

Clasificación de hojas de té al ingreso del proceso de secado mediante redes neuronales con datos supervisados y no supervisados

Luisina A. de Paula¹, Karina B. Eckert¹, Gabriel Guismín¹

¹Departamento de Ingeniería y Ciencias de la Producción - Universidad Gastón Dachary.
Posadas, Misiones, Argentina.

luisinaadp@hotmail.com, {karinaeck, guisming}@gmail.com

Resumen. El proceso de clasificación de la materia prima, es uno de los procedimientos de mayor relevancia en cualquier secadero de té, siendo el responsable de asegurar una buena calidad del producto final. Actualmente, éste proceso está a cargo de una persona experta, que realiza el trabajo de forma manual y bajo su propio criterio, lo cual tiene una serie de inconvenientes asociados. En este trabajo, se propone una solución que incluye el planteo, diseño, desarrollo y prueba de un prototipo que sea capaz de clasificar correctamente fotografías correspondientes a muestras de materia prima arribada a un secadero, utilizando técnicas de Inteligencia Artificial (IA) de tipo supervisadas para la Clasificación mediante Redes Neuronales Artificiales y no supervisadas con Agrupamiento K-means para la preparación de las clases. El prototipo tuvo un correcto desempeño y es una herramienta confiable para clasificar la materia prima arribada a secaderos de té.

Palabras Clave: Clasificación de hojas de Té, Reconocimiento de Imágenes, Clasificación, Agrupamiento, Redes Neuronales Artificiales, K-means.

1 Introducción

En Argentina, la región tealera o productora de té se centra en las provincias de Misiones y Corrientes, con un total aproximado de 45.000 hectáreas, siendo uno de los productos con mayor nivel de producción del país. En la provincia de Misiones se cultiva el 95 % de las 45.000 hectáreas del país, produciendo un valor de exportación anual aproximado de 57 millones de dólares [1].

Desde el año 2006, con el aumento de la informatización y automatización de los procesos, la industria del té se ha visto afectada por diferentes problemáticas por falta de actualización en sus sistemas de clasificación de calidad, selección de materia prima, etc. en los distintos secaderos [1][2].

Según un informe realizado en el año 2012 en conjunto por el Ministerio de Ciencia y Tecnología de la Nación, la Unión Industrial Argentina, el programa para la Federalización de la Ciencia, la tecnología y la innovación productiva y otros organismos nacionales, conocido como “TÉ: Debilidades cuya superación implica un desafío científico tecnológico”, una de las principales problemáticas es la clasificación de materia prima que ingresa a los secaderos, ya que esta proviene de diferentes proveedores, que

pueden tener productos tanto certificados como no y al mismo tiempo variar sus calidades [1].

Las empresas y cooperativas que realizan la segunda transformación, es decir, aquellas que se encargan del secado de las hojas y elaboración posterior de té, suelen abastecerse con sólo un porcentaje de materia prima propia (cosechas/plantaciones particulares), comprando el resto a varios productores locales. La diversidad de proveedores hace que sea muy difícil poder clasificar el tipo de hoja de té que llega al secadero sin la presencia de un especialista [1].

El proceso de clasificación de la materia prima que arriba a los secaderos se encarga de determinar los porcentajes de malezas existentes en los cargamentos, separando éstas de lo que puede ser considerado como materia prima procesable, la cual al mismo tiempo es clasificada en 3 tipos conocidos como brotes verdes, tallos verdes y hojas o tallos no verdes respectivamente. Actualmente, ambos procesos se realizan a mano, a partir de una pequeña muestra tomada de cada cargamento que arriba a la planta elaboradora. Para este proceso, se necesita de un experto que pueda detectar de forma visual las diferencias entre cada tipo correspondiente a la clasificación, ya que de las decisiones clasificatorias de éste dependerá la futura calidad productiva [1] y la obtención de materia prima correcta para poder realizar adecuadamente los siguientes pasos del proceso productivo. Cabe destacar que en caso de ausencia (momentánea o indeterminada) de esta persona especializada, sería difícil reemplazarla de forma rápida y clasificar de la misma manera, manteniendo así la calidad uniforme del producto [3][4].

Debido a esta situación, surge la necesidad de una solución informática que permita una clasificación uniforme, rápida y efectiva de la materia prima arribada a los secaderos de té. La solución a este problema implica, en primer lugar, la digitalización de imágenes y luego la aplicación de técnicas de procesamiento de datos para realizar dicha clasificación.

Para llevar a cabo esto, las imágenes pueden ser representadas de forma numérica mediante la utilización de sus propiedades RGB (*Red-Green-Blue*) de cada uno de sus píxeles y de diferentes descriptores matemáticos que permitan especificar los mismos. A partir de la representación numérica, las técnicas de IA, en este escenario planteado, son una opción adecuada. Esto se debe a que las mismas, tienen la capacidad de realizar diferentes tipos de clasificaciones, sobre todo en lo que respecta al reconocimiento de imágenes, como la llevada a cabo en esta investigación.

Para comprobar la aplicabilidad de este proceso se requiere formalizar el mismo y llevar a cabo una serie de pruebas concluyentes. Para dar solución a la problemática planteada, se ha optado por las Redes Neuronales Artificiales, como técnica de inteligencia artificial para llevar a cabo las clasificaciones pertinentes y la técnica de agrupamiento (*clustering*) K-means, para la obtención de datos de manera no supervisada, logrando de esta manera, medir la aptitud de la técnica clasificatoria.

2 Proceso general de los secaderos de té en la provincia de Misiones

Basándose en la información brindada por el INTA (Instituto Nacional de Tecnología Agropecuaria) [2] y la empresa tealera Don Basilio S.R.L, se puede decir que el

proceso general del té consta de siete etapas bien diferenciadas que serán descritas a continuación y que pueden verse reflejadas gráficamente en la Fig. 1:

- **Arribo de Camiones:** El arribo de camiones se hace de manera diaria (días hábiles) en los secaderos de té, teniendo en cuenta las chacras que se encuentran en mayor cercanía de un determinado secadero o dependiendo de algún pedido especial por parte de alguna empresa tealera hacia alguna chacra que se encuentre certificada por la calidad de sus productos. Los camiones por lo general no poseen horarios o turnos, sino que llegan luego de haber recolectado la materia prima de la plantación correspondiente y sin previo aviso al aserradero en cuestión, lo cual en muchos casos genera retrasos o cuellos de botella que pueden ralentizar el proceso de producción, es por esto que este proceso se determina como muy problemático.
- **Clasificación de Materia Prima:** El proceso de clasificación de materia prima (especificado en mayor detalle en la Sección 2.1), consiste en determinar el tipo de hojas y tallos recibidos en cada cargamento arribado al secadero en cuestión. Este proceso es clave para luego poder llevar a cabo de forma correcta la posterior producción, ya que, en caso contrario, pueden haber pérdidas en las líneas de producción, de dinero importante, de tiempo, de prestigio de la empresa, de calidad del producto final, entre otras problemáticas [5].
Este proceso se lleva a cabo de forma manual y por una persona capacitada y especializada que clasifica en base a sus criterios, experiencia y una documentación guía brindada por el Ministerio del Agro y la Producción de la provincia de Misiones.
La principal problemática reside en que si el operador en cuestión debe ausentarse, quien lo reemplace no llevará a cabo la tarea de igual manera, o en peores situaciones, no existirá personal que pueda reemplazarlo (al menos por un tiempo), lo cual puede llevar a parar la producción por tiempo indeterminado, además de que el procesamiento manual es lento y conlleva en muchos casos a largas esperas o paradas en la línea de producción.
- **Distribución en contenedores:** La materia prima arribada es ubicada en diferentes contenedores para su posterior procesamiento. Este proceso es de gran cantidad de problemas debido a que no existe una forma de control o distribución adecuada que determina qué contenedores se encuentran vacíos o cuáles se encuentran llenos. Al mismo tiempo, la materia prima certificada o de mayor calidad, puede mezclarse con otro tipo de la misma que no se encuentre en las mismas condiciones en caso de que accidentalmente dos cargamentos ingresen al mismo contenedor.
- **Procesamiento de materia prima:** El proceso de materia prima consiste básicamente en 5 procesos: Marchitado, Enrolado, Fermentado, Secado y Clasificación del producto final. Es una etapa que se lleva de forma normal y sin problemas, siempre y cuando no se produzcan condiciones adversas en procesos anteriores (como ser deficiente o no uniforme clasificación de materia prima, contenedores mezclados, etc.).
- **Tipificado:** Esta operación tiene como finalidad separar el té elaborado en fracciones de distinto tamaño de partícula. La tarea se lleva a cabo en las máquinas clasificadoras, las que consisten en un juego de zarandas animadas por un movimiento circular o de vaivén y cuyas mallas se diferencian entre sí por el diámetro de las perforaciones.

Las fracciones de distintos tamaños reciben el nombre de grados o tipos. Los grados primarios están elaborados con yemas, parte tierna del pecíolo de la hoja y tallos de los brotes de té y son directamente extraídos del proceso de elaboración. No obstante, siempre queda una fracción muy gruesa que no alcanza a atravesar las aberturas de la malla. Esta porción se somete a un proceso de quebrado y tipificación posterior, reciben el nombre de grados secundarios y son de calidad inferior.

Los palos y fibras se muelen y tipifican en forma separada, obteniéndose al final de este proceso los tipos de té llamados BT [5].

- **Envasado:** Este proceso consiste en colocar en envases el té ya elaborado y tipificado para su posterior distribución y comercialización.
- **Distribución:** Este proceso consiste en distribuir a diferentes compradores e interesados, el producto ya elaborado y envasado.
Esta etapa de la elaboración del té tiene diferentes complicaciones, entre ellas, una de las más preocupantes es el hecho de que no se establecen rutas eficientes o más cortas para el reparto de los productos; cada camión realiza el recorrido que mejor les parezca a sus conductores o que les sea de mayor comodidad para ellos, sin haber un plan de ruta establecido con anticipación. Esto, lleva a pérdidas de tiempo no necesarias, gasto excesivo de combustible, pérdidas monetarias, etc.

Proceso General – Secadero de Té

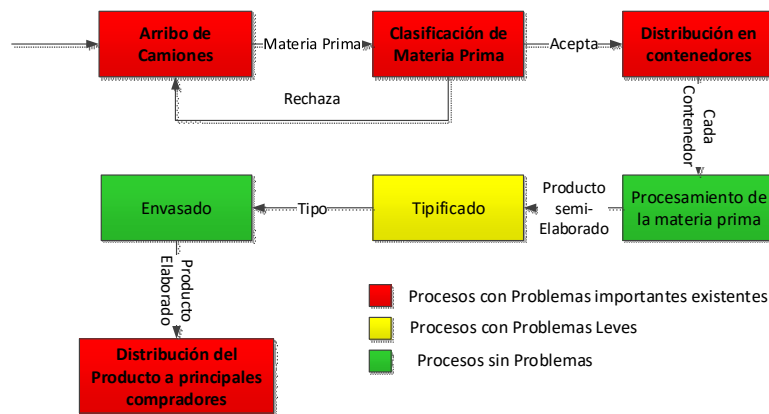


Fig. 1. Proceso general de un secadero de té

2.1 Proceso de clasificación (actual) de la materia prima arribada a los secaderos

Actualmente, varios secaderos de la provincia de Misiones, llevan a cabo el proceso de clasificación de materia prima que se encuentra especificado en el esquema de la Fig. 2. Como se puede observar, luego de la recepción de la última mencionada, se establece el porcentaje de maleza con el que cuenta el cargamento arribado y luego de esto se clasifica a la materia prima en tres posibles tipos (según su contenido y porcentaje), que son brotes, tallos verdes y tallos y hojas no verdes [3][4].

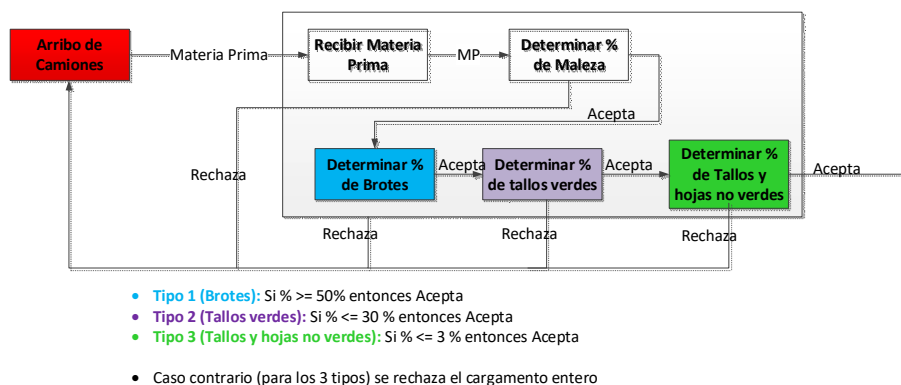


Fig. 2. Proceso actual de clasificación de materia prima

- **Brotos:** Este tipo de materia prima es el que brinda mayor calidad de producto y el que es principalmente utilizado en el proceso de elaboración. La materia prima arribada debe contener más de un 50 % de brotes para ser considerada como de ese tipo. Los brotes son la parte superior de la planta de té, se caracterizan por un color verde claro, pero muy intenso y brillante. Se puede observar claramente su textura y color en la Fig. 3 [3][4].



Fig. 3. Brotos, Tallos Verdes y Hojas No Verdes

- **Tallos Verdes:** Un cargamento que clasifica como tallos verdes y que posee una cantidad menor a un 30 % es aceptada para la elaboración posterior del té. Es de gran importancia determinar si un cargamento pertenece a brote o tallos verdes ya que de esto dependerá el futuro producto elaborado y la calidad del mismo para su distribución. Los tallos verdes, como se puede observar en la Fig. 3, son identificables por ser de un color verde “amarronado” y tener un grosor fino y una longitud similar a la de los brotes [3][4].
- **Tallos y hojas no verdes:** Un cargamento que clasifica como tallos y hojas no verdes es aquel que cuenta con cantidades superiores al 3% de este tipo de materia prima. Los tallos y hojas no verdes son identificables por sus colores marrones, rojizos o verdes muy oscuros (como se puede apreciar en la Fig. 3); los cuales difieren del verde que se puede ver en los brotes o el leve grado de marrón que puede observarse en los tallos verdes. Cualquier cargamento arribado que clasifique como tallos y hojas no verdes, es directamente rechazado, ya que la materia prima arribada no servirá para

poder obtener un producto de buena calidad luego del proceso de elaboración. (En algunos escasos casos, el cargamento se recibe de igual forma a mitad de precio y la materia prima es aprovechada para otro tipo de productos) [3][4].

3 Agrupamiento

El agrupamiento o *clustering* es el proceso de organizar instancias de datos en grupos cuyos miembros sean similares en cierta forma. Un clúster es una colección de instancias de datos que son similares entre sí y no similares a instancias de datos en otros *clusters* (grupos). La tarea del *clustering* es la de encontrar los grupos escondidos entre los datos. Para los humanos, no siempre es sencillo visualizar o detectar un clúster o grupos de clasificación, por lo que el *clustering* es una muy buena herramienta para ayudar en esta tarea. [6]

El clustering es considerado como un tipo de aprendizaje no supervisado; esto se debe a que, a diferencia del aprendizaje supervisado, los valores de clase no denotan una partición o agrupamiento a priori de datos, sino que las instancias o clases se van descubriendo a lo largo de la tarea de partición [7].

Uno de los algoritmos más populares de agrupamiento es K-means, el cual logra dada una cierta cantidad de puntos y la cantidad k de clusters (especificados por el usuario), iterativamente particionar la información en k clusters basados en una función de distancia. Consiste en seleccionar al azar k puntos como centroides semilla que luego se utilizarán para realizar los cálculos de distancia entre cada centroide y cada punto existente en la serie de datos. Cada punto es asignado al centroide que está más cerca de sí. Una vez que todos los puntos son asignados, el centroide de cada clúster se recalcula. El proceso se repite hasta que se conoce o se lleva a cabo un criterio de parada [7].

4 Redes Neuronales Artificiales

“Las redes neuronales son conjuntos de elementos de cálculo simples, usualmente adaptativos, interconectados masivamente en paralelo y con una organización jerárquica que le permite interactuar con algún sistema del mismo modo que lo hace el sistema nervioso biológico” [8].

Debido a su constitución y a sus fundamentos, las RNA presentan un gran número de características semejantes a las del cerebro. Por ejemplo, son capaces de aprender de la experiencia, de generalizar de casos anteriores a nuevos casos, de abstraer características esenciales a partir de entradas que representan información irrelevante, etc. Esto hace que ofrezcan numerosas ventajas y que este tipo de tecnología se esté aplicando en múltiples áreas [9]. Algunas ventajas y características principales de redes neuronales son: [9]

- **Aprendizaje adaptativo:** Capacidad de aprender a realizar tareas basadas en un entrenamiento o en una experiencia inicial.
- **Auto-organización:** Una red neuronal puede crear su propia organización o representación de la información que recibe mediante una etapa de aprendizaje.

- **Tolerancia a fallos:** La destrucción parcial de una red conduce a una degradación de su estructura; sin embargo, algunas capacidades de la red se pueden retener, incluso sufriendo un gran daño.
- **Operación en tiempo real:** Los cálculos neuronales pueden ser realizados en paralelo; para esto se diseñan y fabrican máquinas con hardware especial para obtener esta capacidad.
- **Fácil inserción dentro de la tecnología existente:** Se pueden obtener chips especializados para redes neuronales que mejoran su capacidad en ciertas tareas. Ello facilita la integración modular en los sistemas existentes.

4.1 Redes Neuronales BPN (Back Propagation Net)

Las Redes Neuronales con Propagación hacia atrás (BPN) utilizan un algoritmo de aprendizaje supervisado. Dado un patrón a la entrada de la red como estímulo, se propaga hacia adelante desde la primera capa a través de las capas siguientes de la red, generando una salida, ésta se compara con una salida deseada (que fue estipulada) y se calcula una señal de error para cada una de las mismas obtenidas. Las salidas de error se propagan hacia atrás, partiendo de la capa de salida, hacia todas las neuronas de la capa oculta que contribuyen directamente a ésta. Cabe destacar que las neuronas de capa oculta, solo reciben una fracción de la señal total del error, basándose aproximadamente en la contribución relativa que haya aportado cada neurona a la salida original. Este proceso se repite, capa por capa, hasta que todas las neuronas de la red hayan recibido una señal de error que describa su contribución relativa al error total. [10]

La importancia de las BPN es su capacidad de aprender la relación entre un conjunto de patrones y sus salidas correspondientes para poder aplicar esa misma relación después del entrenamiento a patrones con ruido donde una salida se activa si la nueva entrada es parecida a la presentada en el aprendizaje, lo cual hace a las BPN sistemas capaces de generalizar. [9]

5 Prototipo de clasificación automática de hojas de té, mediante técnicas de inteligencia artificial

El prototipo desarrollado fue implementado en el lenguaje de programación Java, con la característica de tener una interfaz gráfica de usuario y codificación mayoritariamente propia, como así también, la reutilización de pequeñas porciones de código pertenecientes a diferentes autores [11] y librerías [12].

5.1 Obtención de las imágenes

La toma de las muestras fotográficas se realizó en dos ocasiones durante los meses de noviembre a marzo de la temporada de zafra 2015-2016 del té; las mismas se tomaron en los asentamientos de la localidad de Colonia Alberdi de la empresa “Don Basilio S.R.L” a plena luz del día (horas de la mañana) y en la mesa de trabajo utilizada precisamente para el proceso de clasificación manual de la materia prima arribada.

Las fotografías fueron sacadas desde arriba, en línea recta y a una distancia aproximada de 50 cm, permitiendo ser observado el fondo blanco utilizado en la mesa de trabajo, como así también las características de cada una de las muestras.

En total fueron tomadas 630 fotografías, con una cámara digital Sony W320 de 14,1 mega píxeles; de ellas, 420 fueron elegidas según sus condiciones para ser procesadas (esto hace referencia a que cuenten con buena calidad, no estén movidas, con poca luz u otros defectos comunes en las fotografías); al mismo tiempo, de éstas últimas se seleccionaron 210 que posean una representación típica de la materia prima con la que se suele trabajar, como así también situaciones extrañas o diferentes que se podrían llegar a dar en el secadero, analizando de esta manera los puntos extremos (o situaciones límite) que podrían suceder.

De las últimas 210 fotografías citadas, 90 de ellas fueron tomadas para el entrenamiento de las técnicas de inteligencia artificial (30 de brotes, 30 de tallos verdes y 30 de tallos y hojas no verdes, en igual proporción e indicadas correspondientemente por el experto encargado de las tareas de clasificación) y las 120 restantes para las pruebas pertinentes, donde se pretende que 60 sean para situaciones “perfectas” (20 de brotes puros, 20 de tallos verdes puros y 20 de tallos y hojas no verdes puros), 30 para situaciones “comunes” (donde abunda el brote verde) y 30 para situaciones “extrañas” (10 para objetos extraños, 10 en donde abunde tallos y 10 en donde abunden hojas no verdes).

Cabe destacar, que si bien las fotografías fueron sacadas con una cámara de 14,1 mega píxeles (4320x3240 píxeles), éstas fueron adaptadas a un tamaño de 256x256 píxeles (sin pérdidas de calidad o de propiedades alguna) para permitir un procesamiento de las imágenes mucho más ágil y rápido, reduciendo los posibles tiempos de espera.

5.2 Descriptores de imágenes utilizados

Para poder utilizar las fotografías obtenidas, fue necesario encontrar una forma de representarlas de manera numérica, de tal forma que puedan ser utilizadas como datos de los procesos de entrenamiento y prueba propiamente dichos. Para esto, se utilizaron las propiedades RGB de cada una de las imágenes; éstas, indican las características de colores en rojo, verde y azul (*Red*, *Green*, *Blue*) de forma numérica para cada uno de los píxeles de una fotografía; [13] es decir, si cada fotografía tiene un tamaño de 256x256 píxeles se obtiene un total de 65.536 píxeles los cuales, a su vez, tienen 3 valores de color cada uno, por lo cual, el prototipo trabaja con un total de 196.608 datos por fotografía.

Una vez que el prototipo logra la obtención de las características numéricas de cada fotografía en RGB, para poder resumir de cierta manera la información capturada, se procedió a utilizar diferentes descriptores matemáticos: La media, la mediana, la varianza, la desviación estándar, el coeficiente de Curtosis y la correlación lineal. [14]

5.3 Configuraciones para las pruebas

Para realizar las diferentes pruebas se han utilizado tanto datos creados de forma supervisada (proporcionados por el prototipo a partir de instrucciones del usuario experto), como no supervisados (proporcionados por la técnica no supervisada clustering K-Means).

Los valores numéricos que representan a cada una de las fotografías para el entrenamiento y prueba dependen de las configuraciones citadas en la Tabla 1.

Tabla. 1. Configuraciones a utilizar para las pruebas planificadas

Configuración	Características
N° 1	Descriptores: Media, mediana, desviación estándar, varianza, coeficiente de curtosis y correlación lineal. RNA: 6 descriptores x 3 valores (RGB) = 18 neuronas de entrada
N° 2	Descriptores: media, mediana y desviación estándar. RNA: 3 descriptores x 3 valores (RGB) = 9 neuronas de entrada.
N° 3	Descriptores: media, varianza y coeficiente de Curtosis. RNA: 3 descriptores x 3 valores (RGB) = 9 neuronas de entrada.

Se han llevado a cabo 3 (tres) conjuntos de pruebas. La primera de ellas utilizando fotografías puras de las muestras (brote, tallo, hoja no verde), la segunda, utilizando fotografías donde la muestra refleje una situación bastante común en el secadero, que ocurre cuando existan mayoritariamente brotes en la muestra y la tercera, utilizando fotografías con situaciones extrañas que pudieran presentarse (muestras con mayoritaria cantidad de tallos, hojas no verdes o la presencia de objetos extraños). Esto puede observarse de forma resumida en la Tabla 2.

Tabla. 2. Pruebas realizadas

Prueba	Características
N° 1	Situaciones Perfectas: 20 fotos de Brotes, 20 fotos de Tallos, 20 fotos de Hojas no Verdes
N° 2	Situaciones Comunes: 30 fotos de Mayoritariamente Brotes
N° 3	Situaciones Extrañas: 10 fotos de Mayoritariamente Tallos, 10 fotos de Mayoritariamente Hojas no verdes, 10 fotos de Objetos Extraños

5.4 Pruebas de clasificación con Redes Neuronales y datos supervisados

Las pruebas con la técnica de inteligencia artificial de RNA, se llevaron a cabo con el prototipo desarrollado, logrando las clasificaciones que se pueden observar de manera resumida en las Tablas 3, 4 y 5 para aquellas que han utilizado datos supervisados.

Tabla. 3. Resultados de prueba N° 1 con RNA y datos supervisados

Configuración	Clasif. Correctas	Clasif. Incorrectas	% de acierto
N° 1	60	0	100 %
N° 2	60	0	100%
N° 3	58	2	96,67%

Tabla. 4. Resultados de prueba N° 2 con RNA y datos supervisados

Configuración	Clasif. Correctas	Clasif. Incorrectas	% de acierto
N° 1	30	0	100%
N° 2	30	0	100%
N° 3	29	1	96,67%

Tabla. 5. Resultados de prueba N° 3 con RNA y datos supervisados

Configuración	Clasif. Correctas	Clasif. Incorrectas	% de acierto
N° 1	25	5	83,33%
N° 2	28	2	93,33%
N° 3	25	5	83,33%

En la Tabla 3, se presentan los resultados de la prueba N° 1. Como se puede observar, los resultados obtenidos con estas configuraciones arrojaron porcentajes de clasificación superiores o iguales al 96,67%, entre los cuales se destacan las configuraciones 1 y 2 que obtuvieron un 100% de acierto.

La Tabla 4 manifiesta el resumen de los resultados obtenidos en la prueba N° 2. Esta prueba, refleja una de las situaciones más comunes en el escenario de un secadero de té, donde la mayoría del contenido de la muestra arribada representa a brotes de té con pequeña presencia posible de otras partes de la materia prima. Como se puede observar, igual que en la prueba N° 1, los porcentajes de acierto superan o igualan al 96,67%.

En la Tabla 5, pueden verse reflejados los resultados obtenidos de la Prueba N° 3. Ésta prueba evidencia situaciones extrañas que podrían darse en un secadero de té para ver la reacción del prototipo ante la presentación de situaciones límite, como ser la presencia de latas u objetos extraños, mayor cantidad de otras partes de la materia prima que no sean brotes, etc. Al entrenar el prototipo con brotes, tallos y hojas no verdes puras y luego realizar las pruebas con las muestras de situaciones extrañas pertinentes, se puede visualizar que en todas las configuraciones se han obtenido porcentajes de acierto iguales o superiores al 83,33%, existiendo problemas mayoritariamente en la detección de las situaciones donde abundan tallos verdes y hojas no verdes; no obstante, puede notarse, al mirar en detalle cada uno de los resultados obtenidos, prácticamente no existen problemas en el reconocimiento de objetos extraños presentes en las muestras, lo cual es la peor situación que se puede presentar y es de gran importancia que el prototipo fuese capaz de clasificarla de forma correcta.

5.5 Pruebas de clasificación con Redes Neuronales y datos no supervisados

Los resultados obtenidos en las pruebas de clasificación realizadas con Redes Neuronales y datos obtenidos de manera no supervisada se ven reflejados y ordenados en las Tablas 6, 7 y 8; cada uno éstos, fueron originados por la utilización de datos de entrenamiento obtenidos de manera no supervisada a través de la técnica de clustering K-means, por parte del prototipo desarrollado, en donde se especifica la cantidad de clases en las que la mencionada técnica debe agrupar los datos (en este caso 3 muestras).

Tabla. 6. Resultados de prueba N° 1 con RNA y datos no supervisados

Configuración	Clasif. Correctas	Clasif. Incorrectas	% de acierto
N° 1	57	3	95 %
N° 2	53	7	88,33 %
N° 3	53	7	88,33 %

Tabla. 7. Resultados de prueba N° 2 con RNA y datos no supervisados

Configuración	Clasif. Correctas	Clasif. Incorrectas	% de acierto
N° 1	29	1	96,67 %
N° 2	30	0	100 %
N° 3	5	25	16,67 %

Tabla. 8. Resultados de prueba N° 3 con RNA y datos no supervisados

Configuración	Clasif. Correctas	Clasif. Incorrectas	% de acierto
N° 1	20	10	66,67 %
N° 2	10	20	33,33 %
N° 3	17	13	56,67 %

En la Tabla 6, se presentan los resultados de la prueba N° 1 realizada con las diferentes configuraciones y con el uso de los datos creados de forma no supervisada por clustering K-Means. Como se puede observar, con las tres configuraciones, se obtuvieron porcentajes de acierto mayores al 85 %; sin embargo, los resultados no son tan altos como los que se manifestaron con muestras obtenidas de forma supervisada. Esto sin dudas se debe a la forma en que Clustering agrupa las muestras a utilizar, ya que existe cierto grado posible de error al armar las clases de forma no supervisada.

En la Tabla 7, se presentan los resultados de la prueba N° 2. Por una parte, la configuración N° 2 ha brindado una clasificación perfecta, a pesar de que los datos han sido configurados de forma no supervisada, al igual que la configuración N° 1 que presenta únicamente un error a la hora de llevar a cabo la clasificación; sin embargo, la configuración N° 3 arroja un valor realmente bajo que equivale a un 16,67%. Esto se debe a la combinación de los datos obtenidos de forma no supervisada por clustering junto con el uso de diferentes descriptores matemáticos (que no coinciden con las configuraciones 1 y 2).

En la Tabla 8, se presentan los resultados de la prueba N° 3 realizada con las diferentes configuraciones anteriormente citadas y con el uso de los datos creados de forma no supervisada por clustering. Al trabajar con datos no supervisados y situaciones extrañas que no suceden a menudo en el secadero, se puede observar que todos los resultados presentan porcentajes de acierto inferiores al 70%.

5.6 Resultados Finales de las pruebas

La Tabla 9 presenta una comparación entre cada uno de los porcentajes de acierto obtenidos en las 18 corridas totales llevadas a cabo entre las 3 pruebas, las 3 configuraciones y los datos supervisados y no supervisados presentes.

Tabla. 9. Comparativa resultados finales obtenidos

Prueba	Configuración	% acierto D. Sup.	% acierto D. No Sup.
N° 1	N° 1	100 %	95 %
	N° 2	100%	88,33%
	N° 3	96,67 %	88,33 %
N° 2	N° 1	93,33 %	96,67%
	N° 2	100 %	100 %
	N° 3	96,67 %	16,67 %
N° 3	N° 1	83,33%	66,67%
	N° 2	93,33%	33,33 %
	N° 3	88,33%	56,67%

En la Tabla 10, se pueden observar las diferentes denominaciones de las clasificaciones realizadas según su porcentaje de acierto; aquellas que han brindado mejores resultados, están catalogadas como buenas, muy buenas o perfectas, permitiendo de esta forma poder corroborar qué combinación de valores es la más adecuada para realizar clasificaciones mediante redes neuronales y las muestras presentadas.

Tabla. 10. Referencias de las denominaciones a las clasificaciones según su porcentaje de acierto.

Denominación	Rango
Perfecta	100%
Muy Buena	90% - 99,99%
Buena	80% - 89,99%
Regular	50% - 79,99%
Mala	20% - 49,99%
Muy mala	0% - 19,99%

6 Conclusiones

En base a las pruebas realizadas con el prototipo y las diferentes configuraciones, se eligieron aquellas que brindaran mayor cantidad de porcentaje de acierto para cada una de las pruebas. Partiendo de esto, se puede afirmar que la técnica de Inteligencia Artificial (IA) clasificatoria estudiada en este trabajo (RNA) es apta para la clasificación de materia prima ingresada a un secadero de té.

RNA logró obtener resultados muy buenos en cada una de las pruebas, tanto con datos supervisados, como no supervisados (generados por Clustering K-means), a excepción de la tercera de ellas, donde las muestras utilizadas pretendían generar una reacción diferente o poco usual en el prototipo, ya que las mismas coincidían poco y nada con las imágenes utilizadas para el entrenamiento, (pertenecientes a brote, tallo y hoja no verde en estado puro y sin alteraciones).

En cuanto a la generación de los datos y muestras, se utilizaron datos supervisados, los cuales, para su adquisición, demandan la presencia de un usuario humano experto que organice las muestras iniciales correspondientes, indicando cuáles de ellas pertenecen a cada una de las clases. Esta cuestión, lleva un tiempo extra que incluye la organización y preparación de muestras de prueba que se adapten a los escenarios

planteados, para que una vez entrenada y configurada la técnica, esta se base en estos criterios para realizar las clasificaciones correspondientes.

Por otro lado, se propuso, la obtención de datos de forma no supervisada, la cual no precisa de una organización de las muestras o especificaciones de cada uno de los tipos, ya que mediante la técnica de clustering K-means, el prototipo diseñado se encarga de organizar las muestras fotográficas que se le presenten según los criterios de similitud que encontrara en los valores RGB de cada una de ellas (representados por los descriptores matemáticos), reduciendo considerablemente el tiempo que conlleva la clasificación.

Teniendo en cuenta la problemática planteada al principio de este trabajo de investigación, referida a la necesidad de una forma estándar de clasificación y una manera de acelerar las clasificaciones que actualmente se realizan de forma manual en los secaderos de té provinciales, se considera que este prototipo es un acercamiento a una mejora o ayuda a la tarea diaria de clasificación realizada en la temporada de zafra a la hora de recibir materia prima por parte de diferentes proveedores. Esta afirmación se debe a que los resultados obtenidos por el mismo, luego de las pruebas realizadas (con sus respectivas configuraciones) que representan situaciones comunes y poco comunes en el proceso de recepción de materia prima de los secaderos, han sido muy buenos o perfectos en la mayoría de los casos presentados, por lo que se asume que RNA, es de gran ayuda para agilizar este proceso. Bajo esta afirmación, cabe destacar que, el prototipo desarrollado es capaz de agilizar la tarea clasificatoria y de generar un criterio estándar para realizar la misma, solucionando, de esta manera, la problemática existente.

7 Bibliografía

1. Ministerio de Ciencia, Tecnología e Innovación Productiva, Profecyt, Agencia de promoción Científica y Tecnológica, Unión Industrial Argentina. Té en Misiones: Debilidades y Desafíos tecnológicos del sector productivo. Buenos Aires, Argentina (2012).
2. Instituto Nacional de Tecnología Agropecuaria. Té: Procesos de Elaboración. <http://inta.gob.ar/documentos/te-procesos-de-elaboracion> (2011)
3. Ministerio del Agro y la Producción. Patrón de calidad para recepción de hoja de té verde. Posadas, Misiones, Argentina (2011).
4. Ministerio del Agro y la Producción. Procedimiento de muestreo y análisis de Hoja de té verde. Oberá, Misiones, Argentina (2011).
5. Secretaría de Agricultura, Ganadería, Pesca y Alimentos, Subsecretaría de Agroindustria y Mercados, Dirección Nacional de Agroindustria; Guía para la aplicación de Buenas Prácticas de Manufactura en Té Negro. Buenos Aires, Argentina. (2009).
6. Han J., Kamber M., Pei J. Data Mining: Concepts and Techniques. 3rd Edition. Waltham, Massachusetts, Estados Unidos. Editorial Elsevier. (2013)
7. Liu B. Web Data Mining. Chicago, Illinois, Estados Unidos. Springer editorial. (2011).
8. Hilera, J. R., Martínez, J. H. Redes Neuronales Artificiales. Fundamentos, Modelos y Aplicaciones. Ra-Ma ISBN: 8478971556. (1995).
9. Mitchell T. Machine Learning. McGraw-Hill Science/Engineering/Math Editorial. Carnegie Mellon University. Pittsburg. Pensilvania. Estados Unidos. (1997).
10. Haykin S. Neural Networks and Learning Machines. 3rd edition. Pearson Education Editorial. (2009)
11. Heaton J. Encog Machine Learning Framework <http://www.heatonresearch.com/encog/> (2016).
12. Machine Learning Group. Weka 3: Data Mining Software in Java. <http://www.cs.waikato.ac.nz/ml/weka/> (2016)
13. Gonzalez R., Woods R. Digital Image Processing 4th edition. University of Tennessee. United Estates. Pearson Education Editorial (2017)
14. Wlapole R., Myres R., Myres S., Ye K. Probabilidad y estadística para ingeniería y ciencias. 8va edición. Pearson Education Editorial (2007)