different aged Skin for classification of facial Skin Types", *Journal of the Society of Cosmetic Scientists of Korea*, Vol. 15. No. 1, pp. 63-70, 1989
- [3] Kim Ki Young, "A Study on Model of Skin Type Judgment Tool Using Machine Learning Technique", *The Treatise on The Plastic Media*, Vol. 4. No. 21, pp. 115-121, Nov. 2018
- [4] P. Viola and M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, 2001
- [5] Park, Gyeong-Mi and Kim, Young-Bong, "Integrated 3D Skin Color Model for Robust Skin Color Detection of Various Races," *The Journal of the Korea Contents Association*, vol. 9, no. 5, pp. 1-12, May 2009. DOI : 10.5392/JKCA.2009.9.5.001
- [6] Dae-Chul Kim, Cheol_hee Lee, Myong-Hui Choi, Yeong-Ho Ha, "Skin Detection Method using Color Space based Methods and Focus Region", *Journal of Korean Society for Imaging Science & Technology*, Vol. 18. No. 4, pp. 16-22, Dec. 2012
- [7] Hyung-Hoon Kim, Jeong-Ran Cho, "Development of big data based Skin Care Information System SCIS for skin condition diagnosis and management", *Journal of The Korea Society of Computer and Information*, Vol. 27. No. 3, pp. 137-147, Mar. 2022 DOI : <https://doi.org/10.9708/jksoci.2022.27.03.137>
- [8] Salton G. and McGill, M. J., "Introduction to modern information retrieval", McGraw-Hill, ISBN 0-07-054484-0., 1983
- [9] Zheng, Tong, Zha. Dongmiao, Wang, Xingtian, Gao, Weijun, Miao, Sheng, "Architecture Descriptions Analysis Based on Text Mining

and Crawling Technology”, 2021 IEEE 3rd International Conference on Frontiers Technology of Information and Computer , pp. 138-143, Nov. 2021

- [10] Jae-Been Ha, Do-Eun Lee, “A Trend Analysis and Policy proposal for the Work Permit System through Text Mining: Focusing on Text Mining and Social Network analysis”, Journal of Convergence for Information Technology, Vol. 11. No. 9, pp. 17-27, 2021 DOI : 10.22156/CS4SMB.2021.11.09.017
- [11] Yun-Mi Jun, Ki-Jung Kang, “Communitycare Key words and Social Network Analysis based on Big Data”, Korean Journal of Family Welfare, Vol. 24. No. 2, pp. 251-269, June. 2019 DOI : 10.13049/kfwa.2019.24.2.7
- [12] Shineware, KOMORAN, <https://komorandocs.readthedocs.io/>
- [13] Kim, Sun-ji, Chon, HaeJung, “Analysis of the Relationship between Knowledge and Praticce of Skin Moisture and Facial Parts WaterContent ”, Journal of the Korean society for Wellness, Vol. 13. No. 4, pp. 81-89, Dec. 2018 DOI : 10.21097/ksw.2018.11.13.4.81
- [14] Chatzimparmpas, A, Martins, R.M, Kerren, A, “t-viSNE: Interactive Assessment and Interpretation of t-SNE Projections”, IEEE Transactions on Visualization and Computer Graphics, Vol. 26. No. 8, pp. 2696-2714, Aug. 2020 DOI : 10.1109/tvcg.2020.2986996

Authors



Hyung-Hoon Kim received the B.S. degrees in Computer Science from Chonnam National University, Korea, in 1986. He received the M.S. degrees in Computer Science from KAIST, Korea, in 1988.

He received the Ph.D. degrees in Computer Science from Hanyang University, Korea, in 2007. Dr. Kim joined the faculty of Kwangju Womens University, Gwangju, Korea, in 1994. He is currently a Professor in the Department of Cosmetic Science at Kwangju Womens University. He is interested in web programming, medical information system, information security, and AI.



Jeong-Ran Cho received the B.S., M.S. and Ph.D. degrees in Computer Science from Chonnam National University, Korea, in 1987, 1989 and 1999, respectively. Dr. Cho joined the faculty of Kwangju Womens

University, Gwangju, Korea, in 1994. She is currently a Professor in the Department of Health Administration at Kwangju Womens University. She is interested in database, parallel computing, internet and mobile computing, and multimedia contents service.