

Análisis de Consenso Comunitario en la Categorización de Imágenes de Incendios Forestales: Un Enfoque Estadístico para Evaluar la Percepción de las Comunidades de Paillaco y Valdivia

Ing. Sebastián Lara Barría

Facultad de Ciencias de la Ingeniería
Universidad Austral de Chile
General Lagos 2086
Valdivia, Chile
`sebastian.lara@alumnos.uach.cl`

Abstract. La gestión efectiva de incendios forestales requiere una comprensión profunda del riesgo y las percepciones comunitarias. Este proyecto se centra en analizar el consenso de categorización de imágenes de incendios forestales en las comunidades de Paillaco y Valdivia. El objetivo es identificar características visuales compartidas para optimizar la gestión local de incendios y contribuir al conocimiento de la percepción del riesgo. La metodología incluye la evaluación de respuestas comunitarias mediante herramientas estadísticas como la prueba de chi-cuadrado y la tasa de concordancia simple. Se examina la distribución de categorías elegidas, la categoría más frecuente por comunidad, y se evalúa el consenso tanto dentro de cada comunidad como entre Paillaco y Valdivia. Además, se realiza una comparación entre las categorizaciones de las comunidades y las categorizaciones proporcionadas por expertos en incendios forestales. Los datos, recopilados en la plataforma de ciencia ciudadana de percepciones comunitarias de incendios forestales *E-ncendio* (FONDEF ID22I10072), serán fundamentales para caracterizar el consenso comunitario y mejorar la comprensión de la gestión local del riesgo de incendios forestales. Este enfoque permitirá no solo identificar patrones de consenso dentro de cada comunidad, sino también comparar cómo difieren las percepciones entre comunidades y cómo se alinean con la experiencia de expertos en la materia.

Keywords: consensus, categorization, forest fires, community perceptions.

1 Introducción

Los incendios forestales representan una amenaza significativa para los ecosistemas y las comunidades en todo el mundo, siendo Chile particularmente susceptible a estos eventos [1]. La gestión efectiva de estos incidentes requiere no solo de estrategias técnicas, sino también de una comprensión profunda de la percepción comunitaria. En este contexto, el análisis de imágenes de incendios forestales se presenta como una valiosa herramienta para capturar las percepciones de las comunidades.

La ciencia ciudadana ha emergido como un enfoque prometedor para recopilar datos en diversos campos científicos [2] [3]. La participación comunitaria en la categorización de imágenes puede proporcionar información valiosa sobre el consenso territorial en la percepción de los incendios forestales.

Este trabajo se centra en el análisis de consenso comunitario en la categorización de imágenes de incendios forestales en las comunidades de Paillaco y Valdivia. El objetivo principal es evaluar la percepción colectiva en relación con la gestión de incendios forestales y determinar si existe un consenso significativo entre estas comunidades. A través de un enfoque estadístico, se busca analizar la distribución de categorías seleccionadas por los usuarios, identificar patrones de consenso y discrepancia, y comparar las categorizaciones comunitarias con la categorización correcta de cada imagen.

Este enfoque no solo contribuirá a comprender mejor la percepción comunitaria de los incendios forestales, sino que también proporcionará información valiosa para mejorar las estrategias de gestión local. A través de herramientas estadísticas especializadas, se busca ofrecer un análisis robusto y objetivo del consenso en la categorización de imágenes, destacando la importancia de considerar las percepciones locales en la gestión de incendios forestales.

2 Pregunta de Investigación

- **RQ:** ¿Existe un consenso significativo entre las comunidades de Paillaco y Valdivia en la categorización de imágenes relacionadas con incendios forestales, y cómo se compara este consenso con las categorías asignadas por expertos?

La comprensión del consenso comunitario en la categorización de imágenes de incendios forestales se convierte en un aspecto crucial para mejorar la gestión local de estos eventos en comunidades específicas. En este contexto, la pregunta de investigación busca desentrañar las dinámicas de percepción alrededor de las imágenes, tanto dentro de cada comunidad como en comparación con las categorías definidas por expertos en incendios forestales.

Este análisis no solo se centra en la identificación de patrones y tendencias de categorización dentro de Paillaco y Valdivia, sino que también busca establecer la consistencia y divergencia en las elecciones de categorías entre ambas comunidades. Además, la comparación con las categorizaciones expertas proporcionará perspectivas sobre la alineación o discordancia entre la percepción

comunitaria y el conocimiento especializado. La metodología empleada se apoya en herramientas estadísticas como la prueba de chi-cuadrado, tasa de concordancia simple y pruebas post hoc para evaluar la significancia de las diferencias observadas.

3 Descripción datos

3.1 Dataset de Incendios Forestales

A partir de una aplicación móvil de ciencia ciudadana, se obtuvieron un total de 99 imágenes asociadas con la gestión de incendios forestales. Estas imágenes corresponden a actividades asociadas con la prevención, mitigación, respuesta y recuperación frente a incendios forestales. Asimismo, las imágenes tienen una pertinencia territorial, donde la comunidad de usuarios de Paillaco y Valdivia participó en la categorización de estas imágenes relacionadas con incendios forestales.

Cada entrada en el dataset incluye los siguientes atributos:

- **idImage:** Un identificador alfanumérico único para cada imagen.
- **Categoría Elegida por el Usuario:** La categoría seleccionada por los distintos usuarios para cada imagen.
- **Commune:** La comuna en la cual reside el usuario.
- **name:** Corresponde al nombre de usuario.
- **Categoría Correcta:** La categorización correcta de la imagen, que puede ser asignada como "Prevención", "Mitigación", "Respuesta" o "Recuperación".

Estas categorizaciones ya han sido completadas por los usuarios de las respectivas comunidades. En la sección de Resultados, se presentarán y analizarán estas categorizaciones para comprender el consenso y las divergencias en las percepciones de las comunidades de Paillaco y Valdivia en relación con las imágenes de incendios forestales.

Puedes encontrar el repositorio con el dataset y cuadernillo jupyter en el siguiente enlace:

4 Metodología

4.1 Recopilación de Datos

La obtención de datos se llevó a cabo mediante la implementación de una aplicación móvil de ciencia ciudadana. Participantes de las comunidades de Paillaco y Valdivia contribuyeron en la categorización de un conjunto de 99 imágenes vinculadas a incendios forestales. Cada imagen fue etiquetada con la categoría correcta, que incluye Prevención, Mitigación, Respuesta y Recuperación. Además, se registraron las categorías seleccionadas por los usuarios de ambas comunidades.

Preparación y Exploración de Datos La preparación y exploración de datos se llevaron a cabo utilizando un cuadernillo Jupyter con el lenguaje de programación Python. Este enfoque permitió una manipulación eficiente de datos y un análisis exploratorio detallado. La calidad de los resultados del análisis estadístico depende en gran medida de la calidad de los datos utilizados. En este proyecto, la preparación de datos se realizó utilizando el entorno interactivo de Jupyter Notebook, lo que facilita la visualización y documentación de cada paso.

Exploración Inicial de Datos En la etapa inicial, se realizó una exploración detallada del conjunto de datos utilizando las capacidades interactivas de Jupyter Notebook. Esto incluyó el cálculo de la frecuencia de elección de categorías por cada comunidad, identificación de la categoría más elegida y otros patrones relevantes.

Comparación entre Comunidades Se realizó una comparación detallada de las distribuciones de categorizaciones y usuarios entre las comunidades de Paillaco y Valdivia utilizando gráficos interactivos y análisis estadísticos.

Herramientas Utilizadas Estos procesos se llevaron a cabo utilizando Jupyter Notebook con celdas de texto y Python, aprovechando las librerías como Pandas, NumPy, Matplotlib y Seaborn.

Puedes revisar el cuadernillo y dataset en el siguiente repositorio:

4.2 Análisis Estadístico

En el análisis estadístico, se emplearán diversas herramientas para explorar y comprender la percepción comunitaria en Paillaco y Valdivia.

- **Prueba de Chi-cuadrado:** Esta prueba se utiliza para evaluar la independencia entre dos variables categóricas. Se empleará para examinar la distribución de las categorías elegidas por las comunidades, determinando si existe una asociación significativa entre ellas.
- **Tasa de Concordancia Simple:** Esta medida evalúa la proporción de acuerdo entre las comunidades de Paillaco y Valdivia en la categorización de imágenes.
- **Prueba de Correlación de Kendall:** Esta prueba se emplea para evaluar la correlación entre las categorías asignadas por las comunidades y la categoría correcta.
- **Prueba t de Student:** Esta prueba es esencial para comparar las medias de las elecciones de categorías de las comunidades con respecto a la categoría correcta.

Análisis de Comparación con la Categoría Correcta Con el objetivo de evaluar la precisión de las categorizaciones comunitarias, se generó una tabla de contingencia que muestra la frecuencia de las categorías predichas en comparación con las categorías correctas. Esta tabla proporciona una visión general de la alineación entre las elecciones de la comunidad y la realidad del riesgo de incendios forestales.

Adicionalmente, se creó un gráfico de barras apiladas que representa visualmente la distribución de las categorías correctas y predichas. Este gráfico ofrece una perspectiva clara de las elecciones de la comunidad en relación con las categorías correctas.

Asimismo, se generó un mapa de calor que destaca las relaciones entre las categorías correctas y predichas. El uso de colores intensifica la visualización de la concordancia o discrepancia entre las elecciones comunitarias y la categoría correcta.

Estas representaciones visuales se utilizarán para contextualizar y respaldar el análisis estadístico de comparación de categorías, proporcionando una comprensión integral de la percepción comunitaria en Paillaco y Valdivia.

Estas herramientas estadísticas proporcionarán una evaluación integral de la percepción comunitaria sobre las imágenes de incendios forestales en Paillaco y Valdivia, permitiendo identificar patrones, consensos y discrepancias de manera rigurosa y objetiva.

5 Resultados

5.1 Análisis Descriptivo

En esta sección, se presentan los resultados del análisis descriptivo de las categorizaciones comunitarias en Paillaco y Valdivia. Se examinan patrones, tendencias y diferencias observadas en la percepción de las imágenes de incendios forestales por parte de ambas comunidades las cuales se pueden observar en la figura 1.

Distribución de Categorías Para comprender la distribución general de las elecciones de categorías, se calcularon las frecuencias de cada categoría seleccionada por la comunidad en el conjunto de datos. Los resultados proporcionan una visión panorámica de las preferencias y tendencias en la percepción comunitaria.

Consistencia y Consenso Se evaluó la consistencia en las elecciones de categorías dentro de cada comunidad, examinando la cantidad de consenso entre los participantes al categorizar las imágenes. Esta medida de consenso se utilizó para identificar patrones de acuerdo y desacuerdo en la percepción comunitaria.

Después de realizar un análisis descriptivo del conjunto de datos, se obtuvo la siguiente información: 99 imágenes en total, con un total de 504 categorizaciones realizadas por 30 usuarios. De estos, 17 usuarios pertenecen a la comunidad de

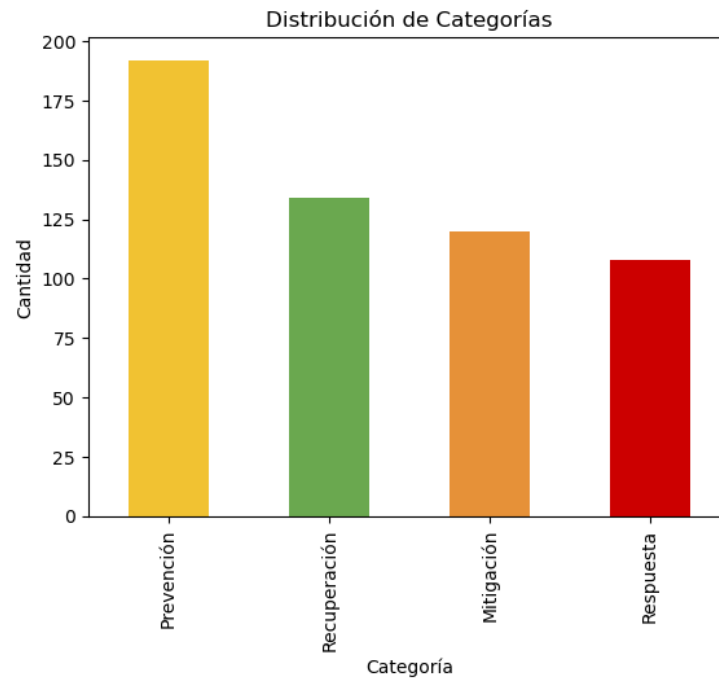


Fig. 1. Distribución de categorizaciones de usuarios (Paillaco y Valdivia).

Valdivia y 6 a la comunidad de Paillaco. En la figura 2 se pueden visualizar la distribución de categorías para cada comuna.

Se verificó que en la comunidad de Valdivia existe un consenso mayoritario en las categorizaciones, a diferencia de Paillaco donde dicho consenso es menos evidente. Es importante destacar que, a pesar de este consenso predominante en Valdivia, las categorizaciones de las imágenes por parte de la comunidad no coinciden consistentemente con las categorías correctas establecidas por los expertos.

5.2 Análisis Estadístico

En esta sección, se presentan los resultados del análisis estadístico realizado para explorar y comparar la percepción comunitaria en Paillaco y Valdivia.

Prueba de Chi-cuadrado Se realizó la prueba de Chi-cuadrado para evaluar la independencia entre las categorías seleccionadas por las comunidades de Paillaco y Valdivia. Los resultados de esta prueba proporcionan información sobre la asociación significativa o falta de ella entre las elecciones de categorías en ambas comunidades.

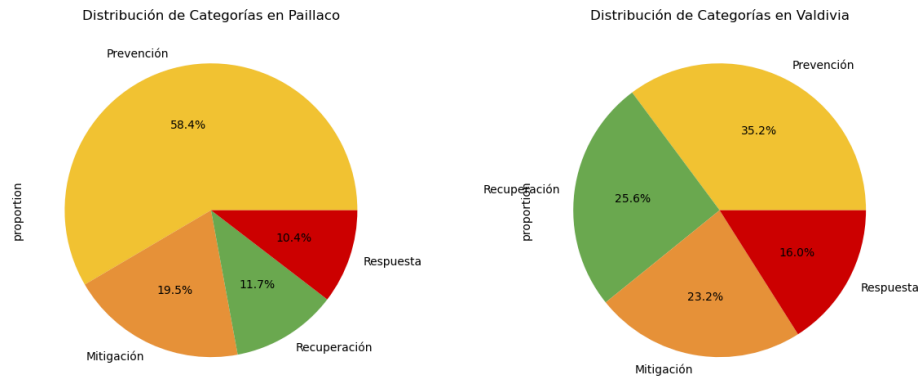


Fig. 2. Distribución de categorizaciones por comuna.

La tabla de contingencia para la prueba de Chi-cuadrado, utilizando las variables 'Categoría' y 'Comuna', arrojó un estadístico Chi-cuadrado de 54.60 con un p-valor muy bajo (2.08×10^{-6}). Este p-valor significativamente bajo sugiere que hay una asociación estadísticamente significativa entre las categorías seleccionadas por las comunidades y las categorías correctas. Por lo tanto, podemos rechazar la hipótesis nula de independencia, indicando que las elecciones de categorías no son independientes de la comunidad a la que pertenecen. Esta relación estadísticamente significativa subraya la importancia de considerar el contexto comunitario al analizar las percepciones de las imágenes de incendios forestales.

Tasa de Concordancia Simple y Correlación de Kendall La Tasa de Concordancia Simple se utilizó para evaluar el grado de acuerdo entre las comunidades en la categorización de imágenes. Asimismo, la prueba de Correlación de Kendall permitió examinar la correlación ordinal entre las categorías seleccionadas por las comunidades y las categorías correctas.

Los resultados obtenidos para la Tasa de Concordancia Simple y la Correlación de Kendall se presentan a continuación:

Resultados para Paillaco:

- Tasa de Concordancia Simple: 0.0779
- Coeficiente de Correlación de Kendall (Tau): -0.0409
- Valor p: 0.7047

Resultados para Valdivia:

- Tasa de Concordancia Simple: 0.224
- Coeficiente de Correlación de Kendall (Tau): 0.0605
- Valor p: 0.2825

Estos resultados indican que la Tasa de Concordancia Simple es baja tanto para Paillaco como para Valdivia. Además, el coeficiente de correlación de Kendall muestra una correlación débil en ambas comunidades. Los valores de p , que son mayores que el nivel de significancia común de 0.05, sugieren que no hay evidencia suficiente para rechazar la hipótesis nula de que no hay correlación entre las categorías seleccionadas por la comunidad y las categorías correctas. Esto podría indicar que la percepción de las categorías entre los participantes y los expertos no muestra un fuerte acuerdo o que la tarea de categorización es subjetiva.

Prueba t de Student La prueba t de Student se aplicó para comparar las medias de las elecciones de categorías entre Paillaco y Valdivia. Este análisis proporciona información sobre posibles diferencias significativas en la percepción comunitaria entre ambas comunidades.

Los resultados de la prueba t de Student indican un estadístico t de -1.85 con un valor p de 0.0664. El valor p es mayor que el umbral de significancia comúnmente utilizado (0.05). Por lo tanto, no hay evidencia suficiente para afirmar que hay una diferencia significativa en las elecciones de categorías entre Paillaco y Valdivia. Los resultados sugieren que, a nivel estadístico, las percepciones comunitarias en ambas comunidades no difieren de manera significativa.

Comparación con la Categoría Correcta A continuación, se presentan los resultados de la comparación entre las categorizaciones comunitarias y las categorías correctas. Esto permitirá evaluar la alineación de la percepción comunitaria con la realidad del riesgo de incendios forestales.

Los porcentajes de precisión para cada comunidad se detallan a continuación:

- Precisión de Paillaco: 7.79%
- Precisión de Valdivia: 22.40%

Estos resultados indican que la precisión de las categorizaciones comunitarias es baja para ambas comunidades. La Figura 3 muestra la distribución de las categorías seleccionadas por los usuarios en comparación con las categorías correctas. Las áreas más oscuras en la tabla de contingencia indican una mayor frecuencia de coincidencias, mientras que las áreas más claras representan discrepancias en las elecciones de categorías.

6 Conclusiones

En el análisis detallado de las categorizaciones comunitarias de imágenes de incendios forestales en las comunidades de Paillaco y Valdivia, se han extraído varias conclusiones clave:

1. **Prevalencia de la Categoría de Prevención:** En general, existe una tendencia entre los participantes a categorizar las imágenes en la clase de "prevención". Esta observación sugiere una fuerte asociación mental de las imágenes con acciones preventivas contra incendios forestales.

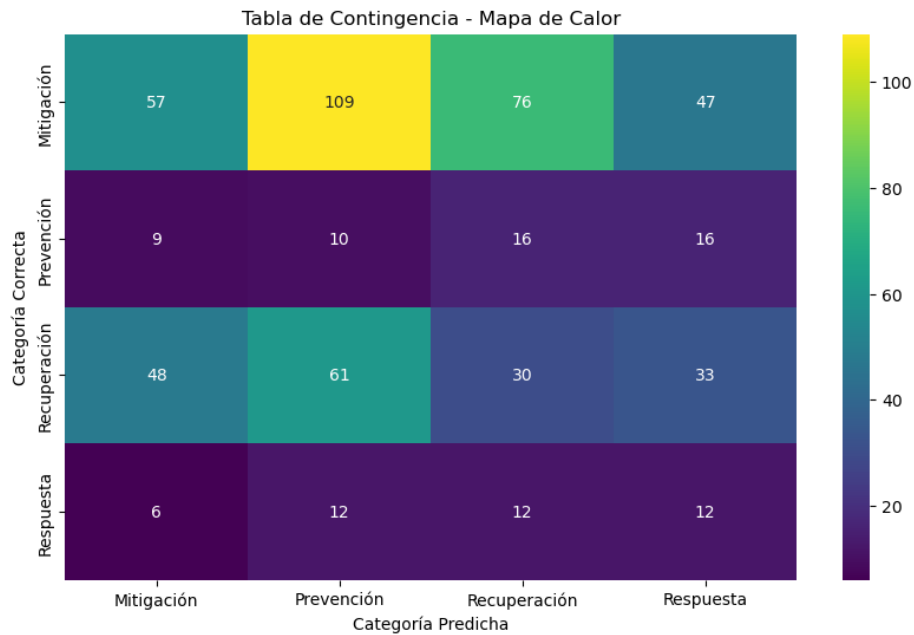


Fig. 3. Tabla de contingencia de las categorizaciones de usuarios y las categorías correctas.

2. **Confusión entre Mitigación y Prevención:** Se identificó una tendencia significativa a confundir las categorías de "mitigación" y "prevención". Este hallazgo resalta la necesidad de una comprensión más clara y educación sobre las diferencias entre estas categorías para mejorar la precisión de las futuras categorizaciones.
3. **Falta de Consenso en las Elecciones:** No se observó un consenso claro en la elección de categorías para cada imagen. La diversidad en las etiquetas asignadas a las imágenes indica la subjetividad de la percepción comunitaria, lo que puede deberse a la interpretación individual de la información presentada.
4. **Diversidad en las Etiquetas:** La diversidad de etiquetas asignadas a cada imagen sugiere que las comunidades tienen perspectivas variadas sobre las imágenes de incendios forestales. Este aspecto resalta la complejidad y la subjetividad asociada con la interpretación de estas imágenes.
5. **Dataset Preliminar:** La cantidad de categorizaciones y usuarios es baja en comparación con la complejidad del tema. Sin embargo, este análisis proporciona una visión preliminar valiosa y sienta las bases para futuras investigaciones más robustas a medida que el conjunto de datos se expande.
6. **Perspectiva para el Futuro:** A pesar de las limitaciones actuales, este estudio sirve como una herramienta inicial para comprender la percepción comunitaria sobre incendios forestales. Con un conjunto de datos más ex-

tenso y una participación más amplia, esta herramienta podría evolucionar para convertirse en un recurso valioso en la planificación y gestión de incendios forestales.

6.1 Respuesta a la Pregunta de Investigación

La pregunta de investigación planteada fue la siguiente: ¿Existe un consenso significativo entre las comunidades de Paillaco y Valdivia en la categorización de imágenes relacionadas con incendios forestales, y cómo se compara este consenso con las categorías asignadas por expertos?

Tras analizar las categorizaciones comunitarias en las comunidades de Paillaco y Valdivia, se han identificado los siguientes hallazgos clave:

- (a) **Consistencia en las Elecciones:** No se observa un consenso marcado en las elecciones de categorías para cada imagen dentro de cada comunidad. La diversidad en las etiquetas asignadas indica que no existe una interpretación uniforme de las imágenes entre los participantes.
- (b) **Comparación de Consenso:** Aunque ambos grupos presentan falta de consenso, se observa un mayor porcentaje de acierto en la comunidad de Valdivia en comparación con Paillaco. Esto sugiere que, en términos relativos, hay una mayor coherencia en las categorizaciones de Valdivia.
- (c) **Comparación con Categorías de Expertos:** La comparación de las categorizaciones comunitarias con las categorías asignadas por expertos revela que ambos grupos presentan porcentajes bajos de acierto. Este hallazgo indica que la percepción comunitaria no se alinea de manera consistente con las categorías correctas establecidas por expertos.

En resumen, no existe un consenso significativo en las elecciones de categorías dentro de cada comunidad, y la comunidad de Valdivia muestra un mayor porcentaje de acierto en comparación con Paillaco. Sin embargo, en ambas comunidades, la alineación con las categorías correctas es limitada, indicando la complejidad y subjetividad en la interpretación de imágenes relacionadas con incendios forestales.

Referencias

1. Pozo, R. A., Galleguillos, M., González, M. E., Vásquez, F., & Arriagada, R. (2022). Assessing the socio-economic and land-cover drivers of wildfire activity and its spatiotemporal distribution in south-central Chile. *Science of the total environment*, 810, 152002.
2. Basile, M., Russo, L. F., Russo, V. G., Senese, A., Bernardo, N. (2021). Birds seen and not seen during the COVID-19 pandemic: The impact of lockdown measures on citizen science bird observations. *Biological Conservation*, 256, 109079.
3. Rappold, A. G., Hano, M. C., Prince, S., Wei, L., Huang, S. M., Baghdikian, C., et al. (2019). Smoke Sense Initiative Leverages Citizen Science to Address the Growing Wildfire-Related Public Health Problem. *GeoHealth*, 3, 443–457.
4. Patagón Supercomputer, Austral University of Chile, 2021. <https://patagon.uach.cl>