

Evaluation of the level of accuracy of multispectral aerial imagery matching through the "Siamese" neural network mode

Abstract– This study evaluates the accuracy of the "Siamese" neural network for matching multispectral aerial images. A set of image data was preprocessed, and the neural network was trained. The precision, recall and F1-score metrics were analyzed, finding that the average precision was 56.35%, with values varying between 25.00% and 91.18%. Recall was more stable with a mean of 52.44%, while precision had a mean of 56.57%. It is concluded that the "Siamese" neural network is effective for matching multispectral aerial images, although the accuracy depends on the training configuration and the characteristics of the data set.

Keywords– Image Matching; Multispectral aerial images; Siamese neural network; classification performance metrics.

Evaluación del nivel de precisión del emparejamiento de imágenes aéreas multiespectrales a través del modelo de red neuronal "Siamesas"

Eder David Pacheco Ramos, Estudiante de Ingeniería de Sistemas e Informática¹, Dios Castillo Christian Abraham, Dr en Administración de la Educación², and, Mariana Chavarry Chankay, Maestría en administración con mención en gerencia empresarial³

^{1,2,3}Universidad Tecnológica del Perú, Perú, U20247648@utp.edu.pe¹, cdios@utp.edu.pe², C17520@utp.edu.pe³

Resumen– Este estudio evalúa la precisión de la red neuronal "siamesa" para hacer coincidir imágenes aéreas multiespectrales. Se procesó previamente un conjunto de datos de imágenes y se entrenó la red neuronal. Se analizaron las métricas de precisión, recuperación y puntaje F1, encontrando que la precisión promedio fue del 56,35%, con valores que variaron entre el 25,00% y el 91,18%. El recuerdo fue más estable con una media del 52,44%, mientras que la precisión tuvo una media del 56,57%. Se concluye que la red neuronal "siamesa" es eficaz para comparar imágenes aéreas multiespectrales, aunque la precisión depende de la configuración del entrenamiento y de las características del conjunto de datos.

Palabras clave: *Coincidencia de imágenes; Imágenes aéreas multiespectrales; Red neuronal siamesa; Métricas de rendimiento de clasificación*

I. INTRODUCCIÓN

Encontrar pares de conjuntos de imágenes con una determinada conexión viene a ser una tecnología elemental en la disciplina de la visión por computadora. La esencia de numerosos campos de estudio radica en la tarea de identificar parejas de imágenes, abarcando aspectos como la concordancia de imágenes y la recuperación de estas, entre otros. En diversas esferas de investigación, el emparejamiento de imágenes desempeña un papel fundamental, extendiéndose a áreas como el seguimiento, el reconocimiento de objetos, la reconstrucción tridimensional de perspectivas múltiples y la estructura derivada del movimiento (SfM), entre otras aplicaciones [1]. En este sentido, también existen métodos que utilizan el aprendizaje discriminativo de modelos BoW (Bag of Words o Bolsa de Palabras) para predecir qué pares de imágenes en el conjunto de datos de entrada coinciden y cuáles no [2].

La revisión de la literatura científica respecto a la problemática de determinar un nivel de precisión del emparejamiento de imágenes aéreas multiespectrales a través del modelo de red neuronal siamesa (SNN), revela que muchos investigadores poseen muy pocas herramientas que soporten la facilidad de analizar datos con relación al emparejamiento de imágenes multiespectrales. Por esto, la clasificación de escenas tradicionales de imágenes de teledetección de alta resolución se basa en una red simple, desde una perspectiva única, utilizando imágenes de teledetección satelital con un modelo de entrenamiento para clasificación y predicción [3]. Así también, muestra la necesidad de aplicar un modelo de inteligencia artificial para

automatizar tareas de emparejamiento. En ese sentido, definimos a una red siamesa como un tipo de red neuronal profunda que, en contexto de procesamiento de imágenes, puede abordar los problemas de emparejamiento de imágenes multiespectrales [4]. Siendo así que, La precisión de estos modelos es esencial para mejorar la calidad de datos utilizados en aplicaciones clave, desde sistemas de asistencia automática basados en reconocimiento facial hasta la detección de enfermedades en hojas de kiwi mediante enfoques de aprendizaje profundo [5].

El estado del arte muestra la necesidad de la aplicación de algoritmos de emparejamiento de imágenes para practicas investigativas, en donde se facilite el control de los datos y la consulta de pares de imágenes. Además, se logra identificar que, los modelos basados en la idea de aprender una función de similitud entre imágenes utilizando técnicas de machine learning permite obtener resultados eficientes y precisos en investigaciones que trata imágenes como datos de análisis [6]. En este sentido, una red neuronal siamesa es un modelo de clasificación único que puede realizar predicciones a partir de pocos datos de entrenamiento, ya que una red siamesa es más resistente al desequilibrio de clases porque requiere menos información [7].

El análisis de teledetección y la investigación de la radiación mediante cámaras multiespectrales se benefician significativamente de las técnicas de emparejamiento de imágenes. Estas no solo facilitan la adquisición de datos clasificados, sino también agilizan el proceso de análisis. Las imágenes multiespectrales ofrecen una comprensión más completa del entorno al capturar información detallada sobre la composición espectral de las superficies. Las aplicaciones de las imágenes multiespectrales abarcan una amplia gama de campos en la investigación científica, abordando aspectos como la monitorización ambiental, la agricultura, la planificación urbana y la defensa [8].

Recientemente, el modelo de red neuronal siamesa se ha aplicado a varios problemas de coincidencia en imágenes, como verificación de rostros, coincidencia de bloques de imágenes locales y coincidencia de imágenes completas [1]. Se puede definir a una red neuronal siamesa como una arquitectura de red que incluye 2 subredes idénticas. Las CNN

(red neuronal convolucional) siamesas tienen una configuración igual, donde los mismos parámetros y pesos compartidos se combinan con una métrica de distancia [9]. Además, La red Siamesa es adecuada para aplicaciones a gran escala con grandes conjuntos de datos, como la detección de cambios en imágenes de teledetección a nivel regional o global [10]. Sin embargo, este proceso requiere una gran capacidad computacional y, por lo tanto, es crucial abordar eficientemente los desafíos computacionales asociados con la implementación de esta arquitectura en entornos prácticos y del mundo real.

En el contexto del presente estudio buscamos abordar el problema del emparejamiento de imágenes, a través de la interrogante: ¿Cuál es nivel de precisión del emparejamiento de imágenes aéreas multiespectrales a través del modelo de red neuronal "Siamesas"? El objetivo es determinar el nivel de precisión del emparejamiento de imágenes aéreas multiespectrales a través del modelo de red neuronal "Siamesas". Los objetivos específicos se han planteado de la siguiente manera: 1) Preparar el dataset para el entrenamiento de la red neuronal, 2) Aplicar la arquitectura del modelo "Siamesas" y entrenarlo con el dataset, y 3) Desarrollar entrenamientos a través del modelo de red neuronal "Siamesas". Tomando como base el avance del aprendizaje profundo, se decidió elegir la arquitectura de red neuronal siamesa en los experimentos. En ese sentido, la clasificación automatizada de imágenes mediante aprendizaje profundo ofrece la posibilidad de mejorar la eficiencia y la precisión del análisis de imágenes. Esto supone un enfoque más dinámico y adaptativo, permitiendo que el modelo aprenda patrones complejos y características relevantes directamente de los datos, lo cual es esencial en el contexto del emparejamiento de imágenes aéreas multiespectrales. En problemas prácticos de búsqueda por similitud, es frecuente recurrir a una sola instancia representativa de cada objeto dentro del conjunto de datos empleado. Esta particularidad presenta un desafío notable al elegir las Redes Neuronales Convolucionales (CNNs) como método para la extracción de características [11]. En este contexto, la arquitectura siamesa se destaca como una solución que mejora considerablemente la búsqueda de pares de imágenes, abordando eficazmente esta limitación. Además, buscamos facilitar la clasificación de imágenes a los investigadores, así como ahorrar tiempos en actividades diferentes del análisis de dichas imágenes, ya que el proceso de captura de relaciones intrínsecas y sutilezas presentes en las imágenes ya estaría automatizado, lo que debería repercutir en mayor precisión y robustez en el emparejamiento.

II. METODOLOGÍA

La unidad de análisis está constituida por los entrenamientos a la red neuronal "Siamesas". La población fue de 29 entrenamientos, de los cuáles se obtuvieron datos para determinar el nivel precisión del emparejamiento de imágenes.

• Medidas de evaluación

Para la evaluación, se usó las métricas de rendimiento de clasificación estándar: Precisión, Recall y F1-score, las cuales se definen en las siguientes formulas:

$$\begin{aligned} \text{precisión} &= \frac{TP}{TP + FP} \\ \text{recall} &= \frac{TP}{TP + FN} \\ \text{F1 - score} &= \frac{2 * \text{precisión} * \text{recall}}{\text{precisión} + \text{recall}} \end{aligned}$$

Donde: El término "verdadero positivo" (TP) indica la cantidad de muestras de prueba positivas que son correctamente predichas como positivas, "falso positivo" (FP) se refiere a la cantidad de muestras de prueba negativas que son incorrectamente predichas como positivas, "verdadero negativo" (TN) representa la cantidad de muestras de prueba negativas que son correctamente predichas como negativas, y "falso negativo" (FN) denota el número de muestras de prueba positivas que son incorrectamente predichas como negativas [12].

• Medidas de correlación:

Las relaciones dependientes entre variables se pueden analizar con métodos como el análisis de distancia y el análisis de correlación [13]. Se utilizó el indicador estadístico de correlación de Pearson, para determinar la relación entre los valores de los indicadores de la Precisión, el Recall y el F1-score; de esta manera se pudo identificar la dependencia entre estos valores. Se utilizó el coeficiente de variación para determinar el nivel de volatilidad o estabilidad de los indicadores Precisión, el Recall y el F1-score, en los 29 entrenamientos.

II. RESULTADOS

A. Preparación del conjunto de datos.

Durante esta fase, se llevó a cabo un meticuloso proceso de preprocesamiento de las imágenes con el fin de asegurar la coherencia y calidad de los datos que alimentarían el modelo de red neuronal siamesa. Se trabajó con un total de 1316 pares de imágenes multiespectrales y sus contrapartes originales, provenientes de dos directorios distintos designados como "Parejas" y "No parejas". Cada uno de estos pares consistía en una imagen multiespectral y su correspondiente imagen original, y fueron sometidos a una serie de transformaciones para su adecuación.

Primero, se ajustó el tamaño de todas las imágenes a una resolución estándar de 224x224 píxeles para asegurar que todas las entradas al modelo tuvieran las mismas dimensiones. Luego, se convirtieron al formato .jpg y se cambiaron al espacio de color RGB para una representación más uniforme y consistente de los datos de imagen.

La biblioteca "os", aplicada al lenguaje de programación Python, se utilizó para navegar por el sistema de archivos y acceder a las carpetas que contenían las imágenes del dataset que cuenta con 658 pares de imágenes "parejas" y 658 pares de imágenes "no parejas", permitiendo la lectura y carga de las imágenes desde el sistema de archivos de Google Drive en el entorno de Colab.

Utilizando la librería OpenCV (cv2) con Python, se cargaron las imágenes del dataset y se aplicaron operaciones de preprocesamiento y manipulación. Esto incluyó la conversión de las imágenes a escala de grises, el ajuste de tamaño a 224x224 píxeles y la normalización de los valores de píxeles. Además, NumPy, también una biblioteca aplicada al lenguaje de programación Python, proporciona estructuras de datos de matriz y rutinas numéricas rápidas relacionadas [14]. en ese sentido Numpy fue fundamental para la manipulación eficiente de datos en forma de matrices, convirtiendo las imágenes preprocesadas en matrices NumPy para su posterior procesamiento por el modelo de red neuronal.

En conjunto, este proceso de normalización y estandarización fue fundamental para garantizar que el modelo de red neuronal siamesa pudiera aprender patrones significativos de las imágenes sin verse afectado por diferencias en la resolución o el formato de archivo, asegurando así la calidad y coherencia de los datos de entrada.

B. Aplicación de la arquitectura del modelo "Siamesas" y entrenamiento con el conjunto de datos.

En esta fase, se procedió a la implementación de la arquitectura de la red neuronal siamesa utilizando la biblioteca Keras en un entorno de desarrollo con Python. En ese sentido, Keras es una herramienta de nivel superior para el aprendizaje profundo que simplifica la creación de modelos complicados, como las redes neuronales siamesas, utilizadas en el autoaprendizaje [15]. Así mismo, esta arquitectura siamesa, diseñada específicamente para problemas de comparación de pares, consta de dos ramas idénticas que comparten los mismos pesos y parámetros. Cada rama toma una de las dos imágenes del par como entrada y pasa a través de capas convolucionales y de pooling (agrupación) para extraer características relevantes. En el proceso, se realiza una agrupación para reducir el valor obtenido en los resultados de convolución de la imagen. El proceso de agrupación dividirá los resultados de la convolución en varias cuadrículas, luego tomará el valor más grande de cada cuadrícula y formará una nueva matriz que contiene los valores más grandes de cada cuadrícula previamente dividida [16].

Posteriormente, estas características se concatenan y se pasan a través de capas densas para la clasificación binaria. La elección de la función de activación ReLU en la capa oculta densa y la función de activación sigmoide en la capa de salida.

Las funciones de activación sigmoidea y ReLU garantizan que el modelo pueda aprender correlaciones no lineales complejas entre los puntos de datos de entrada [12].

Una vez aplicada la arquitectura del modelo, se procedió al entrenamiento utilizando el dataset preparado en la etapa anterior. Se utilizó el optimizador Adam y la función de pérdida de entropía cruzada binaria para minimizar el error de clasificación del modelo durante el entrenamiento. La entropía cruzada binaria es una función de pérdida popular para tareas de clasificación binaria, y se ha demostrado que funciona bien con el optimizador Adam. La combinación de Adam y la entropía cruzada binaria puede lograr una alta precisión y convergencia rápida en diversos tipos de problemas [17].

Se llevaron a cabo múltiples épocas (Epoch) de entrenamiento, ajustando tanto la configuración de las épocas como la tasa de aprendizaje (Lr), con el fin de facilitar que el modelo captara de manera efectiva los patrones en los datos y ajustara sus pesos y parámetros para maximizar la precisión de la clasificación. Después de cada entrenamiento, se guardaron los modelos resultantes para su posterior evaluación y uso en la detección de imágenes duplicadas.

La Figura 1 describe la red neuronal siamesa para el emparejamiento de imágenes. Dos imágenes se procesan en paralelo por dos redes neuronales convolucionales idénticas que extraen características visuales y las reducen de dimensionalidad. Las características se concatenan y se procesan por capas densas para obtener una probabilidad de que las imágenes sean una pareja. La función de activación sigmoide genera una salida binaria: 1 (pareja) o 0 (no pareja).

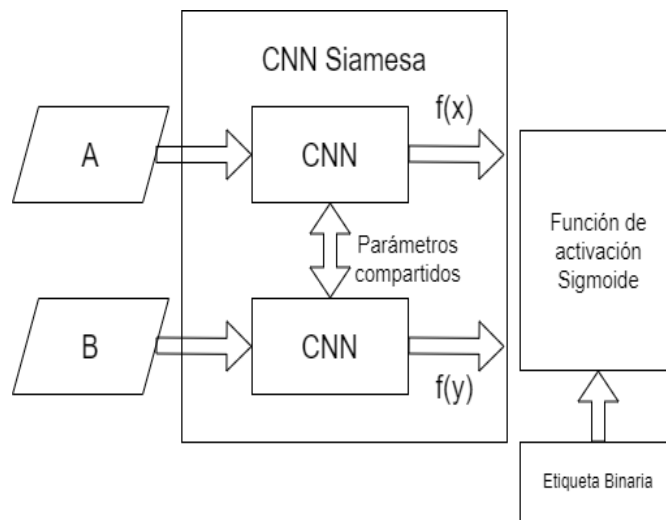


Fig. 1 Arquitectura de red siamesa muestra ambas entradas para las clases pares y no pares las cuales se entran compartiendo parámetros y uniéndose en una función de activación final con salida binaria.

C. Desarrollo de los entrenamientos a través del modelo de red neuronal Siamesa

Para los 29 entrenamientos se ha considerado la misma configuración básica de la red neuronal “Siamesas”: Función de activación Sigmoide y Número de capas ocultas.

Los siguientes valores son solo de tendencia, según el número de entrenamientos realizados:

TABLA I. RESULTADOS DE COEFICIENTES DE CORRELACIÓN.

	Precisión	Recall	F1-score
Media:	56,35%	52,44%	51,27%
Mediana:	55,70%	51,99%	54,33%
DesvEst:	24,99%	4,36%	12,85%
CV:	44%	8%	25%

Nota: Resultados de coeficientes de correlación para las métricas de rendimiento de clasificación: Precisión, Recall y F1-score. De los 29 entrenamientos.

- El entrenamiento en el cual se obtuvo el mayor nivel de precisión (91.18%), fue el número 24, con un valor de Epoch (épocas) de 26, Lr (la tasa de aprendizaje) de 0.001. Sin embargo, en la misma muestra obtenemos un valor bajo de recall (50.42%) Debido a ello se puede afirmar que el modelo no detecta la clase muy bien, pero cuando lo hace es altamente confiable.

- En los 29 entrenamientos, los valores del nivel de Precisión son variantes, mientras del indicador Recall, son estables; con el 44% y 8% de coeficiente de variación, respectivamente. En este sentido, se deduce que, según las pruebas realizadas, los resultados tienden a mostrar que las características iniciales influyen en gran medida al cambio de precisión en contraste con el recall que no se ve muy afectado a pesar del cambio de la configuración inicial, demostrando que la arquitectura de red neuronal Siamesa tiene una tendencia a mejorar el nivel de fiabilidad en la predicción de detecciones verdaderas.

- Existe una relación inversa y escasa entre el Epoch y la Precisión (Coeficiente de correlación de Pearson = -0.15895938). En este sentido, se afirma que el cambio en las épocas(epoch) dentro de la arquitectura no guarda una relación significativa en las predicciones verdaderas estimadas en la precisión.

- Existe una relación directa y media entre el Epoch y el Recall (Coeficiente de correlación de Pearson = 0.645774187). Debido a ello, Se deduce que la variación de las épocas en la configuración inicial del entrenamiento de la red neuronal Siamesa influye en cierta medida frente a las predicciones negativas estimadas en el Recall.

TABLA II. RESULTADOS DEL INDICADOR PRECISIÓN CON UN VALOR DE LR DE 0.001.

Epoch	Precisión
10	38.28%
13	52.58%
15	25.00%
17	81.76%
19	83.89%
22	81.91%
24	89.36%
26	91.18%
28	31.10%

Nota: Resultados de los 29 entrenamientos.

Gracias a los datos recogidos en las 29 predicciones Se puede afirmar que las mejores muestras con altos índices de precisión se encuentran en la configuración de épocas entre el 10 y 28 con una tasa de aprendizaje de 0.001, dando a entender que este es un rango viable y funcional para la arquitectura en el contexto de emparejamiento de imágenes.

IV. DISCUSIÓN

Gracias a los resultados obtenidos, Se puede enfocarnos en varios aspectos clave relacionados con el uso de la CNN Siamesa para la clasificación de imágenes multispectrales, teniendo en cuenta las observaciones y citas proporcionadas por otros autores.

Así, tomando en cuenta el aprovechamiento de la combinación de identificación y modelos de verificación, Wang y Liu (2023) proponen una CNN Siamesa para la clasificación de escenas de teledetección con etiquetas imperfectas. Su método logra resultados comparables o mejores que otros métodos de última generación, incluso con etiquetas de entrenamiento ruidosas o inconsistentes. En contraste, el enfoque usado en la presente investigación se basa en un conjunto de técnicas que incluyen el suavizado de etiquetas y el cambio de parámetros para mejorar la robustez del modelo frente a etiquetas imperfectas y la especificidad del conjunto de datos [18].

Por otro lado, también se analizó el impacto de las etiquetas de verdad imperfectas durante el entrenamiento, donde se encontró que el entrenamiento con etiquetas imperfectas puede afectar negativamente el rendimiento del modelo. Zhang y Zhang (2023) proponen una CNN Siamesa con suavizado de etiquetas para la clasificación de escenas de teledetección con etiquetas imperfectas [19]. Al igual que en la presente investigación, su método reduce el impacto del ruido o presencia de errores en las etiquetas de entrenamiento.

Además, en relación con la especificidad de las características de cada conjunto de datos, Li y Zhang (2023) proponen un método de adaptación de dominio para la clasificación de escenas de teledetección basado en redes neuronales siamesas y aprendizaje supervisado [18]. Similar al enfoque usado en la presente investigación, su método mejora la capacidad de generalización del modelo a diferentes conjuntos de datos. En contraste, también afirmamos que el empleo de pesos de otras redes preentrenadas impacta negativamente dependiendo de la naturaleza del origen de datos, ya que en el contexto de imágenes multiespectrales no se busca siempre una imagen completamente igual. Sin embargo, el preentrenamiento con aprendizaje supervisado es una técnica reciente que ha demostrado ser efectiva para mejorar el rendimiento de los modelos de aprendizaje profundo. Un estudio reciente de Liu et al. (2022) propone un método de preentrenamiento basado en aprendizaje supervisado contrastivo para imágenes de teledetección multifuente [20]. Si bien no se analizó el preentrenamiento en este trabajo, es un área que merece ser investigada en el futuro para mejorar aún más el rendimiento de las CNN Siamesas para la clasificación de imágenes multiespectrales.

En conjunto, las investigaciones de estos autores respaldan la eficacia y la versatilidad de las CNN Siamesas para la clasificación de escenas de teledetección. Nuestro trabajo complementa estas investigaciones al abordar desafíos específicos como la presencia de errores o inconsistencias en los datos de entrenamiento y la especificidad de las características del conjunto de datos.

V. CONCLUSIONES

La aplicación de la red neuronal siamesa en el emparejamiento de imágenes aéreas multiespectrales representa un avance significativo en la mejora de la precisión y la robustez de los modelos utilizados en este campo de estudio. En esta investigación, se abordaron tres objetivos específicos.

En primer lugar, se llevó a cabo un proceso exhaustivo de preprocesamiento de imágenes, garantizando la coherencia y la calidad de los datos de entrada mediante el uso de diversas bibliotecas y técnicas de manipulación de imágenes.

Seguidamente, se implementó la arquitectura de la red neuronal siamesa utilizando la biblioteca Keras, y se entrenó el modelo con el conjunto de datos preparado, utilizando el optimizador Adam y la función de pérdida de entropía cruzada binaria.

Finalmente, se desarrollaron múltiples entrenamientos con diferentes configuraciones de la red neuronal, analizando los resultados mediante métricas de precisión, recall y F1-score, y

observando el impacto significativo de factores como el número de épocas y la tasa de aprendizaje en el rendimiento del modelo. Se identificaron limitaciones relacionadas con el tamaño del conjunto de datos y la necesidad de realizar más investigaciones para validar los resultados obtenidos.

Como recomendaciones futuras, se sugiere explorar técnicas de preentrenamiento para mejorar el rendimiento del modelo, así como investigar el uso de diferentes arquitecturas de red neuronal y evaluar el modelo en una variedad de conjuntos de datos para una validación más completa y robusta.

En síntesis, aunque los resultados presentados representan tendencias observadas en los entrenamientos, este estudio proporciona evidencia sólida de la eficacia de la CNN Siamesa en el emparejamiento de imágenes aéreas multiespectrales, señalando áreas clave para futuras investigaciones y mejoras en el campo.

REFERENCIAS

- [1] H. H. Yu, J. Liu, H. Sun, Z. Wang, and H. Zhang, "GetNet: Get Target Area for Image Pairing," in International Conference Image and Vision Computing New Zealand, 2019. doi: 10.1109/IVCNZ48456.2019.8960995.
- [2] J. Ning, H. Shi, J. Ni, and Y. Fu, "Single-stream deep similarity learning tracking," IEEE Access, vol. 7, 2019, doi: 10.1109/ACCESS.2019.2939367.
- [3] Q. Liu, R. Hang, H. Song, and Z. Li, "Learning multiscale deep features for high-resolution satellite image scene classification," IEEE Transactions on Geoscience and Remote Sensing, vol. 56, no. 1, 2018, doi: 10.1109/TGRS.2017.2743243.
- [4] M. Hazratifard, V. Agrawal, F. Gebali, H. Elmiligi, and M. Mamun, "Ensemble Siamese Network (ESN) Using ECG Signals for Human Authentication in Smart Healthcare System," Sensors, vol. 23, no. 10, 2023, doi: 10.3390/s23104727.
- [5] D. Banerjee, V. Kukreja, S. Hariharan, and V. Sharma, "Fast and Accurate Multi-Classification of Kiwi Fruit Disease in Leaves using deep learning Approach," in International Conference on Innovative Data Communication Technologies and Application, ICIDCA 2023 - Proceedings, 2023. doi: 10.1109/ICIDCA56705.2023.10099755.
- [6] M. Oplenskedal, P. Herrmann, and A. Taherkordi, "DEEPMATCH2: A comprehensive deep learning-based approach for in-vehicle presence detection," Inf Syst, vol. 108, 2022, doi: 10.1016/j.is.2021.101927.
- [7] R. Newlin Shebiah, S. Arivazhagan, S. G. Amrith, and S. Adarsh, "Person Re-Identification by Siamese Network," Inteligencia Artificial, vol. 26, no. 71, 2023, doi: 10.4114/intartif.vol26iss71pp25-33.
- [8] A. S. Laliberte, M. A. Goforth, C. M. Steele, and A. Rango, "Multispectral remote sensing from unmanned aircraft: Image processing workflows and applications for rangeland environments," Remote Sens (Basel), vol. 3, no. 11, 2011, doi: 10.3390/rs3112529.
- [9] M. VAROL ARISOY, "SIGNATURE VERIFICATION USING SIAMESE NEURAL NETWORK ONE-SHOT LEARNING," International Journal of Engineering and Innovative Research, vol. 3, no. 3, 2021, doi: 10.47933/ijeir.972796.
- [10] H. Yang et al., "A Lightweight Siamese Neural Network for Building Change Detection Using Remote Sensing Images," Remote Sens (Basel), vol. 15, no. 4, 2023, doi: 10.3390/rs15040928.
- [11] Andrés J. Pascal, Adrián N. Planas, Federico Lederhos, and Norma E. Herrera, "Extracción de Características utilizando Redes Siamesas y Aumentación aplicada a las Búsquedas por Similitud de Marcas de Ganado," Oct. 2022.
- [12] F. Akbar, M. Hussain, R. Mumtaz, Q. Riaz, A. W. A. Wahab, and K. H. Jung, "Permissions-Based Detection of Android Malware Using Machine

- Learning,” *Symmetry (Basel)*, vol. 14, no. 4, 2022, doi: 10.3390/sym14040718.
- [13] F. Xu, J. Gao, and X. Pan, “Cow Face Recognition for a Small Sample Based on Siamese DB Capsule Network,” *IEEE Access*, vol. 10, 2022, doi: 10.1109/ACCESS.2022.3182806.
- [14] P. Virtanen *et al.*, “SciPy 1.0: fundamental algorithms for scientific computing in Python,” *Nat Methods*, vol. 17, no. 3, 2020, doi: 10.1038/s41592-019-0686-2.
- [15] R. Ji, J. Li, and L. Zhang, “Siamese self-supervised learning for fine-grained visual classification,” *Computer Vision and Image Understanding*, vol. 229, 2023, doi: 10.1016/j.cviu.2023.103658.
- [16] Y. Pratama, S. T. N. Nainggolan, D. I. Nadya, and N. Y. Naipospos, “One-shot learning Batak Toba character recognition using siamese neural network,” *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 21, no. 3, 2023, doi: 10.12928/TELKOMNIKA.v21i3.24927.
- [17] S. J. Reddi, S. Kale, and S. Kumar, “On the convergence of Adam and beyond,” in 6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings, 2018.
- [18] X. Liu, Y. Zhou, J. Zhao, R. Yao, B. Liu, and Y. Zheng, “Siamese Convolutional Neural Networks for Remote Sensing Scene Classification,” *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 8, 2019, doi: 10.1109/LGRS.2019.2894399.
- [19] R. Zhang *et al.*, “Remote sensing image scene classification with noisy label distillation,” *Remote Sens (Basel)*, vol. 12, no. 15, 2020, doi: 10.3390/RS12152376.
- [20] C. Liu, H. Sun, Y. Xu, and G. Kuang, “Multi-Source Remote Sensing Pretraining Based on Contrastive Self-Supervised Learning” *Remote Sens (Basel)*, vol. 14, no. 18, 2022, doi: 10.3390/rs14184632.