Expert System for the Pre-Diagnosis of Skin Diseases

Malla Zhunio Katherine Patricia, Vicente Zapata Edison Santiago, Montaluisa Yugla Franklin Javier

Abstract— **Skin diseases are a common health problem worldwide; this article proposes a method based on deep learning techniques combined with computer vision to detect various types of dermatological diseases. The system relies on m-health, a fundamental component of e-health, which involves the use of mobile devices for diagnosis, thus making it completely non-invasive for the patient and therefore accessible in rural areas where access to dermatologists is limited. Image processing algorithms have been used in the system for the extraction of characteristics of the sample provided by the patient, which serves to feed the convolutional neural network, this network allows to classify images by subdividing them into layers, making it easier to extract patterns through the application of different filters. This expert system works in two phases: the first: analysis and processing of the color image to extract the characteristics and patterns to obtain classified models and then make the prediction or identification of the disease. The second phase of retraining consists of a feedback to the training data of the network, which allows automatic learning of the algorithm. The system successfully detects three types of dermatological diseases: Dermatitis, Pityriasis or Tinea versicolor and Melasma, diseases with the highest incidence in Ecuador, with an average accuracy rate of 90%.**

Index Terms— **skin diseases, deep learning, computer vision, m-health, convolutional neural network, image processing, automatic learning.**

I. INTRODUCTION

The process of diagnosing skin diseases is mainly carried out in two stages: anamnesis, which is a set of data collected from a patient, and physical examination; sometimes also complementary explorations. [1]. This whole process takes too long to obtain a definitive diagnosis, which causes, in several cases, the patient to abandon his treatment. [2]. The process of anamnesis and exploration, in many cases, are carried out with the help of technology, which has evolved over the years.

Medical systems comprise three aspects: management, acquisition and maintenance of software, in order to make preventive diagnoses and maximize the impact on medical decisions, these systems incorporate autonomy and intelligence generating data securely and efficiently. [3].

As a result, a new terminology called *e-Health* emerges, which is defined as "the use of electronic information and ICTs in support of clinical care, health education and public health at a distance". [4].

In this area, mobile devices have become a tool used in conjunction with AI to revolutionize the tracking of conditions in simpler and more reliable ways.

I. SUMMARY DESCRIPTION OF THE PROJECT

The development of a medical expert system capable of pre-diagnosing common diseases in the skin that is carried out following the guidelines of the medical informatics standard for coding diseases, ICD-10; this expert system is made up of two modules: *i)* The research and analysis module, which will diagnose the type of disease presented by the patient and quantify the level of certainty of the diagnosed disease. For this purpose, the semiology of the images of each of the diseases to be treated will be reviewed and samples will be taken, which will be validated by the expert. This will apply Deep Learning algorithms to generate three types of classes that correspond to the diseases treated and a specific class that excludes the aforementioned. In such a way that a training data model is generated; to carry out the pre-diagnosis. The subject of study must provide a photo of the affected area which will be obtained and analyzed by artificial vision so that the existing model can use it to predict what disease resembles and the percentage of certainty. *ii)* The Re-training module consists of providing feedback to the training data, i.e. those corresponding to melasma, dermatitis, pityriasis and the so-called "others"; depending on whether the pre-diagnosis is correct and validated by the expert, the image will be stored in the training data corresponding to the prediction, otherwise it will be stored in "others".

The previously detailed expert system was validated by a specialist in the area of dermatology, who provided all the necessary parameters for the development of the system in order to obtain a reliable functionality at the time of prediction, as well as a group of patients who participated in testing the expert system.

II. BACKGROUND AND RELATED WORKS

A. *Artificial Intelligence and the m-health*

Currently, the impact of Information and Communication Technologies (ICT's) in medicine is decisive and key in the diagnosis of diseases. [4]. Artificial Intelligence is little used

F. A. Katherine P. Malla Zhunio is with the Universidad de las Fuerzas Armadas ESPE, Student, ECUADOR (e-mail: kpmalla@espe.edu.ec).

S. B. Edison S. Vicente Zapata is with the Universidad de las Fuerzas Armadas ESPE, Student, ECUADOR (e-mail: evicente@espe.edu.ec).

T. C. Franklin J. Montaluisa Yugla is with the Universidad de las Fuerzas Armadas ESPE, Professor, ECUADOR (e-mail: fjmontaluisa@espe.edu.ec).

in the field of medicine, specifically in the diagnosis of diseases whose identification could be detected with a simple visual examination.

In Ecuador, the process of incorporating technology and information resources into the health area is in full swing. This initiative has implemented intelligent medical systems projects related to prevention, disease monitoring and communication between patient and home health, but not mobile devices are used to diagnose diseases. [5].

The m-health is "a variant of telemedicine performed with the support of mobile devices such as (smartphones, tablets, PDAs). All of this has contributed to improving accessibility, streamlining diagnostic processes and transcending geographical, economic and political barriers. [4].

B. Automatic Learning Algorithms

There are several types of automatic learning algorithms, supervised, unsupervised and by reinforcements [6]. In automatic learning, a computer first learns to perform a task using a training dataset. The computer then performs the same task with the test set [6]. Supervised learning is about having a set of data tagged with the correct answer, i.e. both input and output data are sent and the result is already known.

Supervised learning is of two types: classification-based and regression-based. In this work, regression-based supervised learning is used. [7]. The Neuronal Convolutional Networks (CNN) is an algorithm that explicitly receives images as input allowing to gain efficiency and reduce the amount of parameters in the network. [8].

C. Artificial vision in the diagnosis of diseases

Although for computer vision it is still difficult to process information like the human brain, which does so in a semantic way, that is, by extracting significant characteristics. [9]. Over the years, computer vision combined with automatic learning algorithms can get more features by applying filters or kernels and deepening the division into layers, making predictions more accurate. [9] [10].

D. Convolutional Neural Networks to build models.

CNNs are similar to neural networks. What differentiates convolutional neural networks is that they explicitly assume that inputs are images; allowing efficiency gains and reducing the number of parameters in the network. Convolutional neural networks solve the problem that ordinary neural networks do not scale well for high definition images.

The CNNs work by modeling consecutively with small matrices of information by applying what is called convolutions, and then combining this information in the deeper layers of the network. One way to understand them is that the first layer will try to detect the edges and set edge detection patterns. Then, the subsequent layers will try to combine them in simpler ways and, finally, in patterns of the different positions of objects, lighting, colors, textures, among others. The final layers will try to match an input image to all patterns and arrive at a final prediction. In this way, convolutional neural networks are able to model complex variations and behaviours giving quite precise predictions.

E. Skin diseases in Ecuador

In the present investigation the three diseases with greater incidence in Ecuador were studied: melasma, pityriasis or tinea versicolor and dermatitis, in addition to carrying out an analysis of the semiology of the images of these diseases.

The word Melasma means *a black spot.* Also known as chloasma or mask of pregnancy. [11]. It is an acquired hyperpigmentation of light or dark brown color that occurs in areas exposed to light, with notable predominance in the female sex. [12]. African Americans, Asians, and Hispanics are the most susceptible populations [11]. It also occasionally affects men. It appears more frequently in the third and fourth decades of life, but sometimes earlier. [13]. Lesions consist of variable sized, asymptomatic, light brown or dark brown macules, with varying degrees of pigmentation, irregular and sometimes well-defined borders. [13].

Tinea versicolor is also called "pityriasis versicolor," a common fungal infection of the skin. It is a mycosis produced by a common saprophytic fungus on the skin, Malassezia furfur. [1]. This fungus predominates in seborrheic areas which makes it exceptional before puberty, since at this stage there is no sebaceous secretion. The fungus affects the normal pigmentation in the skin, which leads to the formation of small discolored spots. These spots may be lighter or darker in color than the surrounding skin, and affect, more often in the trunk and shoulders. [11]. It is more frequent in the juvenile and adult stage and especially in men, it is neither painful nor contagious. However, it can cause emotional distress or insecurity. [14].

The terms eczema and dermatitis are considered synonyms and describe a pattern of inflammatory response of the skin characterized by itching and polymorphic lesions: erythema, edema, blistering lesions, scaling and lichenification. [12]. These characteristics are common to all eczemas, differentiated by aetiology. Eczema may be acute, subacute, or chronic [1]. There are several causes and a wide variety of clinical manifestations: Individuals engaged in recreational or work activities are at risk for dermatitis: domestic workers, hairdressers, medical, dental and veterinary personnel, cleaners, flower shop, agriculture, horticulture, forestry, food preparation and service, printing, painting, metalwork, and infants 1-5 years of age and young women. [1] [11]

F. Related Works

Currently there are several works related to the use of technology in the field of medicine. According to the World Health Organization (WHO), this concept is called cyber health (also known as eHealth or e-health). These days to have systems based on artificial intelligence that can facilitate access to public health by crossing geographical and economic barriers are indispensable in rural areas and where access to a dermatologist is limited. It is for this reason that several resources are allocated for the development of this type of systems around the world. However, these systems also present limitations as they are aware that not everyone has access to technology, which is why alternatives must be sought with the implementation of cloud servers, as proposed in this paper.
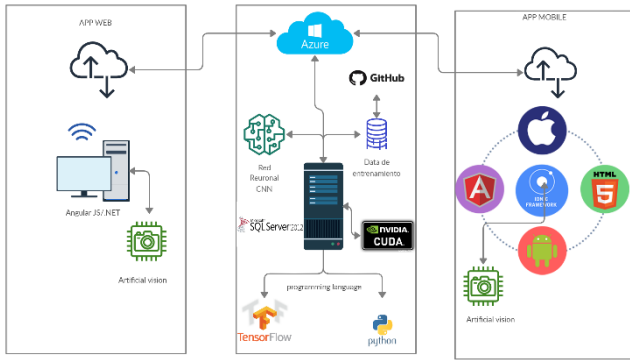
III. ARCHITECTURE AND METHODOLOGY

Fig. 1. Architecture.

The architecture proposed in this research aims to focus the development environment to create and train a convolutional neural network, generating reliable and safe algorithms, and thus predict the three types of diseases raised, as shown in Fig. 1.

A. Training Environment (GPU Tensor Flow, Cuda.).

GPU (Graphics Processing Unit). - For the training of models of a neural network it is fundamental the use of GPU, Basically, to the being another added processor, its function is that of freeing of load to the CPU, increasing the performance of the server that is used to increase the performance and to reduce the load.

TensorFlow. - Much of the algorithm proposed for this research is based on the TensorFlow library that allows building and training neural networks, therefore, facilitates detecting and deciphering patterns and correlations, analogous to learning and reasoning used by humans.

Cuda. - is a computing and programming model invented by Nvidia that allows us to use the computational power of the GPU for any task you want in the proposed model of artificial intelligence development.

B. Applications (web - mobile)

Web. - The purpose of the web administrator is to manage users who suffer from these three types of diseases, is focused on health professionals in dermatology, which is responsible for validating the result of the neural network.

Mobile. - The objective of the mobile application is that a user, patient, provides a photographic sample of the affected area that is redirected to the neuronal network, which returns the name of the disease to which it resembles and the percentage of success it has; this process is validated by the dermatological specialist, who will approve or reject the result of the sample.

C. Development Methodology

For the adequate development of the expert system, the software development methodology "BUCHANAN" was taken into account, which allows for a constant relationship between the knowledge engineer and the area expert. It has six fundamental stages: *i) Identification*, where participants were defined who will play the role of patients, resources and sources of knowledge, in this case, the dermatology specialist, in addition to bibliographic and graphic sources. In addition, in this phase the computational facilities, budget and objectives and goals were identified. *ii) Conceptualization*, together with the expert an investigation was carried out on the main foundations of the diseases that were studied. *iii) Formalization*, the prototype of the knowledge base is designed and built with images that will be previously classified and validated by the expert. *iv) Implementation,* the designed prototype was implemented in Python programming language, where a model was generated after applying a multilayer neuronal network for image processing through TensorFlow-GPU libraries. *v) Testing*, tests were performed to observe the behavior of the prototype, the functioning of the knowledge base and where the level of certainty or performance of the system was verified. *vi) Prototype revision,* at this stage the prototype was validated by the expert. [15].

D. Medical Standard

Medical standards are of great help for Medical Informatics (IM) because they allow to harmonize the methods of information management and analysis. These methods are based on the use of a common language and the use of specific biomedical terminology. [16].

The ICD is the acronym for the International Classification of Diseases and Health-Related Problems. The ICD 10, represents the tenth version and determines the classification and coding of diseases and a wide variety of signs, symptoms, abnormal findings, complaints, social circumstances and external causes of damage and/or disease. [17].

Each affection can be assigned to a category and receive a code up to six characters long (in X00.00 format). For the present research, Chapter XII is of interest, which deals with diseases of the skin and subcutaneous tissue. [18].

IV. DATA SET

In this investigation, the first thing was to obtain statistical data on the index of most common skin affections in Ecuador.

According to the National Institute of Statistics and Censuses of Ecuador, until 2017, about 0.11% of the population suffered from some type of disease or infection in the skin and subcutaneous tissue. This rate has been increasing approximately 0.01% each year; due to various climatic, hereditary, hormonal, or contact conditions; presenting with a greater incidence in men. [19]. The most common skin diseases recorded are: melasma occurring in 0.17% of cases, mainly in pregnant women. Dermatitis in 80.65% where babies and children are most affected, usually between 3 months and 5 years of age, with a maximum of 80% of cases occurring before the first year of life. And pityriasis occurring at 7.02% usually in hot weather in 15-40 year olds. [20].

The following table details the number of samples of the three selected disease types.

TABLE I. Data Collection

Skin disease	Number of Samples
Melasma	387
Pityriasis or Tinea Versicolor	250
Dermatitis	472
Other	1257

In this research 2270 images of skin diseases were taken from repositories and dermatology books and 96 patients suffering from any of the diseases to be treated.

A specialist in the area of dermatology validated the set of training data used in this research and later also corroborated the results of the expert system.

V. IMPLEMENTATION

A. Algorithm based on Supervised Learning

The research aims to develop and implement an expert system that allows an electronic device to learn for itself and make decisions according to a knowledge base, which is used to obtain a training data model as a result of the application of the algorithm based on supervised learning. In this way, the system will obtain a pre-diagnosis with an acceptable and efficient level of certainty that was later validated by the expert in the area of dermatology.

In this research, convolutional neural networks were used as a supervised learning training algorithm for the classification of the three types of diseases to be treated.

In TABLE I, there is a fourth classification called "Others", this set of data serves to exclude other types of skin diseases that do not correspond to Melasma, Ringworm or Pityriasis and Dermatitis, or simply no type of disease.

When working with learning algorithms, it is necessary to validate the models with which it is going to be predicted, for this it is recommended to divide the dataset in a set of training, validation and tests as can be seen in Fig. 2.

Fig. 2. Machine Learning Modeling Methodology

The training set is the data with which the model is constructed, the validation data is a portion that is used to validate the model, the test set is data that is used to evaluate the model.

B. Image Processing with Convolutional Neural Networks

For the research, a convolutional neural network was constructed, which allows us to obtain patterns and characteristics of the input image in different layers after each convolution and which finally allows us to classify them.
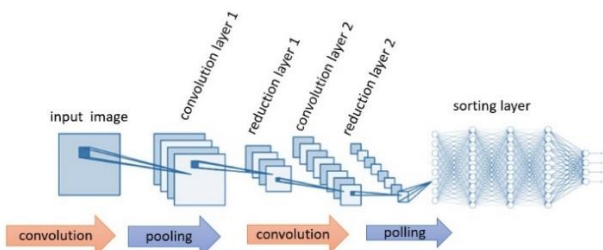

Fig. 2. Functioning of a CNN

A convolutional neural network has two fundamental operations, convolution and reduction. The convolution operation receives the image as input and then applies on it a filter or kernel that returns a map of characteristics of the original image, reducing the size of the parameters forming the convolutional layer, as can be seen in Fig. 3.
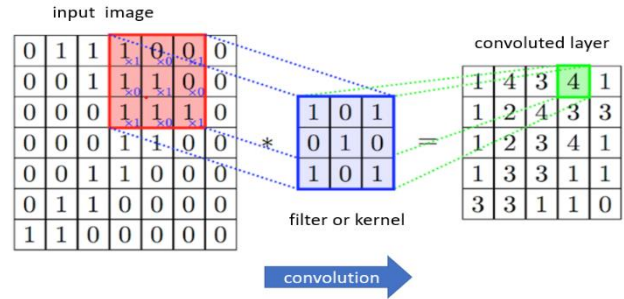

Fig. 3. Convolution

The reduction operation is performed after the convolutional layer, its main utility is to reduce the spatial dimensions of the image (width and height) for the next convolutional layer without affecting the depth of the volume, as shown in Fig. 4.
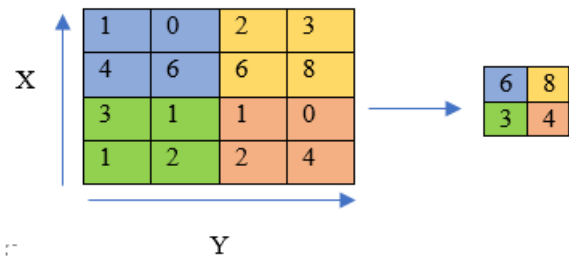

Fig. 4. Reduction.

C. Prediction

For this investigation, in the prediction stage computer vision was used, which allows to extract characteristics of one or several images such as color code, average infected area, border of the infected area.

The diagram of the image processing system is shown in Fig. 5. below.
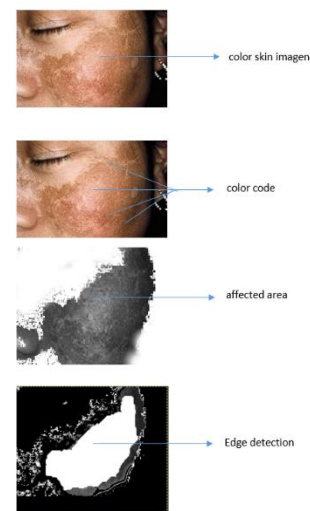

Fig. 5. Characteristics obtained with computer vision.

Then, the input images are processed in the convolutional neural network, which applies several filters or kernels to each of these, in this process patterns or characteristics are obtained to classify the input images. The filters shown in Fig. 6, are some examples, in this case the neural network applies 20 different filters.
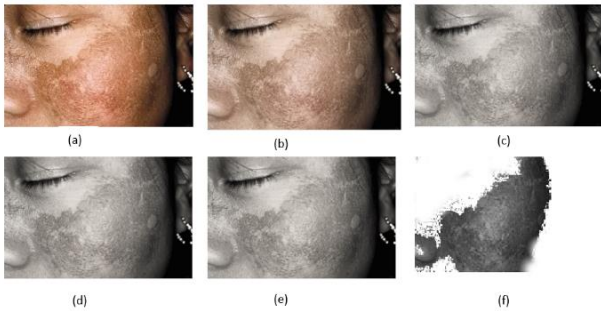

Fig. 6. Result of the algorithm. (a) Original color image, (b) Transforms the image into grayscale, (c) Image sharpness increase, (d) Medium filter, (e) Smooth filter, (f) Binary photography.

The input image is transformed into grayscale, this filter is used so that the neural network can identify the affected area in the skin, therefore, applies a filter that increases the sharpness and removes noise from each image, this helps the neural network has a better performance at the time of training.

D. Re-training

The Re-training module will be treated by the dermatology specialist from the web administrator, consists of performing a retro feeding to each of the classes created, ie those corresponding to melasma, dermatitis, pitiriasis and the so-called "others", depending on whether the pre-diagnosis is correct and validated by the expert, the image will be stored in the class corresponding to the prediction, otherwise it will be stored in the class "others" and recreate the models with new images.

VI. RESULTS

In TABLE II, the results are expressed, with an example of a color image of the three skin diseases, the total number of images used for the tests, the number of images whose diagnosis was successful, and the percentage of certainty of each.

TABLE II. Disease Detection Rate

Disease Name	Sample Image	Total Image	Disease Detected	Detection Rate %.
Eczema/ Dermatitis		71	64	90,14%
Melasma/ Cloasma		58	54	94,82%
Pityriasis/ Tinea versicolor		38	33	86,84%

The proposed system can successfully detect three different dermatological diseases with an accuracy of 90.6%, as shown in Fig. 7. 15% of the entire dataset has been used, as indicated for testing, 15% of images for validation and 70% of images for training purposes.
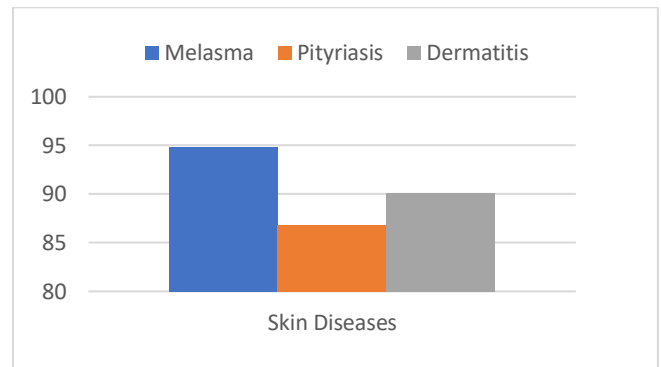

Fig. 7. Detection Rate

VII. CONCLUSIONS AND FUTURE WORK

In this research we present a system for detecting skin diseases, the expert system designed worked successfully for all three types of diseases. For this type of images, the system can generate binary layers with high precision and can detect infected areas with better results.

The continuation of this research work in the future is the development of a version that allows to add more types of diseases in the training data of the neuronal network, as well as a medical history of the users, which would be important for the prediction of other types of diseases, being able, the system, to increase its level of precision at the time of making a diagnosis.

CONFLICT OF INTEREST

"The authors declare no conflict of interest".

AUTHOR'S CONTRIBUTIONS

A. in charge of bibliographic research, B. developed the architecture of the expert system; C. oversaw all research and implementation, also directed system testing, A, B analyzed the data, implemented the system, and wrote the paper, C analyzed the data; all authors approved the final version.

ACKNOWLEDGMENT

REFERENCES

[1] J. Serna, M. Vitales, M. López and A. Molina, "Dermatología," de *Farmacia Hospitalaria - Tomo II*, Madrid, SEFH, 2002, pp. 842-856.

[2] M. Ribera, "Semiología Dermatológica," from *Dermatología Clínica 3era Edición*, Barcelona, Elsevier, 2009, pp. 11-20.

[3] M. Silano, "La Salud 2.0 y la atención de la salud en la era digital," *Rev. Médica Risaralda,* vol. 19, no. 3, pp. 1-14, 2013.

[4] J. Salvador, "Sanidad del Futuro (e-Health, m-Health, e Inteligencia Ambiental)," de *Gestión sanitaria integral: pública y privada*, Madrid, Centro de Estudios Financieros, 2010, pp. 561-575.

[5] R. López, M. Chiriboga y A. Carrera, «The present situation of e-health and mHealth in Ecuador,» *Latin American Journal of Telehealth,* vol. 4, nº 3, pp. 261-267, 2017.

[6] P. Louridas and C. Ebert, "Machine Learning," *IEEE Software,* pp. 110-115, 2016.

[7] K. Manoj y S. Shylaja, «Scene Classification with Deep Neural Nets Using Background Suppression,» de *2018 International Conference on Information Technology (ICIT)*, Bhubaneswar, India, 2018.

[8] T. Sharma, V. Singh, S. Sudhakaran y N. K. Verma, «Fuzzy based Pooling in Convolutional Neural Network for Image Classification,» de *2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, New Orleans, LA, 2019.

[9] B. Zhang, «Computer vision vs. human vision,» de *9th IEEE International Conference on Cognitive Informatics (ICCI'10)*, Beijing, China, 2010.

[10] K. Azhar, F. Murtaza, M. H. Yousaf y H. A. Habib, «Computer vision based detection and localization of potholes in asphalt pavement images,» de *2016 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, Vancouver, BC, Canada, 2016.

[11] H. Thomas, Skin Diseases. Fifth Edition, Madrid: Elsevier, 2009.

[12] K. Wolff, R. Johnson and A. Saveera, Fitzpatrick. Atlas de Dermatología Clínica, Mexico City: McGRAW-HILL, 2014, pp. 18-48.

[13] K. Porat, "Etiología Clínica y Manejo del Melasma," *Revista Médica de Costa Rica y Centro América,* vol. 68, pp. 57-60, 2011.

[14] R. Falabella, J. Victoria, M. Barona and L. Domínguez, Fundamentos de Medicina - Dermatología Sexta Edición, Medellín: Corporación para investigaciones biológicas, 2002.

[15] L. Montenegro, A. Vásquez and E. Prentice, "Aplicación de metodología Buchanan para la construcción de un sistema experto con redes bayesianas para apoyo al diagnóstico de la Tetralogía de Fallot en el Perú," *Revista de Investigación UNMSM,* vol. 18, nº 1, pp. 135-147, 2015.

[16] F. Suárez-Obando and J. C. Sánchez, "Estándares en informática médica: generalidades y aplicaciones," *Revista Colombiana de Psiquiatría,* vol. 42, nº 3, pp. 295-302, 2013.

[17] K. Gomez, "International Health Standards for Medical Information Management," *MegaPractical,* 18 May 2016.

[18] "ICD-10," *International Statistical Classification of Diseases and Related Health Problems,* vol. 3, no. 10, 2008.

[19] Instituto Nacional de Estadísticas y Censos, "Camas y Egresos Hospitalarios," 2017. [Online]. Available: http://www.ecuadorencifras.gob.ec/camas-y-egresos-hospitalarios/. Last login: 20 February 2019.

[20] Instituto Nacional de Estadísticas y Censos, "Camas y Egresos Hospitalarios por Tiña (Pitiriasis), Dermatitis y Melasma," 2017. [Online]. Available: _COPY18@outlook.com=48349. Last login: 21 February 2019.

[21] F. Sacoto, "El país requiere una estrategia de e-Health articulado al sistema nacional," Sanitaria 2000, Quito, 2017.

[22] K. Wolff, R. Johnson, and A. Saveera, "Eczema/Dermatitis," by *Fitzpatrick. Atlas of Clinical Dermatology*, Mexico City, McGRAW-HILL, 2014, pp. 18-48.

[23] F. M. Deaborn, Skin Diseases, Delhi: B. Jain Publishers, 2005.

First A. Katherine Patricia Malla Zhunio, was born in Quito, Ecuador on October 21^{st,} 1995. She is currently studying her semester degree in Software Engineering at the Universidad de las Fuerzas Armadas ESPE and the last semester of the Technology degree in Quality Systems Management at the Edwards Deming Corporate Technological Institute (Instituto Tecnológico Corporativo Edwards Deming). She has attended several congresses on Science, Innovation and Technology. She is fluent in several programming languages and has experience as a scrum project manager.

Second B. Edison Santiago Vicente Zapata, was born on October 14, 1993 in the city of Latacunga, Ecuador. on October 14^{th,} 1993. He is currently studying the semester of the Software Engineering degree at the Universidad de las Fuerzas Armadas ESPE. He has high knowledge in several programming languages, artificial intelligence, Iot, software infrastructure, as well as experience in several development environments and methodologies.

Third C. Franklin Javier Montaluisa Yugla, professor at the Universidad de las Fuerzas Armadas ESPE.
Engineer in Systems and Informatics; he also has a master's degree in Software Engineering. He has knowledge of Java, C, C#, Kotlin, Dart, knowledge in artificial intelligence.
Expert in mobile application development.