Detección de temperatura corporal mediante imágenes de rostro usando una cámara teléfono inteligente para la contención de COVID-19 utilizando una red neuronal convolucional (CNN)y maquina de vector de soporte (SVM)

Juan José Laínez Bolaños
jjlainez@espol.edu.ec

Resumen*— En este proyecto, proponemos contener el COVID-19 y otras enfermedades contagiosas usando una APP obteniendo la temperatura corporal a partir de una foto de una cámara digital, una cámara web de un computador o una cámara de un teléfono inteligente, siendo como objetivo extraer características faciales con un modelo de red neuronal convolucional (CNN) y entrenar un clasificador de máquina de vector de soporte (SVM) para predecir la temperatura a partir de una imagen. CNN es un tipo especial de redes neuronales multicapa diseñadas para reconocer patrones visuales directamente de las imágenes naturales de píxeles, mientras que SVM puede realizar de manera eficiente una clasificación no lineal, especialmente con un pequeño conjunto de datos. El resto de este documento está organizado de la siguiente manera:La Seccion 1 Introducción , La Sección 2 Procesamiento de la imagen,La Sección 3 Modelo de CNN y SVM, La sección 4 Metodología y resultados, La sección 5 Conclusiones .

Keywords—COVID-19, CNN, SVM

I. INTRODUCCION

El coronavirus (COVID-19) se está extendiendo rápidamente por todo el mundo desde diciembre de 2019. Una de las medidads de contención es el monitoreo de lo sintomas más comunes en un población. La temperatura del cuerpo humano es un signo vital importante y varias enfermedades se caracterizan por un cambio en la temperatura corporal. La fiebre, que es uno de los síntomas más importantes del COVID-19, es un aumento temporal de la temperatura corporal. Es una parte de la respuesta general del sistema inmunitario del cuerpo. Por lo general, la fiebre se debe a una infección. El rango normal de temperatura del cuerpo humano generalmente se establece como 36,5 a 37,5 °C.

Actualmente la temperatura se mide con varios tipos de termómetros médicos, ver Figura 1, así como lugares en el cuerpo utilizados para la medición. Una forma típica es poner un termómetro digital en la boca (temperatura oral). Esto es relativamente conveniente en contraste con la medición de la temperatura axilar o la temperatura rectal. Sin embargo, todas estas mediciones se realizan por contacto, lo que no es adecuado para monitorear una multitud de enfermedades contagiosas. Dada la pandemia es importante que la temperatura corporal sea de una medicion sin contacto un termómetro frontal infrarrojo sin contacto es un dispositivo portátil asequible para medir la temperatura individual, mientras que un sistema de cámara térmica es viable (pero muy costoso) para monitorear multitudes en sitios públicos por ejemplo, en centros comerciales, aeropuertos , universidades.

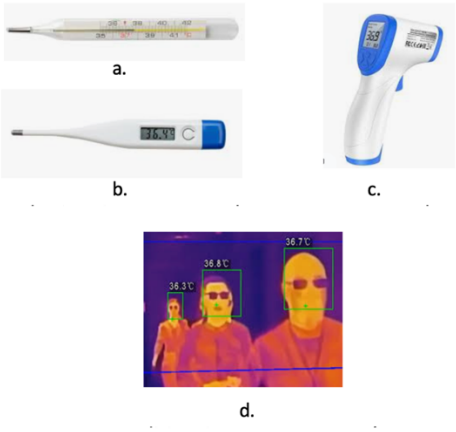

Figura 1. Equipos para la medición de temperatura, a. termómetro analógico, termómetro digital de contacto, c. termómetro infrarrojo, d. cámara térmica.

II. PROCESAMIENTO DE IMÁGENES

El proceso completo se divide en tres pasos: obtención de imagen, detección de rostros, extracción de características y reconocimiento de temperatura de rostros. La detección de rostros se centra en la detección de rostros humanos frontales. Las técnicas de extracción de características ayudan a localizar y extraer características faciales únicas. El sistema de reconocimiento facial utiliza métodos basados en aprendizaje automático para clasificar las características extraídas en una de las clases (perteneciente a un individuo en particular) en la base de datos.

La obtención de la imagen para el procesamiento es por medio de una cámara digital de cctv, un cámara web de computadora o una cámara de teléfono inteligente. Esta imagen se envia a una ubicacion en el servidor para ser procesada por el algoritmo del software que utiliza la CNN y la SVM. La detección de rostros consiste en localizar el rostro en la imagen, lo cual es un paso importante para el proceso sucesivo. Se han desarrollado numerosas técnicas para detectar rostros en una sola imagen, por ejemplo, métodos basados en el conocimiento, enfoques faciales invariantes de características, métodos de coincidencia de plantillas, métodos basados en la apariencia.

Muchos métodos utilizan información de color, es decir, identifican las regiones (mapa de piel) cuyo color es similar al color de la piel, lo que puede restringir significativamente el área de búsqueda.

Figura 2. Procesamiento de Imágenes

Uno de los algoritmos más exitosos en imágenes visuales es el algoritmo Viola-Jones propuesto en 2001. Los métodos de detección de rostros están bien desarrollados y marcan rápidamente múltiples rostros en una imagen, independientemente de sus tamaños y fondos, utilizando cambios espaciales. El algoritmo Viola-Jones (VJ) se ha convertido en un método muy común de detección de objetos, incluida la detección de rostros.

Viola y Jones propusieron este algoritmo como un enfoque de aprendizaje automático para la detección de objetos con énfasis en obtener resultados rápidamente y con altas tasas de detección.

III. Modelo de CNN y SVM

A. *CNN MTCNN*

Las redes convolucionales en cascada multitarea (MTCNN) es un marco desarrollado como una solución tanto para la detección de rostros como para la alineación de rostros. El proceso consta de tres etapas de redes convolucionales que pueden reconocer rostros y ubicaciones de puntos de referencia, como ojos, nariz y boca. El documento propone MTCNN como una forma de integrar ambas tareas (reconocimiento y alineación) utilizando el aprendizaje multitarea. En la primera etapa, utiliza una CNN poco profunda para producir rápidamente ventanas candidatas. En la segunda etapa refina las ventanas candidatas propuestas a través de una CNN más compleja. Y, por último, en la tercera etapa utiliza una tercera CNN, más compleja que las demás, para refinar aún más el resultado y generar posiciones de puntos de referencia faciales.

- *Etapa 1: La Red de Propuestas (P-Net)*

  Esta primera etapa es una red totalmente convolucional (FCN). La diferencia entre una CNN y una FCN es que una red totalmente convolucional no utiliza una capa densa como parte de la arquitectura. Esta red de propuestas se utiliza para obtener ventanas candidatas y sus vectores de regresión de cuadro delimitador.

  La regresión de cuadro delimitador es una técnica popular para predecir la localización de cuadros cuando el objetivo es detectar un objeto de alguna clase predefinida, en este caso caras. Después de obtener los vectores del cuadro delimitador, se realiza un refinamiento para combinar regiones superpuestas. El resultado final de esta etapa son todas las ventanas de candidatos después del refinamiento para reducir el volumen de candidatos.

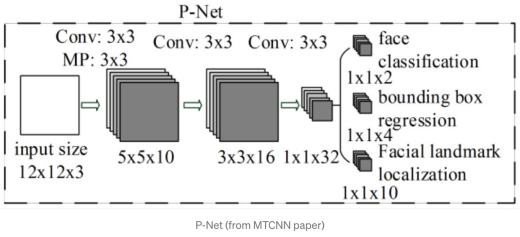

Figura 3. Procesamiento de Imágenes, P-Net

- *Etapa 2: La red Refine (R-Net)*

  Todos los candidatos de P-Net ingresan a Refine Network. Tenga en cuenta que esta red es una CNN, no una FCN como la anterior, ya que hay una capa densa en la última etapa de la arquitectura de la red. R-Net reduce aún más el número de candidatos, realiza la calibración con regresión de cuadro delimitador y emplea supresión no máxima (NMS) para fusionar candidatos superpuestos.

  Las salidas de R-Net, ya sea que la entrada sea una cara o no, son un vector de 4 elementos que es el cuadro delimitador para la cara y un vector de 10 elementos para la localización de puntos de referencia faciales.

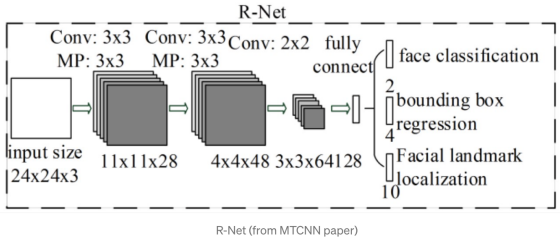

Figura 4. Procesamiento de Imágenes, R-Net

- Etapa 3: La Red de Salida (O-Net)

  Esta etapa es similar a R-Net, pero esta red de salida tiene como objetivo describir la cara con más detalle y generar las posiciones de los cinco puntos de referencia faciales para los ojos, la nariz y la boca.

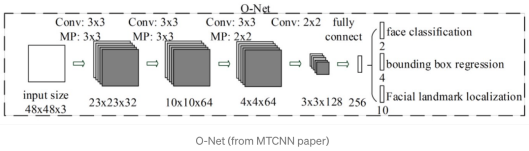

Figura 5. Procesamiento de Imágenes, O-Net

Las tres tareas de MTCNN, la tarea de la red es generar tres cosas: clasificación de rostros/no rostros, regresión de cuadro delimitador y localización de puntos de referencia faciales. 1.- Clasificación de caras: este es un problema de clasificación binaria que utiliza pérdida de entropía

cruzada. 2. Regresión de cuadro delimitador: el objetivo de aprendizaje es un problema de regresión. Para cada ventana candidata, se calcula el desplazamiento entre la candidata y la realidad del terreno más cercana. La pérdida euclidiana se emplea para esta tarea. 3. Localización de puntos de referencia faciales: la localización de puntos de referencia faciales se formula como un problema de regresión, en el que la función de pérdida es la distancia euclidiana. Hay cinco puntos de referencia: ojo izquierdo, ojo derecho, nariz, comisura izquierda de la boca y comisura derecha de la boca.

B. *SVM*

La máquina de vectores de soporte (SVM) puede realizar de manera eficiente una clasificación no lineal, especialmente con un conjunto de datos pequeño. Las SVM son modelos de aprendizaje supervisado con algoritmos de aprendizaje asociados que analizan los datos utilizados para la clasificación y el análisis de regresión. Dado un conjunto de ejemplos de entrenamiento, cada uno marcado como perteneciente a una u otra de dos (o múltiples) categorías, un algoritmo de entrenamiento SVM construye un modelo que asigna nuevos ejemplos a una categoría u otra, convirtiéndolo en un binario lineal no probabilístico.

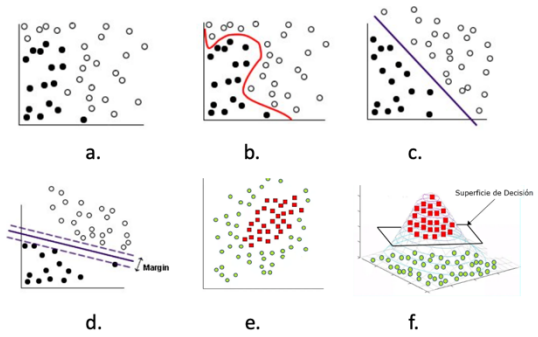

Figura 6. Datos originales y transformados usando SVM

Una SVM construye un hiperplano o un conjunto de hiperplanos en un espacio de dimensión alta o infinita para clasificación o regresión. Intuitivamente, una buena separación la logra el hiperplano que tiene la mayor distancia al punto de datos de entrenamiento más cercano de cualquier clase (el llamado margen funcional), ya que en general cuanto mayor es el margen, menor es el error de generalización del clasificador. Mientras que el problema original definido en un espacio de dimensión finita puede no ser linealmente separable en ese espacio. Por esta razón, el espacio original de dimensión finita se puede mapear en un espacio de dimensión mucho más alta, lo que presumiblemente facilita la separación en ese espacio. SVM multiclase tiene como objetivo asignar etiquetas a las instancias mediante el uso de SVM, donde las etiquetas se extraen de un conjunto finito de varios elementos.

El enfoque dominante para hacerlo es reducir el problema multiclase único a múltiples problemas de clasificación binaria.

En La Figura 6, a. muestra los datos originales, b. datos originales con separador añadido, c. muestra los datos transformados, la función matemática utilizada para la transformación se conoce como función kernel. Los tipos de kernel son :Lineal, Polinómico, Función de base radial (RBF), Sigmoide, d. muestra los puntos de datos que están en los márgenes y se conocen como vectores de soporte, e. muestra los datos que no se pueden separar con hiperplano, f. muestra los datos con aumento de dimensionalidad y separados con una superficie de decisión .

IV. METODOLOGIA Y RESULTADOS

Para la creación de las bases de datos se utilizan fotos de teléfonos inteligentes de diferentes marcas y modelos como Iphone XI , Iphone XII Promax, Iphone X, Iphone VIII, Iphone VI de MAC IOS, Samsung J7, Xiaomi, Huawei de usuarios de la App Salug, que junto con la foto enviaron información del teléfono, del termómetro y de su temperatura al momento de tomarse la foto para asociar la foto con la categoría de temperatura . La App Salug creada para la contención del Coronavirus que obtiene información de los síntomas asociados, y que forma parte de un proyecto de investigación previo donde se concluye que el teléfono inteligente podría constituirse en una herramienta muy útil para el monitoreo de la salud de una población . Desde la plataforma Instagram de la App Salug, se obtuvieron 200 imágenes recolectadas y guardadas en la nube en Google Drive, en un carpeta raíz y clasificándolas con la temperatura. para ser utilzadas con un programa en Python de Google Colab.

Figura 7. Obtención de Imágenes para Base de Datos

Las fotos enviadas con sus respectivas temperaturas se clasifican y almacenan en las respectivas carpetas de categorías de las temperaturas, en este caso nos interesan las categorías en el rango de 36.5º C a 37.5º C.

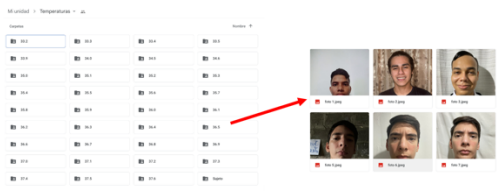

Figura 8. Obtención de Imágenes para Base de Datos

La CNN arroja el rostro de la imagen que se obtiene y la SVM que entrena con las categorías de las temperaturas de acuerdo con las imágenes previamente almacenadas .

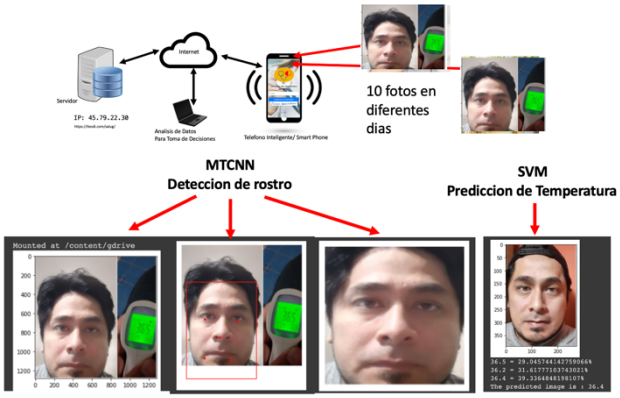

Figura 9. Predicción de la temperatura usando MTCNN y SVM

Se realizan 30 pruebas con 3 categorías de temperaturas para probar la eficacia y eficiencia de los algoritmos, obteniendo un 100% de eficacia en la detección del rostro que está a cargo de la CNN MTCNN y hasta el 54% de eficacia de la SVM en la predicción de la temperatura. Los resultados se pueden ver en la Tabla 1.

Categoria de Temperatura	Prueba 1	Prueba 2	Prueba 3	Prueba 4	Prueba 5	Prueba 6	Prueba 7	Prueba 8	Prueba 9	Prueba 10	Promedio de Acierto	
36.2	33,7	34,3	33,8	35,4	36,4	34,5	52	45,2	62,6	70,4	**43,84**	%
36.4	39,3	44,5	54,7	37,8	67,8	35,8	63	54,3	48,9	80,5	**52,62**	%
36.5	56,7	35,59	56,9	54,8	45,5	46,8	39	63,2	78,3	63,8	**54,02**	%

Tabla 1. Cuadro de resultados de predicciones exitosas

V. Conclusiones

Este estudio permite concluir que usando las imágenes faciales tomadas con un teléfono inteligente, elimina la necesidad de cualquier dispositivo especial como un termómetro infrarrojo o una cámara térmica que son equipos costosos y están instalados a la entreda de lugares de concurrencia masiva como lo son los centro comerciales , aeropuertos o instituciones educativas , siendo este el punto de control, después de haber recorrido un camino desde la residencia de la persona hasta el sitio de destino, incrementando el riesgo de contagiar a otras personas si este individuo tiene la enfermedad. Esta solución podría ser una implementación económica y conveniente para monitorear la temperatura, especialmente para enfermedades contagiosas relacionadas con la fiebre, como el COVID-19 de una población de estudiantes o trabajadores de la industria. Con un conjunto de datos a gran escala, se mejorará aún más la precisión y confiabilidad de la predicción de la temperatura usando un modelo CNN-SVM. Las imágenes usadas fueron tomadas de teléfonos inteligentes de diferentes marcas y modelos . Las imágenes faciales fueron extraidas con una CNN y el entrenamiento con un modelo SVM para predecir la temperatura. Los resultados experimentales podrian mejorar (actualmente las casos estan a la baja) con mas datos y son muy alentadores para un nuevo método para medir y monitorear la temperatura corporal.

References

[1] Hutchison, James S.; et al., “Hypothermia therapy after traumatic brain injury in children”. New England Journal of Medicine. 358 (23): 2447–2456 (2008).

[2] ImageNet, http://www.image-net.org.

[3] M. H. Yang, D. J. Kriegman, N. Ahuja, “Detecting faces in images: a survey,” IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (1) 34-58 (2002).

[4] P. Viola , M. Jones, “Robust real-time object detection,” International Journal of Computer Vision, (2001)

[5] Viola, P. and Jones, M., “Rapid Object detection using a boosted cascade of simple features.” Proceedings of CVPR, vol. 1 pp. 511–518, (2001).

[6] Viola, P. and Jones, M., “Robust Real-time Object Detection,” International Journal of Computer Vision, Vol. 57, Iss. 2, pp. 137–154, (2001).

[7] Papageorgiou, C., Oren, M., and Poggio, T., “A general framework for object detection,” Sixth International Conference on Computer Vision, pp. 555-562, (1998).

[8] Freund, Y. and Schapire, R., “A decision theoretic generalization of on-line learning and an application to boosting,” Computational Learning Theory: Eurocolt’95, pp 23-37, (1995).

[9] C. Burges, “A Tutorial on Support Vector Machines for Pattern Recognition,” Data Mining and Knowledge Discovery 2, 121- 167, 1998.

[10] Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome, “The Elements of Statistical Learning: Data Mining, Inference, and Prediction” (Second ed.). New York: Springer. p. 134, (2008).

[11] Duan, Kai-Bo; Keerthi, S. Sathiya, “Which Is the Best Multiclass SVM Method? An Empirical Study”, Multiple Classifier Systems. LNCS. 3541. pp. 278–285, (2005).

[12] Krizhevsky, A., Sutskever, I., Hinton, G.E., “Imagenet classification with deep convolutional neural networks,” Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1, Pages 1097-1105, Lake Tahoe, Nevada, (2012).

[13] Russakovsky, O., Deng, J., Su, H., et al., “ImageNet Large Scale Visual Recognition Challenge,” International Journal of Computer Vision (IJCV). Vol. 115, Issue 3, pp. 211–252 (2015).

[14] Simonyan, K., Zisserman, A., “Very Deep Convolutional Networks for Large-Scale Image Recognition,” arXiv technical report, (2014).

[15] University of Oxford, Visual Geometry Group, http://www.robots.ox.ac.uk/~vgg/research/very_deep/.

[16] He, K., Zhang, X., Ren, S., & Sun, J., “Deep Residual Learning for Image Recognition,” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770-778, (2016).