

SISTEMA PARA EL DIAGNÓSTICO DE HERIDAS SUPERFICIALES EN LA PIEL MEDIANTE PROCESAMIENTO DE IMÁGENES UTILIZANDO IA

*SYSTEM FOR THE DIAGNOSING OF SUPERFICIAL SKIN WOUNDS
THROUGH IMAGE PROCESSING BY USING AI*

Aldonso Becerra Sánchez

Universidad Autónoma de Zacatecas, México
a7donso@uaz.edu.mx

Armando Rodarte Rodríguez

Universidad Autónoma de Zacatecas, México
armandorodarte19@gmail.com

Alejandro Pinedo Barrios

Universidad Autónoma de Zacatecas, México
apinedobarrios@live.com

Gustavo Zepeda Valles

Universidad Autónoma de Zacatecas, México
gzepeda_@uaz.edu.mx

Elda García Mayorga

Universidad Autónoma de Zacatecas, México
emayorga3@hotmail.com

Santiago Esparza Guerrero

Universidad Autónoma de Zacatecas, México
chago@uaz.edu.mx

Recepción: 30/octubre/2020

Aceptación: 27/noviembre/2020

Resumen

Frecuentemente resulta difícil para un médico poder dar un diagnóstico certero y en poco tiempo sobre heridas superficiales en la piel, así como su evolución y mejoría, además esto puede ser subjetivo. Existen factores que pueden complicar esta valoración médica: experiencia, estudios y circunstancias de la herida. El objetivo de este trabajo es desarrollar un sistema que dé soporte al personal médico en el proceso de diagnóstico de heridas superficiales en la piel. Para cumplir el cometido, se propuso implementar diversos algoritmos para

extracción/representación de características, como SIFT, SURF, BoW; así como varios modelos de clasificación: KNN, SVM y CNN. Los resultados experimentales aún no son tan satisfactorios, al momento alcanzaron exactitudes del 31.25%. Se requiere modelar y realizar pruebas adicionales para obtener precisiones superiores. Sin embargo, los datos obtenidos alientan a probar variantes de los modelos y la conjunción de bancos de datos más robustos que generen mejores resultados.

Palabras Clave: Clasificación de heridas, heridas en la piel, inteligencia artificial, redes neuronales convolucionales, visión computacional.

Abstract

Sometimes it can be difficult for a physician, in a limited amount of time, to make an accurate diagnosis regarding superficial skin wounds, which comprises their evolution and improvement. There are factors that can complicate the medical evaluation, such as a physician's experience, previous studies, and the circumstances under which the wound occurred. The purpose of this work is to develop a system that supports physicians in the process of diagnosing superficial skin wounds. To achieve this goal, algorithms such as SIFT, SURF and BoW were implemented, as well as classification models such as KNN, SVM and CNN. Experimental results have yet to be satisfactory, so far they have only reached an accuracy level of 31.25%. Additional effort is required to obtain higher accuracy levels. However, the data obtained so far encourages testing variants of the models and the combination of more robust datasets that can eventually yield better results.

Keywords: Artificial intelligence, computer vision, convolutional neural networks, skin wounds, wound classification.

1. Introducción

En la actualidad la tecnología comienza a tener un impacto mayor en la vida cotidiana de las personas, desplazando cada vez más al ser humano en el desarrollo de tareas. Esto era imposible de imaginar hace una década debido a las limitaciones de la época. Existen cada vez más sistemas, aplicaciones y robots

basados en inteligencia artificial (IA) con el propósito de sustituir y apoyar al ser humano en la realización de ciertas actividades. La IA consiste en volver inteligente a un sistema artificial, principalmente a un sistema informático. Entendiendo por inteligente a un ente con la facultad de comprender y establecer relaciones entre elementos [Mathivet, 2017].

La visión por computadora es una de las disciplinas de la IA que permite procesar, analizar e interpretar imágenes, textos, videos y audios del mundo real en forma automática [García, 2002]. Si bien puede resultar fácil para un humano encontrar un elemento en una fotografía, lamentablemente para una computadora no lo es, debido a que la arquitectura cerebral es diferente a la de los algoritmos.

Por otra parte, el uso de las redes neuronales artificiales (RNA) ha venido a solucionar los problemas en los que una computadora suele ser menos eficiente que el cerebro humano. En las tareas de bajo nivel como el reconocimiento de patrones, de percepción y control, el cerebro es más eficiente. Es así que las RNA buscan emular la arquitectura biológica del cerebro humano [Del Brío, 2006].

Las redes neuronales convolucionales (CNN) son un tipo de RNA que tratan de emular las neuronas de la corteza visual del cerebro humano. Son usadas actualmente en la resolución de muchos problemas en diversas áreas de la visión artificial; además estos modelos aprenden a encontrar patrones y a realizar tareas de clasificación directamente a partir de imágenes, videos, textos y audios. En ocasiones llegan a superar la capacidad del ojo humano en este tipo de tareas.

En el área de medicina, existen varios factores que complican el diagnóstico de heridas superficiales: experiencia, especialidad de estudios, características de las heridas, entre otros. Estos factores influyen para clasificar una herida superficial y en identificar si ha tenido alguna mejoría, ya que para la capacidad del ojo humano puede resultar difícil detectar pequeños cambios, por lo cual se hace uso de la IA para apoyar en la solución de estas problemáticas.

En el rubro de la medicina, se pueden encontrar proyectos enfocados a la clasificación de imágenes. Uno de estos proyectos es el desarrollado por Wang et al., donde lograron reconocer úlceras mediante cápsulas endoscópicas inalámbricas utilizando redes neuronales convolucionales. En un principio utilizaron

un dataset con imágenes de 11 pacientes. El sistema consiste en una cápsula endoscópica, en donde el dataset está compuesto por 1,557 videos de úlceras y 259 videos que no poseen úlceras. Estos datos fueron probados en varios modelos de clasificación como HANet, ResNet, Inception-Resnet-V2, dense y SPM; donde lograron comprobar que las CNN tienen un mejor desempeño en este tipo de imágenes. En estos hallazgos, HANet obtuvo un 92.05% de precisión, mientras que las SVM lograron un porcentaje de 61.3% [Wang, 2019].

En este mismo sentido, Tran et al. desarrollaron un proyecto para el reconocimiento de grados de quemaduras en la piel tomado como entrada imágenes, a las cuales se les procesó mediante una red neuronal que ellos mismos propusieron, llamada Burn Convolutional Neural Network o B-CNN. Esta red utiliza la distribución de probabilidad de los datos de entrenamiento correspondientes a los grados de quemadura, para calcular la probabilidad de las muestras una por una correspondientes a las imágenes de entrenamiento de los grados de quemadura. Al aplicar este modelo, se obtuvieron resultados con tasas de error menores al 0.6%, lo cual resulta muy alentador [Tran, 2016].

Incluso existen proyectos enfocados a clasificar imágenes de heridas, como el desarrollado por Ohura et al., en donde utilizando CNN, lograron reconocer diversas heridas en la piel como úlceras en pie diabético, úlceras de presión y úlceras venosas en las piernas. Utilizaron un dataset que consta de 440 imágenes de estas heridas. En la fase de pruebas implementaron 4 modelos de CNN para comparar los resultados obtenidos por cada uno. de esta forma obtener el mejor modelo para este caso. Estos modelos fueron SegNet, LinkNet, U-Net y U-Net con VGG16 pre-entrenados con ImageNet (para lograr que ambas redes U-Net alcanzaran los mejores resultados, los cuales fueron del 99% de precisión) [Ohura, 2019].

Viendo estos trabajos previos, debería ser posible llevar este tipo de diagnóstico de manera eficiente y rápida mediante una categorización no sólo a casos específicos, sino llevarlos a un nivel más alto de abstracción. Un nivel donde sea posible clasificar heridas de diferentes tipos como, cortadas, mordeduras úlceras y quemaduras, además de sus diferentes niveles de gravedad.

Este trabajo tiene como objetivo desarrollar un sistema que dé soporte a personal médico diverso, permitiéndole un apoyo para un diagnóstico más rápido y preciso de los tipos más comunes de heridas superficiales en la piel. Para lograr eso, se implementan y prueban diferentes algoritmos, comenzando con mecanismos de extracción/representación de características como SIFT (Scale Invariant Feature Transform), SURF (Speed up Robust Feature), y BoW (bag of words). Las salidas generadas por estos son usadas por las técnicas de clasificación empleadas, tales como KNN (k-Nearest Neighbor), SVM (Support Vector Machines) y CNN; eligiendo los que brindan resultados prometedores para formar el sistema de diagnóstico. Los experimentos obtenidos aún tienen camino que recorrer, por lo que se requiere modelar y realizar pruebas adicionales para obtener precisiones superiores. No obstante, los datos generados brindan iniciativas para probar variantes de los modelos con nuevos bancos de datos que reditúen en mejoras considerables.

2. Métodos

La arquitectura general del sistema completo a proponer se basa en la figura 1, en donde el usuario proporciona como entrada una fotografía de la herida del paciente en turno, y en donde esta entrada es suministrada al sistema de diagnóstico de heridas superficiales.

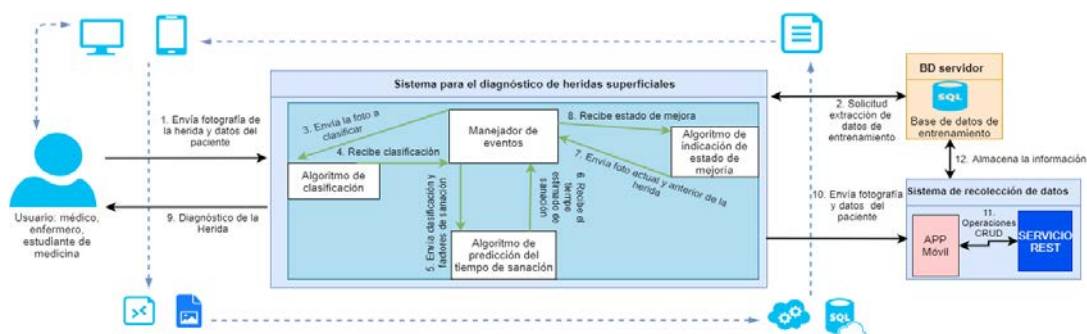


Figura 1 Diagrama general del sistema de diagnóstico de heridas superficiales.

Dentro del sistema se genera el proceso de clasificación, que interactúa con el mecanismo de manejador de eventos, el cual canaliza el flujo hacia el algoritmo de predicción de tiempo de sanación; posteriormente se regresa el flujo al manejador

para que pueda ser canalizado al algoritmo de indicación de mejoría. Todo este proceso se ve reflejado a través de la aplicación móvil y servicio REST para generar/almacenar la información respectiva, la cual será enviada como salida al usuario en un apoyo en el diagnóstico solicitado.

Los tipos de heridas superficiales en la piel a diagnosticar por el sistema se observan en la tabla 1. El total de clases que puede diagnosticar el sistema son 11. En el caso de las quemaduras y úlceras, el sistema no sólo indicará en concreto que tipo de herida es, sino también dirá el nivel de gravedad de la herida; y por ende, estos grados de gravedad se convierten en clases a clasificar.

El dataset usado para entrenar los algoritmos y realizar sus pruebas contiene 800 imágenes. Estas imágenes fueron re-clasificadas para ser agrupadas en las 11 clases que se mencionan en la tabla 1; en donde cada clase contiene una cantidad diferente de imágenes. Las imágenes fueron obtenidas del sitio web Medetec (banco abierto de imágenes médicas).

Tabla 1 Heridas superficiales en la piel.

Tipo de herida superficial en la piel	Gravedad o subtipo (clases a clasificar)	Cantidad de imágenes
Quemadura	Grado 1, grado 2, grado 3, grado 4	283
Úlcera	Grado 1, grado 2, grado 3, grado 4	213
Herida superficial	Mordedura, cortada lineal	185
Herida sana	Herida cicatrizada	119

Clasificación de imágenes

El proceso de clasificación de imágenes consiste en asignar etiquetas a una imagen; aunque el resultado puede ser muy subjetivo [Alegre, 2016]. Las etiquetas permitirán describir su contenido, como es el caso de las imágenes de heridas superficiales en la piel. Se puede ver en la figura 2 el proceso de clasificación, el cual está dividido en 3 subprocesos.

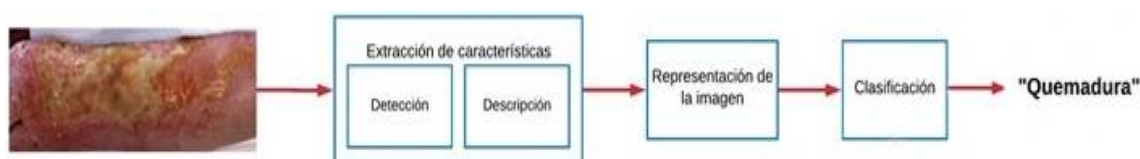


Figura 2 Esquema de clasificación.

El primer paso consiste en la extracción de características visuales y representativas de la imagen. El segundo paso obtiene una representación de la imagen, donde todas las características extraídas representarán de forma única a la imagen para el paso final. En el último paso se clasificará la imagen distintivamente de las otras clases con base en su representación [Valveny, 2016]. En las secciones posteriores se habla de forma más detallada de esta metodología utilizada para la clasificación de heridas en la piel, y cómo se formaron los distintos esquemas de clasificación.

Extracción de características

Este primer paso consiste en extraer las características visuales de la imagen. Las características deben ser representativas en varias imágenes, a la vez que son discriminativas. Cada característica es un vector matemático que describe una región de la imagen. Esta propiedad describe las características de un pixel en relación a sus pixeles vecinos. Da información sobre estructuras relevantes de la imagen, tales como regiones o contornos de la imagen; las cuales son obtenidas mediante un análisis matemático. Este paso se divide en dos subprocesos diferentes. En el primero se detectan los puntos de interés (características locales); zonas que dan información relevante sobre el contenido de la imagen. Los puntos de interés estarán localizados en zonas significativas, ver figura 3 [Valveny, 2016]. Los puntos de interés pueden encontrarse también en cambios significativos de contraste o bien en los contornos de los objetos. El segundo subproceso consiste en la descripción de estos puntos de interés (cada característica tendrá su descriptor); es decir, una descripción visual de una región de la imagen. Igualmente, el descriptor será un vector matemático, como se observa en ecuación 1.

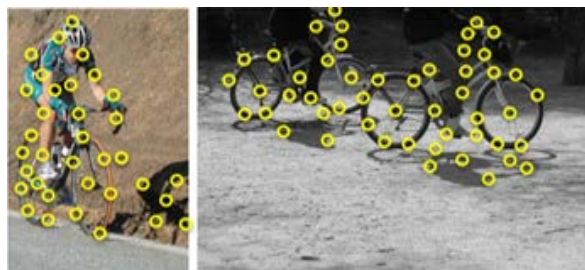


Figura 3 Detección de puntos de interés.

$$\begin{aligned}
 P_1 &= (p_{11}, p_{12}, \dots, p_{1n}) \\
 P_2 &= (p_{21}, p_{22}, \dots, p_{2n}) \\
 &\vdots \\
 P_n &= (p_{m1}, p_{m2}, \dots, p_{mn})
 \end{aligned}
 \tag{1}$$

Donde:

- P_n : Punto de interés.
- p_{mn} : Vector numérico que describe el punto de interés.

Una manera de detectar los puntos de interés de una forma genérica es mediante la aplicación de un filtro o kernel a cada pixel de la imagen. La forma matemática para aplicar este kernel o filtro es mediante la operación de la convolución.

Un filtro, como se muestra en la figura 4, define una serie de pesos, los cuales se aplicarán a los vecinos de un determinado pixel. Se obtendrá un valor que dará la respuesta a la característica local que está definiendo el filtro [Valveny, 2016].

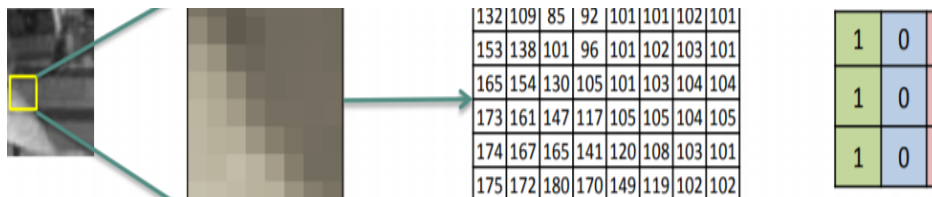


Figura 4 Aplicación del filtro o kernel.

La convolución es usada en la teoría de la señal para descomponer una señal en una serie de señales más simples (Ecuación 2).

$$(f * g)(x, y) = \sum_{x', y'} g(x', y') * f(x - x', y - y')
 \tag{2}$$

Donde:

- f : Imagen.
- g : Filtro.
- x' : Coordenada en x del filtro.
- y' : Coordenada en y del filtro.
- x : Coordenada en x de la imagen.
- y : Coordenada en y de la imagen.

En visión artificial, la función f representa la imagen y la función g es el filtro que representa la característica local que se quiere detectar. El sumatorio está definido

por el tamaño del filtro; en pocas palabras, la convolución mide el grado de solapamiento entre un punto de la imagen y el filtro.

Representación de la imagen

El segundo paso (Figura 2) para lograr la clasificación de la imagen de entrada consiste en obtener una representación única de la imagen mediante un vector numérico. Este resultado se obtiene combinando o agregando los puntos de interés más relevantes de la imagen. Este procedimiento se puede ver en la ecuación 3.

$$X = (x_1, x_2, \dots, x_n) \quad (3)$$

Donde:

- X : Vector numérico representante de la imagen.
- x_n : Vector numérico que describe el punto de interés.

Clasificación de la imagen

El paso final consiste en aprender cómo distinguir la imagen de las diferentes categorías a partir de un conjunto de imágenes de entrenamiento. Este proceso parte de la representación visual en cuestión, asignándole la o las etiquetas correspondientes que describan la imagen. El objetivo es poder asignar la etiqueta correspondiente a la herida superficial en la piel (en la cual se quiere saber su tipo) tomando como referencia su fotografía como entrada. Por ejemplo, en la figura 2 se ingresa la imagen de una quemadura, y se le asigna correctamente la etiqueta “quemadura”, indicando adecuadamente el tipo de herida en la piel [Valveny, 2016]. En las siguientes secciones se hablará de cómo se formaron los distintos esquemas de clasificación que fueron implementados; los cuales siguen el proceso de clasificación de imágenes.

Esquemas de clasificación implementados

En la actualidad existen muchos algoritmos genéricos para cada parte (etapa) del esquema de clasificación, como se observa en la tabla 2. Se formaron diferentes enfoques aplicando distintos algoritmos en cada paso del esquema, a la vez que trataba de especializar en la clasificación de heridas superficiales.

La elección y ajuste de los hiperparámetros está basada en diferentes técnicas como optimización para algoritmos genéticos, optimización bayesiana y para aprendizaje automático. Además, se consideró para su elección un conjunto de trabajos similares ya realizados [Loussaief, 2018][Wu,2019] [Tran, 2016].

Se ponen a prueba diferentes esquemas de clasificación de imágenes, donde el objetivo es obtener como solución el esquema de clasificación más efectivo en el contexto de heridas superficiales. La métrica para elegir el mejor clasificador toma como base la exactitud de clasificación, rapidez y el coste computacional del esquema. Algunos de estos esquemas implementan varios algoritmos.

Tabla 2 Algoritmos de clasificación de imágenes.

Detector y descriptor de características	Representación de la imagen	Clasificación
SIFT	Conjunto de puntos	K-NN
SURF	BoW, descriptor a color	SVM
CNN	Fusión de representaciones	Combinación de clasificadores
	CNN	CNN

Primer esquema de clasificación: KNN

Lo que realiza la computadora para clasificar una imagen es obtener sus puntos de interés y descriptores (vectores numéricos representantes de la imagen).

Después se realiza una búsqueda de patrones coincidentes con el objeto en los vectores numéricos representantes de cada imagen. El esquema básico para la clasificación (Figura 5) utiliza el algoritmo SIFT y SURF para la extracción de puntos de interés y sus descriptores. Para la representación de la imagen se usa la técnica de conjunto de puntos, y como clasificador se utiliza el algoritmo KNN.

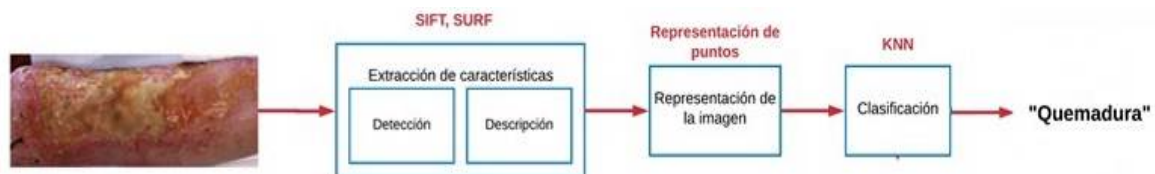


Figura 5 Primer esquema de clasificación.

En estos esquemas se usó $k = 11$ vecinos más cercanos para poder aplicar la prueba de razón explicada por D. Lowe [Lowe, 2004], con un valor de 0.9.

Se puede ver en la figura 6, que, en ambos casos, la mayoría de los puntos de interés se encuentran dentro de la zona de la herida (con los mejores resultados). SIFT es capaz de detectar detalles menores de la imagen (ver figura 6a); en otras palabras, puntos de interés que abarcan zonas pequeñas y descriptivas de la imagen. SURF [Bay, 2006] trabaja en zonas más grandes de la imagen, lo cual puede ser una ventaja contra SIFT [Lowe, 2004]. Sin embargo, las zonas de interés más grandes suelen ser menos representativas que las pequeñas.

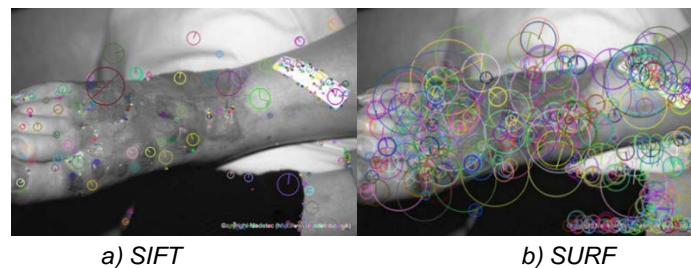


Figura 6 Obtención de puntos de interés.

En la figura 7a se pueden ver gráficamente las mejores k correspondencias de los puntos de interés detectados usando SIFT. En la figura 7b se observa lo mismo, pero usando SURF. Ambos algoritmos obtienen k mejores coincidencias.

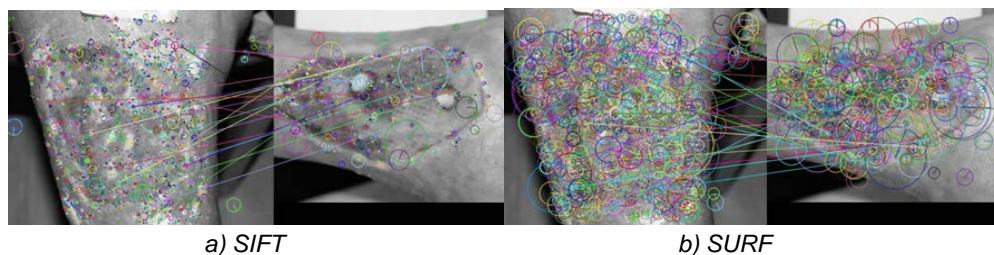


Figura 7 Representación de puntos en correspondencia.

Segundo esquema de clasificación: SVM

En este segundo esquema de clasificación, como se muestra en la figura 8, también se hace uso de diferentes algoritmos. Para la fase de extracción de características se usó el algoritmo SIFT y SURF. Para la representación de la imagen se utilizó el algoritmo BoW (bag of words), en cuanto a la parte de clasificador se usó SVM (Support Vector Machines).

Al igual que en el esquema anterior, se hace uso de la detección de puntos de interés. A diferencia del esquema pasado, no solo se obtiene el punto de interés con su descriptor, también se obtiene su descriptor de color para la formación del vocabulario visual de puntos de interés para el BoW.

El número de palabras visuales es $k = 72$, y el número de puntos de interés para formar las palabras visuales es de $n = 70,000$. El factor de regularización para el SVM es $C = 1$. Con estos valores se obtuvieron los mejores resultados.

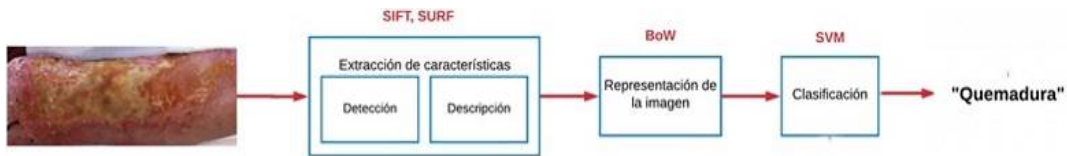


Figura 8 Segundo esquema de clasificación.

Tercer esquema de clasificación: CNN

En la figura 9 se observa la propuesta inicial del modelo de CNN. Este modelo recibe una imagen de $64 \times 64 \times 3$, en donde se procesa con la capa a color. La CNN tiene dos capas en la fase de entrenamiento de características. Cada capa tiene dos elementos de convolución con un kernel de tamaño 3×3 , con 32 filtros; cada convolución tiene la función de activación ReLu. Después se aplica un pooling de 2×2 en cada capa. En la fase de clasificación se tiene una capa oculta totalmente conectada de 128 neuronas con la función de activación ReLu.

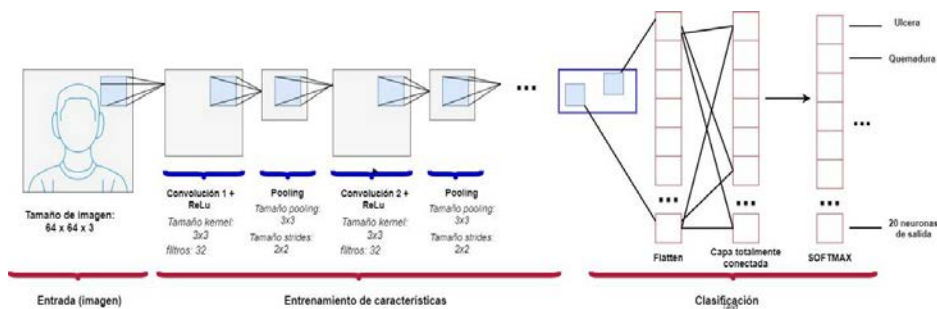


Figura 9 Esquema de CNN propuesto.

La capa de salida está compuesta de 11 neuronas (cantidad de clases) con una función de activación softmax. Se usó el algoritmo Adam para el aprendizaje de la

red, con una tasa de aprendizaje de 0.0001 y un momentum de 0.0. La cantidad de épocas de entrenamiento fue de 25; en donde no se utilizó algún pre-entrenamiento.

Cuarto esquema de clasificación: Inception V3

Inception V3 (Figura 10) es un modelo de reconocimiento de imágenes muy utilizado, y en donde se puede alcanzar una exactitud considerable. El modelo representa la culminación de muchas ideas que desarrollaron varios investigadores durante años [Szegedy, 2015].

El modelo Inception tiene una combinación de bloques de compilación simétricos y asimétricos, y que incluyen: convoluciones, avg pooling, max pooling, concatenaciones, dropout, capas totalmente conectadas y capas softmax. Se usó el algoritmo RMSprop para el pre-entrenamiento de la red con el dataset por defecto del modelo, y se aplicó una tasa de aprendizaje con un valor de 0.0008 y un momentum de 0.0. La cantidad de épocas de entrenamiento fue de 25.

Después se realizó el entrenamiento con el conjunto de imágenes médicas utilizando el algoritmo SGD (Stochastic Gradient Descent) para el aprendizaje de la red, empleando una tasa de aprendizaje con un valor de 0.009 y un momentum de 0.9. La cantidad de épocas de entrenamiento fue de 20. Se usó una capa oculta de 1,024 neuronas con la función de activación Relu, y una capa tipo softmax con 11 neuronas de salida.

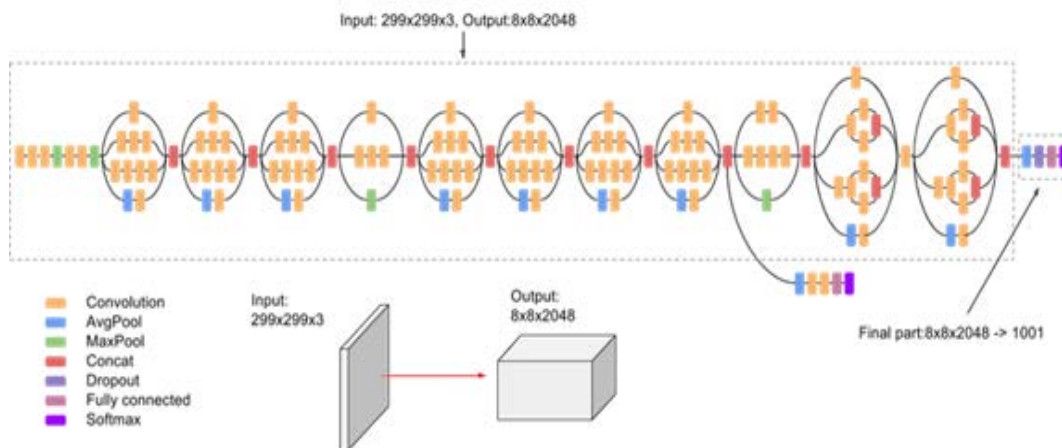


Figura 10 Modelo Inception V3.

Aplicación de recolección de datos

Dado que el dataset no es tan amplio, fue necesario el desarrollo de una aplicación móvil (Figura 11) que permita recolectar los datos necesarios para la formación del nuevo dataset. Esta aplicación permitirá a los médicos recolectar información sobre casos clínicos tratados con heridas superficiales, tal que esta información permita robustecer el entrenamiento de los algoritmos de IA.

En la figura 12 se muestra el diseño propuesto de la base de datos mediante un diagrama relacional. El modelo contiene las tablas para la aplicación de recolección de datos. En este esquema se almacenarán datos de pacientes, médicos, centros clínicos, así como sus respectivos domicilios. Además de almacenar casos médicos, revisiones, tipos de heridas y factores de sanación.

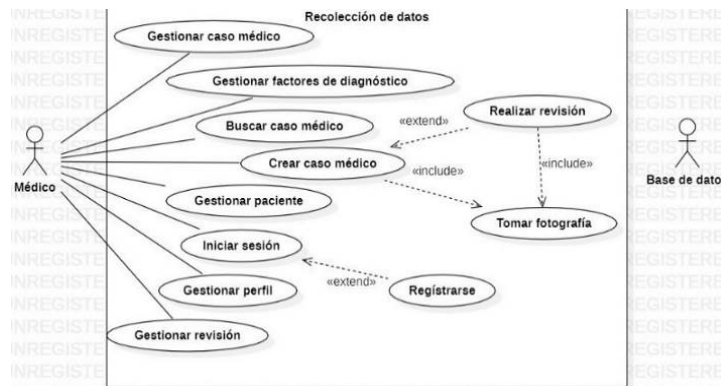


Figura 11 Casos de uso de la aplicación.

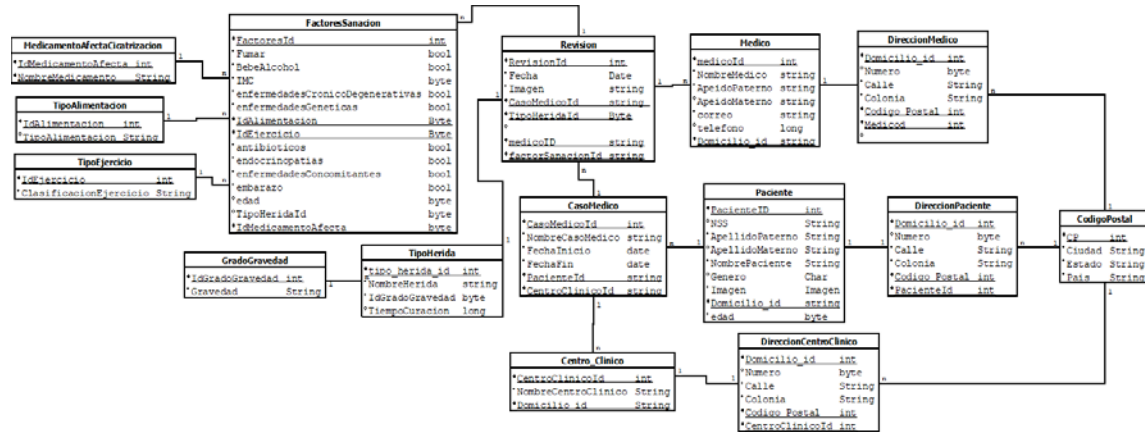


Figura 12 Modelo relacional de BD.

Se puede apreciar que los casos médicos pueden tener diversas revisiones, y a su vez cada nueva revisión puede ser realizada por un médico diferente. Además, en cada revisión se va almacenando el tipo de herida y los factores de sanación contemplados. Esto se obtiene para monitorear el estado de las heridas de una forma más precisa y para identificar si estos cambios se deben a que el paciente tiene hábitos que pueden influir en la cicatrización de la herida, para bien o mal.

3. Resultados

Los resultados de los diferentes esquemas mostrados en esta sección son parciales hasta el momento, dado que se seguirá trabajando en mejorar las técnicas de clasificación y de formación del corpus de datos. Además, se hará la implementación de los módulos aún en desarrollo (estado de mejoría, predicción del tiempo de sanación). El corpus de imágenes utilizado fue el descrito con anterioridad, el cual fue dividido en dos conjuntos (entrenamiento y prueba). El conjunto de entrenamiento contiene el 75% de las imágenes de cada categoría, el cual corresponde a un total de 600 imágenes. Mientras que el conjunto de prueba contiene el 25% restante del corpus de datos original, es decir, 200 imágenes.

La figura 13a representa los resultados del primer esquema básico de clasificación utilizando SURF y SIFT como detectores de puntos de interés, y KNN como clasificador. Donde se obtuvo una exactitud de 16.31% de imágenes clasificadas. Para cierto tipo de imágenes, este clasificador básico puede resultar muy efectivo. Sin embargo, en este caso resulta ineficiente, por lo cual es necesario buscar otros esquemas de clasificación más robustos. Estos resultados hacen considerar que sí es necesario tomar características espaciales como: color, ubicación, etc.

```
categoria: meningitisMeningococica imagenes clasificadas correctamente: 0.0%
categoria: ulcerasPresion Imagenes clasificadas correctamente: 0.0%
categoria: wounds Imagenes clasificadas correctamente: 20.0%
categoria: heridas Imagenes clasificadas correctamente: 38.775510204081634%
categoria: ulcerasArterialesPlena Imagenes clasificadas correctamente: 0.0%
categoria: heridasMiscelaneas Imagenes clasificadas correctamente: 0.0%
categoria: extravasacion Imagenes clasificadas correctamente: 0.0%
categoria: heridasLeves Imagenes clasificadas correctamente: 0.0%
categoria: ortopedicas Imagenes clasificadas correctamente: 0.0%
categoria: dedosInfectados Imagenes clasificadas correctamente: 11.111111111111111%
categoria: ulcerasPies Imagenes clasificadas correctamente: 0.0%
categoria: hemangioma Imagenes clasificadas correctamente: 0.0%
16.317991631799163% de img clasificadas correctamente
Done in 9413.793374300003 secs.
```

a) Primer esquema

```
Extracting local descriptors
Done in 30.350790977478027 secs.
Computing kmeans on 70000 samples with 52 centroids
Done in 1.5853638648986816 secs.
Extracting visual word representations
Done in 0.315227746963501 secs.
Extracting Local Descriptors
Done in 11.551965236663818 secs.
Extracting visual word representations
Done in 0.08343887329101562 secs.
Training and Testing a linear SVM
Done in 0.09159111976623535 secs.
Accuracy BOVW: 26.953125
```

b) Segundo esquema

Figura 13 Resultados obtenidos con esquemas que utilizan SIFT y SURF.

La figura 13b muestra los resultados para el segundo esquema de clasificación (con Bow y SVM). En él se obtuvo una exactitud del 26%, casi el doble en relación con el primer esquema. Ambos algoritmos han proporcionado resultados muy por debajo de lo esperado, siendo ineficientes para este tipo de imágenes. Estos resultados dan indicativo de que es posible clasificar estas heridas, ya que tomando una característica espacial (color) mejoraron los resultados. Además, para tener mejores resultados es necesario probar con otros modelos como las CNN, las cuales tienen una mayor robustez.

Esquema CNN propuesto y modelo InceptionV3

En la figura 14a se muestran los resultados obtenidos por época por el modelo de CNN, en donde se puede observar un 28.91% de exactitud con los datos de prueba, correspondientes a su última época; y un 56.31% de exactitud en los datos de entrenamiento. Se observa que existe sobre-ajuste, ya que la diferencia de exactitud entre los datos de entrenamiento y prueba es considerable.

En la figura 14b se muestran los resultados obtenidos por el modelo de InceptionV3. En donde se puede observar que se obtuvo, en su última época, un 31.25% de exactitud con los datos de prueba. Existe una muy grande mejoría en la exactitud de los datos de entrenamiento (un 94.88%), superando al modelo anterior. Los resultados mejoraron en comparación con los modelos anteriores, pero aún se contempla que todavía existe sobre-ajuste, ya que la diferencia de exactitud entre los datos de entrenamiento y prueba es considerable.

```
loss: 2.000 - categorical_accuracy: 0.3302 - val_loss: 2.3485 - val_categorical_accuracy: 0.2227
loss: 2.1004 - categorical_accuracy: 0.3039 - val_loss: 2.3866 - val_categorical_accuracy: 0.2422
loss: 1.9825 - categorical_accuracy: 0.3546 - val_loss: 2.4067 - val_categorical_accuracy: 0.2381
loss: 1.7526 - categorical_accuracy: 0.4112 - val_loss: 2.5948 - val_categorical_accuracy: 0.2656
loss: 1.8057 - categorical_accuracy: 0.3848 - val_loss: 2.5062 - val_categorical_accuracy: 0.2461
loss: 1.7035 - categorical_accuracy: 0.4375 - val_loss: 2.7871 - val_categorical_accuracy: 0.2656
loss: 1.6137 - categorical_accuracy: 0.4645 - val_loss: 2.7936 - val_categorical_accuracy: 0.2734
loss: 1.4689 - categorical_accuracy: 0.5263 - val_loss: 2.5988 - val_categorical_accuracy: 0.2824
loss: 1.4628 - categorical_accuracy: 0.4812 - val_loss: 2.7145 - val_categorical_accuracy: 0.2969
loss: 1.3488 - categorical_accuracy: 0.5529 - val_loss: 2.7088 - val_categorical_accuracy: 0.2969
loss: 1.2833 - categorical_accuracy: 0.5666 - val_loss: 2.7465 - val_categorical_accuracy: 0.2812
loss: 1.2643 - categorical_accuracy: 0.5631 - val_loss: 2.9421 - val_categorical_accuracy: 0.2891
```

a) CNN

```
loss: 0.1701 - categorical_accuracy: 0.9326 - val_loss: 3.7438 - val_categorical_accuracy: 0.3086
loss: 0.0963 - categorical_accuracy: 0.9519 - val_loss: 4.1301 - val_categorical_accuracy: 0.2691
loss: 0.1465 - categorical_accuracy: 0.9454 - val_loss: 4.1315 - val_categorical_accuracy: 0.3086
loss: 0.1514 - categorical_accuracy: 0.9317 - val_loss: 4.2853 - val_categorical_accuracy: 0.3086
loss: 0.1308 - categorical_accuracy: 0.9556 - val_loss: 4.3629 - val_categorical_accuracy: 0.3283
loss: 0.1717 - categorical_accuracy: 0.9317 - val_loss: 3.7657 - val_categorical_accuracy: 0.2530
loss: 0.1021 - categorical_accuracy: 0.9539 - val_loss: 3.6224 - val_categorical_accuracy: 0.2891
loss: 0.1584 - categorical_accuracy: 0.9283 - val_loss: 4.9007 - val_categorical_accuracy: 0.2695
loss: 0.1284 - categorical_accuracy: 0.9488 - val_loss: 4.5630 - val_categorical_accuracy: 0.3125
```

b) Inception V3

Figura 14 Resultados de las redes neuronales.

A pesar de que los resultados están por debajo de los esperados con ambos esquemas, se puede apreciar que es posible la clasificación para estos 11 tipos de heridas utilizando CNN. Sin embargo, para poder disminuir el porcentaje de error, es necesario utilizar un dataset más robusto, homogéneo y con mayor cantidad de datos (imágenes) del que se utilizó en su momento. Haciendo uso de este, se podría evitar el sobre-ajuste de la CNN y aumentar significativamente su exactitud.

4. Discusión

Las pruebas realizadas para los diferentes algoritmos no mostraron los resultados esperados, dado su bajo porcentaje de exactitud. Los primeros dos esquemas (KNN y SVM) tuvieron un resultado ineficiente con la clasificación, lamentablemente carecen de robustez, y no toman en cuenta características espaciales, lo cual es algo importante en este contexto de clasificación.

Al trabajar con las CNN e InceptionV3, se notó una mejoría en los porcentajes de exactitud durante la fase de entrenamiento; lamentablemente en la fase de pruebas no se mantuvo la misma exactitud. Ambos algoritmos obtuvieron resultados muy por debajo de los obtenidos durante el entrenamiento, en donde InceptionV3 clasificó correctamente cerca del 31%, contra un 28.91% obtenido de la CNN. Este comportamiento radica en la existencia de sobre-aprendizaje.

En los distintos esquemas implementados se puede observar que es posible clasificar las heridas en la piel a este nivel de abstracción (contemplando heridas de diferente tipo y grados de gravedad) a pesar de sus bajos resultados. Sin embargo, para mejorar considerablemente la exactitud de clasificación, para alcanzar los resultados óptimos y evitar el sobre-aprendizaje en los algoritmos es necesario formar un banco de imágenes más grande, homogéneo, más representativo y robusto. También es necesario tomar en cuenta características espaciales para este tipo de imágenes y así mejorar la exactitud de clasificación, además de probar nuevas técnicas. Se puede ver que los esquemas que toman en cuenta características espaciales tienen una mayor exactitud sobre los esquemas básicos, los cuales no tienen esa robustez. Por todo esto, se optó por crear una aplicación móvil para la recolección de otro dataset en busca de prevenir el sobre-aprendizaje

y continuar con el desarrollo del trabajo, el cual incluye predicción de tiempos de sanación y mejoría de la herida.

5. Conclusiones

En este proyecto se propuso la realización de un sistema que permite diagnosticar heridas superficiales. Para el módulo del clasificador de heridas médicas se probaron diferentes esquemas de clasificación, comparando los resultados para ir bosquejando el modelo que se pueda considerar como base. Para este trabajo se hizo uso de un conjunto abierto de 800 imágenes médicas para el entrenamiento de los algoritmos y para sus pruebas. Dado que los resultados de los esquemas implementados no cumplen las expectativas, se planteó a futuro proponer nuevos esquemas, mejorar las técnicas y optimizar los algoritmos implementados. También se propone la construcción de un nuevo conjunto de datos más robusto para el entrenamiento de estos modelos; esto mediante la aplicación móvil de recolección de datos. Un conjunto de datos más representativo y robusto mejorará la exactitud de los algoritmos implementados. Por último, se propone en un futuro la implementación de los módulos que indican la sanación de una herida y predicción del tiempo de sanación de la misma.

6. Bibliografía y Referencias

- [1] Alegre E., Pajares G., et al., (2016). *Conceptos y métodos en Visión por Computador*. España: CEA.
- [2] Bay H., Tuytelaars T., et al., (2006). SURF: Speeded Up Robust Features. *Lecture Notes in Computer Science*, vol. 3951. Springer, Berlin, Heidelberg.
- [3] Del Brío B. M., & Sanz A., (2006). *Redes neuronales y sistemas borrosos*, España: Universidad de Zaragoza.
- [4] García E. M., (2002). *Visión Artificial*, Fund. Univ. Oberta Catalunya p. 115.
- [5] Lowe D. G., (2004). Distinctive Image Features from Scale-Invariant Keypoints, *International Journal of Computer Vision*, no. 60, 91–110.
- [6] Mathivet V., (2017). *Inteligencia Artificial para desarrolladores conceptos e implementación en Java*. España: Ediciones ENI.

- [7] Loussaief S., & Abdelkrim A., (2018). Convolutional Neural Network Hyper-Parameters Optimization based on Genetic Algorithms. *International Journal of Advanced Computer Science and Applications*. Vol. 9, 252-266.
- [8] Ohura N., Mitsuno R., et al., (2019). Convolutional neural networks for wound detection: the role of artificial intelligence in wound care. *Journal of Wound Care*, no. 28, 13-24.
- [9] Szegedy C., Vanhoucke V., et al., (2015). Rethinking the Inception Architecture for Computer Vision.
- [10] Tran H.S., Le T. H., et al., (2016). The Degree of Skin Burns Image Recognition Using Convolutional Neural Network. *Indian J. Sci. Technol.* Vol. 9, no. 45.
- [11] Valveny E., (2016). Clasificación de imágenes: ¿Cómo reconocer el contenido de una imagen?, Universidad Autónoma de Barcelona, [Online]: <https://www.coursera.org/learn/clasificacion-imagenes/home/welcome>.
- [12] Wang S., Xing Y., et al., (2019). Deep Convolutional Neural Network for Ulcer Recognition in Wireless Capsule Endoscopy: Experimental Feasibility and Optimization. *Computational and Mathematical Methods in Medicine*, Vol. 2019.
- [13] Wu J., Chen X., et al., (2019). Hyperparameter Optimization for Machine Learning Models Based on Bayesian Optimization. *Journal of Electronic Science and Technology*. vol.17, 26-40.