

ASAI 2015, 16° Simposio Argentino de Inteligencia Artificial.

Implementación de un Sistema de Análisis de Cortes Vacunos Basado en Redes Neuronales

Ulises Bussi¹, Giselle Mazzarello², Marta Ochoa², Fabián Sanfilippo¹, and
Damián Oliva¹

¹ Universidad Nacional de Quilmes, Carrera de Ingeniería en Automatización y Control Industrial, ulisesbussi@gmail.com

² Universidad Nacional de Quilmes, Carrera de Ingeniería en Alimentos Roque Sáenz Peña 352, Bernal, Buenos Aires, Argentina.

Abstract. Los sistemas de visión por ordenador permiten automatizar procesos de medición relacionados con la industria alimentaria, asistiendo a los operadores humanos en tareas supervisión comúnmente tediosas y realizando clasificaciones objetivas. En este trabajo describimos el desarrollo de un sistema de visión artificial para la clasificación de tejidos en cortes vacunos en condiciones de una línea de producción. El sistema utiliza redes neuronales multicapas que aprenden, a través del entrenamiento supervisado por un experto, a clasificar cada pixel de la imagen del corte vacuno en seis categorías de interés: fondo, bandeja, sombra, carne con coloración deseada, carne con coloración no deseada y grasa. Los resultados obtenidos muestran la factibilidad del sistema propuesto. Predice las clases de interés con un 2% de error para los cortes de Nalga y Bife de Chorizo. Además, el sistema puede cuantificar el deterioro en la coloración músculo en función del tiempo.

1 Introducción

Las técnicas visión por computador incluyen la adquisición, procesamiento y análisis de imágenes, facilitando la evaluación objetiva y no destructiva de las características visuales de calidad en productos alimenticios. Estos sistemas, están siendo utilizados con mayor frecuencia en la industria alimentaria para fines de control de calidad, ya que ofrecen la posibilidad de automatizar las prácticas manuales de clasificación eliminando las tareas de supervisión humana tediosas. Los recientes avances en hardware y el software han ayudado en esta expansión, proporcionando soluciones potentes en rentabilidad, consistencia, velocidad y exactitud[1].

En este trabajo desarrollamos un sistema de visión automático que adquiere imágenes de cortes vacunos depositados sobre una bandeja y mide parámetros de interés para el consumidor, como ser la cantidad de grasa y el estado de la coloración del músculo.

Existe una bibliografía extensa relacionada con la aplicación de técnicas de visión artificial para el análisis de cortes cárnicos (por ejemplo cortes vacunos [3], de cerdo [4], de cordero [5], etc). Una revisión detallada sobre esta temática puede

encontrarse en [6]. Es notable que la mayoría de estos estudios se realiza en condiciones de laboratorio donde el corte se deposita sobre una superficie que genera un buen contraste respecto al fondo, simplificando la clasificación de la grasa. Una vez eliminados los píxeles asociados a la grasa, se predicen propiedades complejas del músculo como ser la ternura, la jugosidad ó la aceptación. Utilizando modelos lineales se alcanzan muy buenas correlaciones entre las propiedades medidas en la imagen y las variables de interés [7].

En este trabajo, estudiamos tres aspectos que consideramos importantes en relación a la industrialización de estos procesos en una cadena de producción que no fueron estudiados previamente: 1) El diseño de un clasificador supervisado que distinga correctamente las clases de interés. Además que aprenda a separar la bandeja (que generalmente tiene una coloración similar a la grasa) de la grasa en sí. De este modo no será necesario el ajuste por un experto, de los pasos de preprocesamiento previos a la clasificación de los tejidos de interés alimentario; 2) la implementación de un sistema de iluminación que disminuya el efecto de las sombras introducidas por la curvatura de la bandeja y los reflejos en la superficie del corte 3) una medición del aumento de la coloración no deseada en el músculo durante un lapso de cinco días.

Es importante destacar que la información visual es una de las vías más relevantes en la decisión del consumidor de adquirir el producto. Un cambio en el color de los tejidos puede relacionarse con una mala conservación y el subsecuente desarrollo microorganismos.

Otro aspecto muy importante en la coloración es el contacto del oxígeno, que en tiempo prolongado le imprime un color marrón a la carne. Los consumidores ante la detección en el cambio de coloración, prefieren no comprar el producto. Es necesario destacar que el cambio en la coloración de la carne en función del tiempo no fue caracterizado en trabajos previos [7].

Estas razones nos llevaron a desarrollar un sistema para analizar automáticamente imágenes de cortes vacunos. Para esto, un entrenador humano (que denominaremos supervisor o experto) clasificará algunos píxeles de la imagen en las siguientes categorías: {1-fondo, 2-bandeja, 3-sombra, 4-carne con coloración deseada, 5-carne con coloración no deseada, 6-grasa}. Nuestro objetivo es hallar una red neuronal que aprenda el patrón incorporado por el operador y pueda generalizar el comportamiento del experto a nuevas imágenes. De éste modo, la empresa podrá utilizar ésta información para tomar una decisión acerca del funcionamiento de su línea de producción y del proveedor (penalización y/o devolución de la mercadería dañada).

Este trabajo se organiza de la siguiente forma: En la Sección 2 introducimos el dispositivo utilizado y describimos los clasificadores con redes neuronales que utilizaremos. En la Sección 3 caracterizamos el error de clasificación para las distintas redes testeadas y mostramos su utilidad para medir la cantidad de grasa y el deterioro del músculo a través de los días. Finalmente en la Sección 4 discutimos posibles mejoras a introducir en el futuro.

2 Metodología Experimental

2.1 Dispositivo Experimental

La iluminación de una escena es un factor importante que afecta a menudo a la complejidad de los algoritmos de visión. En este trabajo, se utilizó el enfoque de iluminación difuso (ver Fig. 1) que se puede emplear en los objetos caracterizados por superficies lisas y regulares (disminuyendo reflejos especulares y sombras).

El sistema de adquisición de imágenes utiliza una cámara digital Fujifilm modelo FinePix S4000 con longitud focal y la exposición fijas. Las imágenes fueron registradas en 24 bits de color. Resolución de la imagen fue 574x720 píxeles. El área registrada en cada imagen fue 20.3cm de ancho y 16.1cm de largo. De este modo, un centímetro corresponde aproximadamente a 35 pixels.

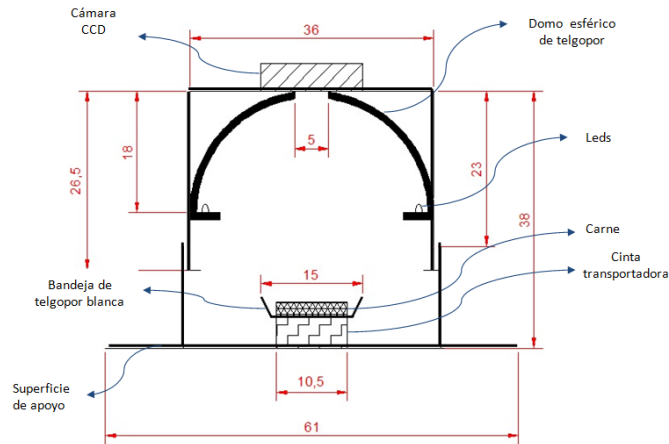


Fig. 1. Dispositivo de iluminación y adquisición de imágenes. Las dimensiones son en centímetros.

2.2 Clasificación de Imágenes de Cortes Vacunos Utilizando Redes Neuronales Multicapas

El objetivo de este trabajo es construir un clasificador que clasifique los píxeles de una imagen del corte de carne en las seis categorías (ver Fig. 2). Luego de entrenar el clasificador con la información ingresada por un experto, este debe emular al experto.

Utilizaremos redes neuronales artificiales (RNAs) multicapas. En la Fig. 3 mostramos una red de dos capas y el cómputo que realiza. Como se ve en la figura, la entrada de la RNA está formada por un vector p de dimensión $R \times 1$. La capa i de la RNA tiene S^i neuronas y su estado a^i está descrito por el vector: $a^i = f^i(W^i * a^{i-1} + b^i)$, donde W^i es la matriz de pesos sinápticos (de dimensión $S^i \times S^{i-1}$), b^i es el vector de desvíos y f^i es la función de activación.

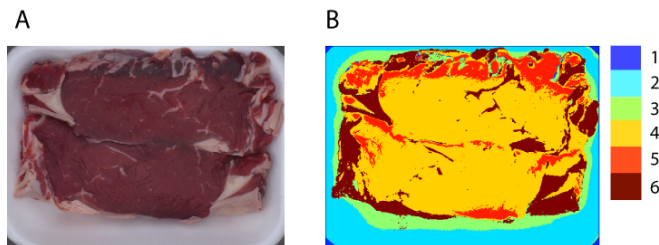


Fig. 2. Clasificación de la imagen de un corte vacuno. A) Imagen original. B) Imagen clasificada por la red neuronal. Las clases son: {1-fondo, 2-bandeja, 3-sombra, 4-carne con coloración deseada, 5-carne con coloración no deseada, 6-grasa}.

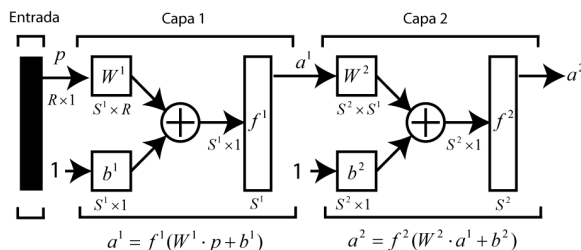


Fig. 3. Redes neuronales multicapas [2: Matlab].

Si tenemos que resolver problemas de clasificación que no sean linealmente separables debemos utilizar RNAs con más de una capa.

Un ejemplo de vector de entrada puede ser: $p = [R_n, G_n, B_n, x_n, y_n]^T$, donde R_n, G_n, B_n representan los valores (normalizados a 1) de los tres canales de color del pixel 'n' y x_n, y_n las coordenadas (x,y) en la imagen.

Como dijimos, cada pixel será clasificado en una de las seis categorías. El aprendizaje de la asociación de la clase correcta (dada por la salida de la red 'a') para cada entrada p se hace utilizando la información ingresada por el supervisor. El supervisor selecciona un pixel de la imagen e indica la clase correcta (que notaremos con el vector t). La salida deseada t de la red neuronal es un vector columna de (6×1) que vale: $[1 \ 0 \ 0 \ 0 \ 0 \ 0]^T$ si es clase 1; $[0 \ 1 \ 0 \ 0 \ 0 \ 0]^T$ si es clase 2, y así sucesivamente a la clase 6.

Ésta información de entrenamiento es guardada en una base de datos para producir los entrenamientos y testeos de RNAs con distintas configuraciones de entrada y de geometrías de red (ver Sección 3).

El entrenamiento de la RNA consiste en encontrar los parámetros $\{W, b\}$ de cada red propuesta para lograr que el error cuadrático medio entre la salida de la red a y el valor correcto t ingresado por el supervisor sea mínimo. Se define el error cuadrático medio de la predicción de la red como:

$$E(W, b) = \frac{1}{N_{train}} \sum_{n=1}^{N_{train}} (t_n - a_n)^T \cdot (t_n - a_n).$$

Los valores iniciales de $\{W, b\}$ se eligen al azar y se utiliza la técnica de Levenberg-Marquardt para minimizar E ([2]). Los algoritmos se implementaron en el toolbox de Redes Neuronales de Matlab (que es un software privativo) y los casos óptimos se implementaron

en Python (que es un software libre). En los dos casos sus resultados fueron similares.

2.3 Esquemas de Entrenamiento-Testeo

Se analizaron muestras de cortes de ‘Bife de Chorizo’ ($N_{BF} = 6$) y ‘Nalga’ ($N_N = 5$), donde N_{BF} y N_N es el número de muestras de cada corte. El conjunto de datos consta de $N_{cortes} = N_N + N_{BF} = 11$. Se tomaron imágenes de cada muestra durante 5 días (por lo tanto la base de datos tiene 55 imágenes). Por cada corte y cada día (o sea, por cada imagen) el entrenador ingresó 50 ejemplos de cada clase (esto resulta en 300 ejemplos por imagen). El conjunto total de datos ingresados por el supervisor $\{p, t\}$ (recordar que p es el vector de entradas y t el de salidas deseadas) se dividió en dos conjuntos. El conjunto $\{p_{train}, t_{train}\}$ formado por $(N_{cortes} - 1)$ y $\{p_{test}, t_{test}\}$ formado por el corte restante. Además, el toolbox de RNA divide el conjunto $\{p_{train}, t_{train}\}$ en 2 grupos, de 70% para entrenamiento y 30% para validación. Con este esquema, el toolbox frena el entrenamiento cuando el error con los ejemplos de validación aumenta sostenidamente por 6 épocas. Una vez encontrada la mejor red neuronal que predice la clase correcta para $\{p_{train}, t_{train}\}$, se usó $\{p_{test}, t_{test}\}$ para testear la predicción de la red con datos no usados previamente.

3 Resultados

3.1 Error de Clasificación para Distintas Geometrías de la RNA

Utilizamos el esquema de entrenamiento-testeo de la sección anterior para estudiar el error de clasificación de distintas geometrías de redes multicapas. Se testearon redes de una y dos capas y se varió el número de neuronas S^1 y S^2 en $\{5, 10, 20\}$ neuronas. Se utilizaron funciones de transferencia del tipo *logsig*.

Otro factor importante a la hora de definir la red es proponer cuales son las características de la imagen que se utilizarán como entrada de la RNA. Propusimos tres casos: La configuración ‘1’ tiene un vector $p = [R_n, G_n, B_n]^T$ donde R_n, G_n, B_n representan los valores (normalizados a 1) del canal rojo, verde y azul del pixel ‘ n ’. En la configuración ‘2’ tenemos cinco entradas $p = [R_n, G_n, B_n, x_n, y_n]^T$ donde x_n, y_n corresponden a las coordenadas (x,y) del pixel ‘ n ’. Esta propuesta se hizo debido a que las clases ‘fondo’, ‘bandeja’ y ‘sombra’ generalmente se encuentran en los límites externos de la imagen. Finalmente en la configuración ‘3’ se agregaron los valores del canal rojo de los primeros vecinos del pixel ‘ n ’ (por lo tanto, en este caso $p \in \mathfrak{R}^{(9 \times 1)}$).

En la Tabla 1 mostramos el error de clasificación utilizando los datos de testeo para todas las configuraciones analizadas en el caso de los dos tipos de cortes.

Podemos ver que a medida que aumentamos la complejidad de la red, el error de testeo para la clasificación disminuye. Esto quiere decir que el control de overfitting utilizado (ver sección anterior) resulta exitoso. Por otro lado, vemos que el error de testeo para el corte 1 es menor que para el corte 2. De

Tabla 1. Resultados del entrenamiento de distintas geometrías de RNA con las muestras de Nalga y de Bife de chorizo. Las columnas 5 y 6 informan el porcentaje de predicciones erróneas.

Entradas	N° de capas	N° de neuronas en la primer capa	N° de neuronas en la segunda capa	Error medio nalga (%)	Error medio bife de chorizo (%)	
3	1	5	-	2.7	6.5	
		10	-	2.1	5.7	
		20	-	1.9	5.4	
	2	5	5	5	2.0	6.1
			10	10	2.0	6.2
			20	20	2.0	5.9
		10	5	5	2.0	5.8
			10	10	2.0	5.8
			20	20	2.0	5.9
		20	5	5	1.9	5.7
			10	10	2.0	6.0
			20	20	2.0	6.0
5	1	5	-	8.3	15.6	
		10	-	2.4	5.3	
		20	-	1.9	4.8	
	2	5	5	5	2.0	5.4
			10	10	2.0	4.9
			20	20	1.9	4.9
		10	5	5	2.0	4.9
			10	10	2.0	4.6
			20	20	1.5	4.5
		20	5	5	2.3	5.4
			10	10	1.9	4.4
			20	20	1.9	4.9
9	1	5	-	5.2	10.7	
		10	-	2.6	5.7	
		20	-	1.9	4.8	
	2	5	5	5	2.0	5.1
			10	10	2.0	5.0
			20	20	1.8	5.0
		10	5	5	1.8	4.4
			10	10	1.6	4.8
			20	20	1.5	4.3
		20	5	5	1.5	4.6
			10	10	1.6	4.2
			20	20	1.4*	3.9*

este modo, concluimos que (si la línea de producción lo permite) resulta apropiado utilizar redes optimizadas para cada tipo de corte. Sin embargo, también testeamos la clasificación de todos los cortes con una sola red neuronal con la geometría número 20×20 . En este caso obtuvimos un error de clasificación de 2% y esto es razonable ya que es un valor promedio entre los dos cortes analizados.

Para analizar que clases producen mayor error de clasificación, calculamos la matriz de confusión en el testeo para la RNA 20×20 (ver * Tabla 2). En ambos casos la mayor confusión se produce entre las clases 'grasa' y 'sombra'.

Otro aspecto importante es que, dado que nuestro principal objetivo es medir parámetros de interés para los consumidores (tales como porcentaje de grasa y coloración del músculo) se probó agrupar las tres clases que no son de interés para el consumidor (bandeja, fondo y sombra) en una sola para simplificar el clasificador y, de este modo, el entrenamiento necesario. Es notable que la RNA 20×20 puede resolver el problema aumentando a 2.6% de error promedio. Esto se debe a un aumento en las confusiones se produce nuevamente con la clase grasa. Finalmente se compararon los resultados obtenidos por la RNA 20×20

con otros dos tipos de clasificadores: un clasificador basado en el análisis discriminante (CAD) y otro basado en la técnica k vecinos cercanos (KNN). En los tres clasificadores se utilizó el caso de 9 entradas y 6 salidas. El error de testeo (para ambos cortes) con el CAD fue del 8.7%. La técnica basada en KNN se testeó con valores de k entre 1 y 15 obteniéndose un error del 2.9%. Por lo tanto ambos clasificadores tuvieron un desempeño inferior al basado en RNAs.

Tabla 2. Matriz de confusión correspondiente al corte Nalga (izquierda) y Bife de Chorizo (derecha) para la RNA20 × 20 con 9 entradas.

clase predicha	target						clase predicha	target					
	1	2	3	4	5	6		1	2	3	4	5	6
1	0.995	0	0	0	0	0	1	1	0	0	0.001	0.001	0
2	0	0.998	0.004	0	0	0	2	0	0.99	0.022	0	0	0.001
3	0	0.002	0.954	0.001	0.002	0.02	3	0	0.01	0.951	0.002	0	0.035
4	0	0	0.001	0.976	0.007	0	4	0	0	0.003	0.953	0.064	0.003
5	0.005	0	0.009	0.019	0.989	0	5	0	0	0	0.036	0.934	0
6	0	0	0.032	0.004	0.002	0.98	6	0	0	0.024	0.008	0.001	0.961

3.2 Proporción de Clases y Deterioro del Corte en Función del Tiempo

Utilizamos el esquema de entrenamiento-testeo de la Sección 2 para estudiar la evolución temporal de la clasificación durante un periodo de 5 días. En este estudio se entrenó la RNA 20 × 20 con $(N_{cortes} - 1)$ de muestra y se separó una muestra para testeo (tener en cuenta que cada muestra tiene asociada 5 imágenes). En forma similar a la sección anterior, una vez entrenada la red, se realizaron las predicciones para todas las imágenes de los datos de test. En la Fig. 4 (A) mostramos un ejemplo de clasificación en función del tiempo para un corte de testeo. Como se ve en la RNA detecta un aumento sostenido de la clase ‘músculo con coloración no deseada’ en detrimento de la clase ‘músculo con coloración deseada’. Para cuantificar mejor el efecto mostrado en la Fig. 4 ,

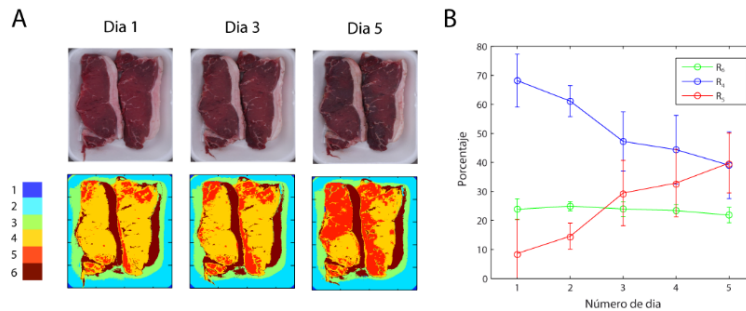


Fig. 4. Clasificación para un corte de testeo y medición del deterioro de la calidad de la carne en función de los días (A). Medición de parámetros de interés del corte (R_4 , R_5 y R_6) en función de los días. Los círculos corresponden al valor medio, y las barras al error estándar (B).

definimos A_T como el área total de la muestra $A_T = A_4 + A_5 + A_6$, donde A_i ,

corresponde al área de la clase i . Definimos los porcentajes de interés siguientes para cada clase como: $R_i = A_i/A_T * 100$.

En la Fig. 4 (B) mostramos la evolución temporal de R_4 , R_5 y R_6 promediado para todos los cortes de testeo. En este caso se observa nuevamente la tendencia que mostramos en la Fig. 4.

4 Discusión y Conclusiones

Este trabajo describe la implementación de un clasificador supervisado basado en redes neuronales multicapas para la clasificación y medición de parámetros de interés en cortes vacunos. Los resultados obtenidos muestran la factibilidad del sistema propuesto ya que con una RNA 20×20 (de dos capas con 20 neuronas por capa) permite predecir las clases de interés con un 1.4% de error para los cortes de Nalga y un 3.9% de error para los cortes de Bife de Chorizo. Además, nuestro estudio muestra que el sistema puede cuantificar el deterioro del músculo en función del tiempo.

En el futuro, existen muchas mejoras que se pueden implementar en este sistema. La primera consiste en una automatización del alineado de los cortes para la toma de imágenes. También el testeo del método a otros cortes cárnicos y a otras clases de interés como ser tendones o huesos.

Respecto a la iluminación, se debe mejorar la uniformidad y difusividad para disminuir las sombras en la imagen y así reducir las confusiones entre la clase ‘sombra’ y ‘grasa’. También testear la posibilidad de extender el espectro a otras bandas como ser la ultravioleta o infrarroja que pueden implementarse con leds.

Finalmente debe estudiarse cómo generar un método de toma de decisión en función de los estándares de calidad del productor.

References

1. Gunasekaran, S.: Computer vision technology for food quality assurance. Trends in Food Science and Technology, 7(8), 245–256.(1996)
2. Demuth, H., & Beale, M.: Neural network toolbox for use with MATLAB. (1993)
3. Jackman, P., Sun, D.-W. & Allen, P.: Automatic segmentation of beef longissimus dorsi muscle and marbling by an adaptable algorithm. Meat Science , 83(2) , 187–94.(2009)
4. Jackman, P., Sun, D.-W., Allen, P., Valous, N. A., Mendoza, F. & Ward, P.: Identification of important image features for pork and turkey ham classification using colour and wavelet texture features and genetic selection. Meat Science , 84(4) , 711–17. (2010)
5. Chandraratne, M. R., Kulasiri, D. & Samarasinghe, S.: Classification of lamb carcass using machine vision: Comparison of statistical and neural network analyses. Journal of Food Engineering , 82(1) , 26–34. (2007)
6. Sun, D. W.: Computer vision technology in the food and beverage industries. Elsevier.(2012)
7. Jackman, P., Sun, D.-W., Allen, P., Brandon, K. & White, A. M. Correlation of consumer assessment of longissimus dorsi beef palatability with image colour, marbling and surface texture features. Meat Science , 84(3) , 564–8. (2010)