Evaluación automática de escritura a mano usando características visuales de forma y clasificación supervisada

Lic. Jorge Durán Carmona[1], Dra. Rocío Alfonsina Lizárraga Morales[2]
[1]Departamento de Estudios Multidisciplinarios, División de Ingenierías, Campus Irapuato – Salamanca, Universidad de Guanajuato
[2]Departamento de Arte y Empresa, División de Ingenierías, Campus Irapuato – Salamanca, Universidad de Guanajuato
j.durancarmona@ugto.mx, ra.lizarragamorales@ugto.mx

Resumen/Abstract

La escritura a mano constituye una habilidad fundamental en las primeras etapas del aprendizaje escolar, ya que favorece la comunicación escrita y el desarrollo de habilidades motrices finas. La evaluación de la escritura manual infantil suele depender de criterios observacionales, rúbricas o instrumentos psicométricos. Este trabajo presenta un prototipo de clasificación automática de imágenes de escritura manuscrita en dos clases: escritura correcta e incorrecta. Se trabajó con 80 imágenes obtenidas a partir de la escritura de una misma frase: “El veloz murciélago hindú comía feliz”, escrita en un espacio en blanco sin renglones ni líneas guía. Esta condición permitió mantener constante la longitud textual y hacer más comparables las características geométricas de las muestras. La metodología incluyó captura y recorte digital manual de la región de interés, binarización mediante el método de Otsu, inversión de la imagen para representar los trazos como pixeles blancos sobre fondo negro, detección de bordes con el método de Canny y extracción de características geométricas. A partir de estas características se entrenaron y compararon cinco modelos supervisados: máquinas de soporte vectorial (SVM), máquina de vectores de soporte con kernel RBF, k vecinos más cercanos, Random Forest y perceptrón multicapa. Los modelos fueron evaluados mediante validación cruzada de cinco particiones. Los resultados mostraron que el SVM con kernel lineal obtuvo la mayor exactitud global, con 81.25%, mientras que el SVM con kernel RBF alcanzó 80.00% y redujo el número de falsos positivos respecto al SVM lineal. Estos resultados sugieren que las características geométricas extraídas de una frase manuscrita bajo condiciones controladas contienen información útil para diferenciar parcialmente entre escrituras correctas e incorrectas, aunque se requiere ampliar el conjunto de datos y refinar los criterios de etiquetado para fortalecer la generalización del sistema.

Palabras clave: escritura manual; aprendizaje automático; procesamiento de imágenes; características de forma; características extraídas.

Introducción

La escritura a mano continúa siendo una habilidad relevante en el ámbito escolar, ya que permite comunicar aprendizajes, completar actividades académicas y desarrollar destrezas relacionadas con la motricidad fina. (Feder y Majnemer, 2003) señalan que los instrumentos de evaluación de escritura no solo deben considerar la dificultad observada, sino también propiedades psicométricas diversas que se dan en el transcurso del aprendizaje. Asimismo, también destacan que estos procesos de evaluación son inherentemente desafiantes debido a la subjetividad del evaluador, lo que se suma a la evidente escasez de herramientas de evaluación que sean plenamente válidas, fiables y normadas.

En el contexto actual educativo en México, desde el marco de la Nueva Escuela Mexicana, la evaluación de la escritura infantil se concibe como parte de un proceso formativo amplio, vinculado con el uso del lenguaje escrito en situaciones significativas (SEP, 2023). Sin embargo, en la práctica escolar, la valoración de la escritura a mano suele apoyarse en observaciones cualitativas del docente, especialmente en aspectos como la legibilidad, la alineación, la separación entre palabras y la regularidad del trazo. Esto puede representar una dificultad para clasificarse para profesores con menos experiencia docente en los primeros grados de la educación primaria, quienes podrían enfrentar mayores retos para distinguir entre escrituras con características gráficas adecuadas y aquellas que presentan problemas de organización, separación o legibilidad. En este sentido, los sistemas automáticos de análisis de escritura pueden complementar las evaluaciones tradicionales al ofrecer mediciones objetivas, cuantitativas y reproducibles.

La evaluación tradicional de la escritura incluye instrumentos que analizan legibilidad, velocidad, alineación, espaciamiento y calidad de la forma escrita. (Feder y Majnemer, 2003) revisaron herramientas como el Minnesota Handwriting Assessment (MHA) (Reisman, 1999), el Evaluation Tool of Children's Handwriting (ETCH) (Amundson, 1995) y otros instrumentos orientados a evaluar el desempeño de escritura a mano infantil. Este antecedente justifica que variables como la alineación, tamaño, espaciamiento y regularidad sean consideradas dimensiones relevantes para el análisis de la escritura a mano.

Desde el campo del reconocimiento de patrones, la escritura a mano se ha estudiado tanto en escenarios en línea como fuera de línea. (Tappert *et al.*1990) describieron que la escritura fuera de línea se procesa después de haber sido producida, generalmente a partir de una imagen digitalizada, lo que la relaciona con el reconocimiento óptico de caracteres y con tareas de preprocesamiento como segmentación, limpieza y extracción de características. Mientras los métodos en línea, que capturan las variables dinámicas durante el trazo, emplean dispositivos para obtener información adicional a la imagen pura, como en el caso de (Bublin *et al.* 2023), donde emplearon un lápiz sensorizado (SensoGrip) acoplado a redes LSTM para predecir la calidad caligráfica a partir de la presión y aceleración. Por otro lado, en el ámbito de las plataformas digitales, (Provenzale *et al.* 2022) desarrollaron en MATLAB un sistema de evaluación semiautomática para analizar la geometría y espaciado de letras en tablets interactivas. Asimismo, (Hedmi *et al.* 2022) propusieron un modelo híbrido basado en CNN y descriptores morfológicos para evaluar la dirección y forma del trazo en tiempo real. (Bin Durayhim *et al.* 2023), propusieron modelos de redes convolucionales para reconocimiento de escritura árabe infantil y destacaron la importancia de aplicaciones educativas capaces de proporcionar retroalimentación al estudiante. Después de hacer la revisión del estado del arte, se identifican dos corrientes principales en el estudio de la evaluación de escritura a mano: el de los datos capturados en línea y el de los datos analizados fuera de línea a partir de imágenes digitalizadas y procesadas. También se identifica que los trabajos realizados con redes profundas suelen requerir mayores volúmenes de datos.

La presente investigación se enfoca en un procesamiento fuera de línea sobre imágenes estáticas, teniendo como objetivo desarrollar y evaluar un prototipo de clasificación automática de imágenes de escritura a mano mediante características geométricas extraídas en MATLAB. A diferencia de un sistema de reconocimiento de caracteres, el propósito no fue transcribir la frase escrita, sino clasificar la muestra completa como escritura correcta o incorrecta a partir de patrones de alineación, separación y regularidad de los trazos.

Metodología

El flujo metodológico general se presenta en la Figura 1. El proceso inició con la captura de las imágenes para crear un dataset, continuó con el preprocesamiento y la extracción de características, y finalizó con el entrenamiento y evaluación de distintos modelos supervisados.

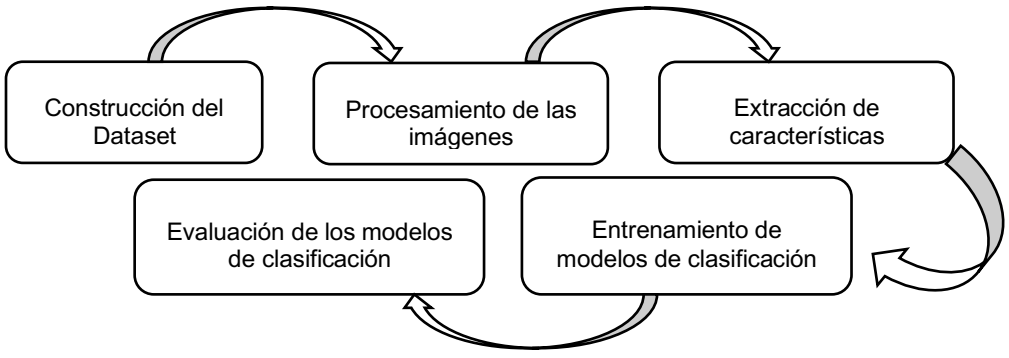


Figura 1. Flujo metodológico general.

1. Construcción del Dataset

Se trabajó con 80 imágenes de escritura manuscrita, distribuidas en dos clases relativamente balanceadas: 41 imágenes etiquetadas como escritura correcta y 39 imágenes etiquetadas como escritura incorrecta. Para construir el conjunto de datos, se recabaron muestras físicas de escritura a mano en hojas impresas con un espacio en blanco destinado a la escritura de una misma frase: “El veloz murciélago hindú comía feliz cardillo y kiwi”. Todas las muestras fueron escritas dentro de un mismo espacio en blanco, semejante a un renglón sin líneas guía, con el propósito de mantener constante la longitud textual y las condiciones espaciales de escritura entre las imágenes.

El proceso de etiquetado de las 80 muestras fue realizado por un Licenciado en Pedagogía con 14 años de experiencia en el ámbito educativo. La clasificación cuantitativa se fundamentó en la evaluación con criterios gráficos de la escritura infantil y de adultos, tales como el mantenimiento de la alineación horizontal, la regularidad del tamaño de cada grafía, la proporcionalidad del trazo y el mantenimiento de un espaciado uniforme entre letras y palabras. En la Figura 2 se muestra un ejemplo de una imagen de la muestra de escritura etiquetada como correctas.

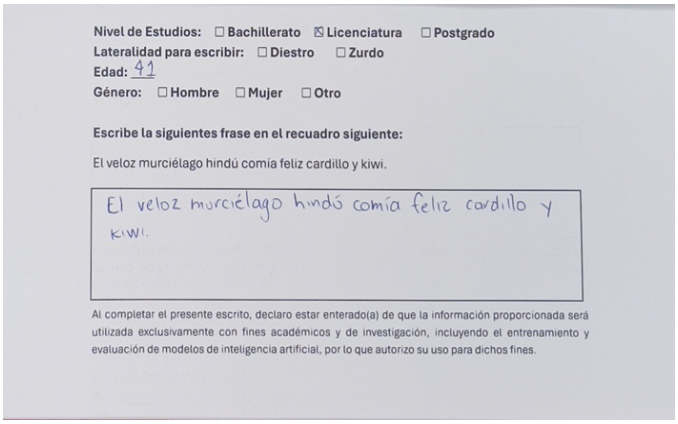


Figura 2. Ejemplo de una imagen de la muestra etiquetada como correctas.

Las fotografías fueron tomadas fijando el dispositivo móvil en un atril estabilizador a una distancia y altura constantes respecto a una base de captura, lo cual crea condiciones controladas de inclinación, iluminación, cámara y configuración. Las fotografías fueron tomadas en un tamaño de 50MP desde una cámara de un teléfono celular marca CUBOT modelo P80, mediante la aplicación CAMARA que viene instalada por defecto en el dispositivo. Las imágenes se guardaron posteriormente en formato PNG para evitar pérdidas por compresión. Posteriormente, cada imagen fue recortada digitalmente para eliminar el máximo espacio en blanco externo y conservar únicamente el fragmento de la frase “El veloz murciélago hindú comía feliz”. Esta delimitación se realizó debido a las variaciones en el tamaño de la letra de los participantes, y de esta manera, conservando únicamente el área ocupada por la escritura en una sola línea continua. Este recorte no modificó la forma de los trazos ni alteró la escritura original; únicamente delimitó la región de interés correspondiente al área efectiva de escritura en una sola línea para todas las imágenes. Esta decisión permitió que las dimensiones finales de la imagen, pixeles de ancho por pixeles de alto, fueran utilizadas como características geométricas adicionales durante la clasificación. A continuación, se muestra en la Figura 3 un ejemplo de una imagen recortada de la muestra, en este caso clasificada dentro del grupo “correctas”.

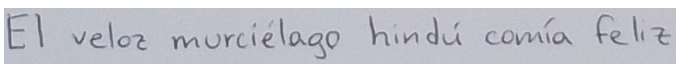


Figura 3. Ejemplo de una imagen recortada de la muestra etiquetada como correctas.

2. Procesamiento de las imágenes

Antes de extraer características, las imágenes fueron binarizadas mediante el método de Otsu. Este método selecciona automáticamente un umbral a partir del histograma de niveles de gris, buscando separar los pixeles en dos clases mediante un criterio de varianza (Otsu, 1979). Posteriormente, las imágenes fueron invertidas para representar los trazos de escritura como pixeles blancos sobre fondo negro. Esta inversión, tal como se muestra en la Figura 4, permitió que las funciones de análisis de componentes trabajaran sobre los trazos y no sobre el fondo completo de la imagen.



Figura 4. Imagen binarizada y con los valores invertidos de pixeles blancos y negros.

Después de la binarización e inversión, se aplicó el detector de bordes *Canny* mediante la función *Edge,* de MATLAB. Esta etapa permitió obtener una representación delgada de los contornos de los trazos, reduciendo el problema de que las palabras o la frase completa fueran identificadas como un solo objeto conectado. A continuación, en la Figura 5, se muestra la imagen de salida de una muestra, después de haberle aplicado el detector de bordes *Canny.*



Figura 5. Imagen de salida después de aplicarle el detector de bordes Canny de la función Edge en MATLAB.

Debido a que todas las imágenes se capturaron con la misma frase escrita en una sola línea, el procesamiento de bordes permitió analizar de manera más comparable la distribución espacial de los componentes de la escritura a mano.

3. Extracción de características

Las características se obtuvieron con la función *regionprops* de MATLAB sobre la imagen de bordes. Para cada componente conectado se recuperaron el centroide, la caja delimitadora y el área. A partir de esos datos se calcularon variables de alineación, separación y regularidad. La alineación se estimó mediante la pendiente de una recta ajustada sobre los centroides ordenados de izquierda a derecha. La separación se calculó con las distancias euclidianas entre centroides consecutivos. La regularidad se representó mediante promedios y desviaciones estándar de alto y ancho de las cajas delimitadoras, además del alto máximo y mínimo detectado. En la Tabla 1, se pueden apreciar las variables extraídas correspondientes a cada una de las características geométricas del texto de la imagen; mientras que la Figura 6, Figura 7 y Figura 8 corresponden a la descripción visual de cómo se obtuvo la pendiente estimada a partir de centroides, la altura de los componentes mediante *BoundingBox* y la separación máxima y separación mínima entre centroides consecutivos, respectivamente.

Tabla 1*. Características geométricas extraídas de las imágenes.*

Característica	**Variable extraída**
Dimensión de la imagen	Ancho Imagen
	Alto Imagen
Alineación	Pendiente
Separación	Separación Máxima
	Separación Mínima
Regularidad	Alto Promedio
	Ancho Promedio
	Desviación Alto
	Desviación Ancho
	Alto Máximo de Letra
	Alto Mínimo de Letra



Figura 6. Pendiente estimada a partir de centroides.



Figura 7. Altura de componentes mediante BoundingBox.



Figura 8. Separación máxima y mínima entre centroides consecutivos.

4. Entrenamiento de Modelos de Clasificación

Se evaluaron cinco modelos de clasificación supervisada. El SVM lineal se incluyó como modelo base de margen máximo (Cortes y Vapnik, 1995). El SVM con kernel RBF se incorporó para permitir una frontera de decisión no lineal, apoyada en el principio de transformar implícitamente el espacio de características mediante funciones kernel (Boser *et al.,* 1992), (Aizerman *et al.,* 1964). Además, se evaluó k vecinos más cercanos, método no paramétrico basado en la clase de las muestras más cercanas (Cover y Hart, 1967); Random Forest, ensamble de árboles de decisión construido mediante aleatorización y agregación (Breiman, 2001); y un perceptrón multicapa entrenado mediante aprendizaje supervisado, relacionado con el algoritmo de retropropagación (Rumelhart *et al.*, 1986).

Los hiperparámetros utilizados en cada modelo se muestran en la Tabla 2. Estos valores corresponden a la configuración empleada en MATLAB durante la evaluación experimental.

***Tabla 2**. Hiperparámetros utilizados en los modelos evaluados*

Modelo	Hiperparámetros utilizados en MATLAB	Interpretación
SVM con kernel lineal	KernelFunction = linear Standarize = true	Usa la frontera de decisión lineal. Transforma las variables para que tengan las escalas comparables.
SVM con kernel RBF	KernelFunction = rbf KernelScale = auto Standarize = true	Usa la función kernel radial para construir una frontera de decisión no lineal. Permite que MATLAB estime automáticamente la escala del kernel.
kNN	NumNeighbors = 5 Distance = euclidean Standarize = true	Clasifica cada muestra según las 5 muestras más cercanas. Utiliza la distancia euclidiana para medir la

		cercanía entre vectores de características.
Random Forest	Method = Bag NumLearningCycles = 100 MaxNumSplits = 20	Utiliza bagging para entrenar un ensamble de 100 árboles, y limita a 20 la complejidad de cada árbol base.
MLP	LayerSize = [10 5] IterationLimit =1000 Standarize = true	Utiliza una red neuronal con dos capas ocultas de 10 y 5 neuronas. Define el máximo de iteraciones de entrenamiento a 1000.

Todos los modelos se entrenaron y evaluaron con las mismas características extraídas de las imágenes. También se fijó la semilla del generador pseudoaleatorio con *rng(1)*, con el fin de mejorar la reproducibilidad del experimento y reducir variaciones asociadas a procesos aleatorios internos.

5. Evaluación de los Modelos de Clasificación

La validación del prototipo se realizó con el propósito de medir el desempeño de los modelos supervisados en la clasificación de imágenes de escritura a mano en dos clases: escritura incorrecta y escritura correcta. Debido a que el conjunto de datos estuvo conformado por 80 imágenes, se utilizó validación cruzada de cinco particiones, con el fin de aprovechar todas las muestras tanto para el entrenamiento de los modelos como para las pruebas.

La evaluación se realizó mediante validación cruzada de cinco particiones. En cada iteración, cuatro particiones fueron utilizados para entrenar el modelo y el pliegue restante se utilizó para prueba. Este procedimiento se repitió hasta que todas las particiones fueron empleados como conjunto de prueba una vez. Las predicciones obtenidas en las cinco particiones se integraron para construir una matriz de confusión por cada modelo evaluado. La clase positiva se definió como escritura correcta, representada con la etiqueta 1, mientras que la clase negativa correspondió a escritura incorrecta, representada con la etiqueta 0.

A partir de la matriz de confusión se calcularon los verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos. Con estos valores se obtuvieron las métricas de exactitud, error de clasificación, precisión, sensibilidad, especificidad, puntuación F1 y exactitud balanceada. Estas métricas permiten analizar el comportamiento de los clasificadores desde distintos criterios, evitando depender únicamente de la exactitud global (Sokolova y Lapalme, 2009).

Las métricas principales, se calcularon de la siguiente manera:

$$Exactitud = \frac{TP+TN}{TP+TN+FP+FN}, \quad (1)$$

$$Precisión = \frac{TP}{TP+FP}, \quad (2)$$

$$Sensibilidad = \frac{TP}{TP+FN}, \quad (3)$$

$$Error = \frac{FP+FN}{TP+TN+FP+FN}, \quad (4)$$

$$Especificidad = \frac{TN}{TN+FP}, \quad (5)$$

$$Puntuación\ F1 = \frac{2(Precision)(Recall)}{Precision+Recall}, \quad (6)$$

$$Exactitud\ Balanceada = \frac{Recall + Especificidad}{2}, \quad (7)$$

donde TP corresponde a las muestras de escritura correcta clasificadas como correctas; TN a las muestras de escritura incorrecta clasificadas como incorrectas; FP a las muestras de escritura incorrecta clasificadas como correctas; y FN a las muestras de escritura correcta clasificadas como incorrectas.

Para comparar el desempeño de distintos enfoques de clasificación, se evaluaron cinco modelos supervisados bajo las mismas condiciones experimentales: SVM lineal, SVM con kernel RBF, k vecinos más cercanos, Random Forest y perceptrón multicapa. Todos los modelos utilizaron el mismo conjunto de características geométricas extraídas de las imágenes.

La Tabla 3 muestra las métricas generales obtenidas por cada modelo.

***Tabla 3**. Métricas generales de los modelos evaluados*

Modelo	Exactitud (%)	Error (%)	F1 (%)	Exactitud Balanceada (%)
MLP	66.25	33.75	67.47	66.20
Random Forest	71.25	28.75	72.29	71.20
kNN	72.50	27.50	78.00	71.92
SVM RBF	80.00	20.00	80.00	80.05
SVM lineal	81.25	18.75	83.52	80.96

La tabla 4 presenta las métricas relacionadas con la clasificación por clase, así como los errores de falsos positivos y falsos negativos. Estas medidas permiten observar con mayor detalle el tipo de error cometido por cada modelo.

***Tabla 4**. Métricas por clase y errores de clasificación*

Modelo	Precisión (%)	Sensibilidad (%)	Especificidad (%)	FP	FN
SVM lineal	76.00	92.68	69.23	12	3
SVM RBF	82.05	78.05	82.05	7	9
kNN	66.10	95.12	48.72	20	2
Random Forest	71.43	73.17	69.23	12	11
MLP	66.67	68.29	64.10	14	13

A continuación, la Figura 9, Figura 10, Figura 11, Figura 12 y Figura 13, muestran las matrices de confusión para los modelos SVM con kernel lineal, SVM con kernel RBF, kNN, Random Forest y perceptrón multicapa, respectivamente. En este trabajo, el orden de clases utilizado fue *[0,1],* donde *0* corresponde a escritura incorrecta y *1* a escritura correcta.

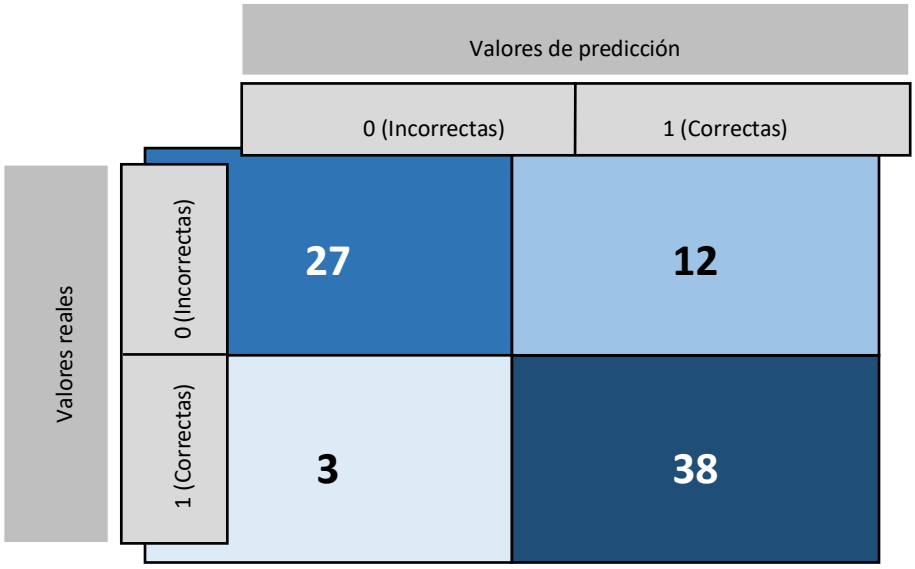


Figura 9. Matriz de confusión del modelo SVM con kernel lineal.

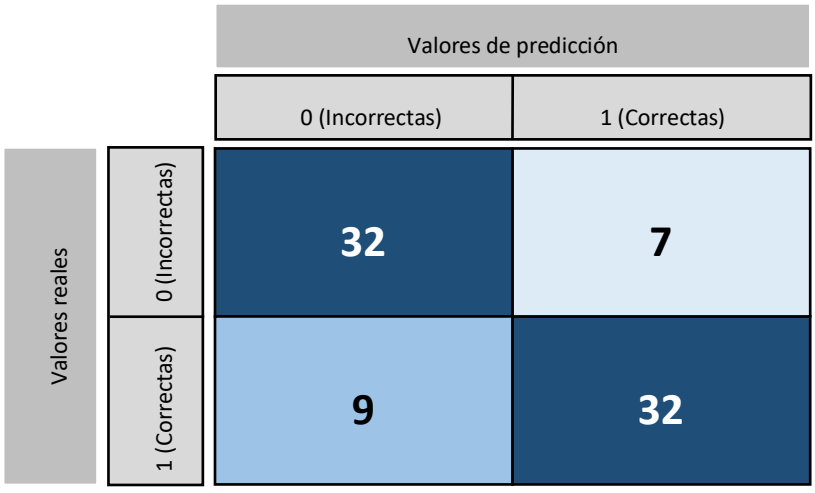


Figura 10. Matriz de confusión del modelo SVM con kernel radial.

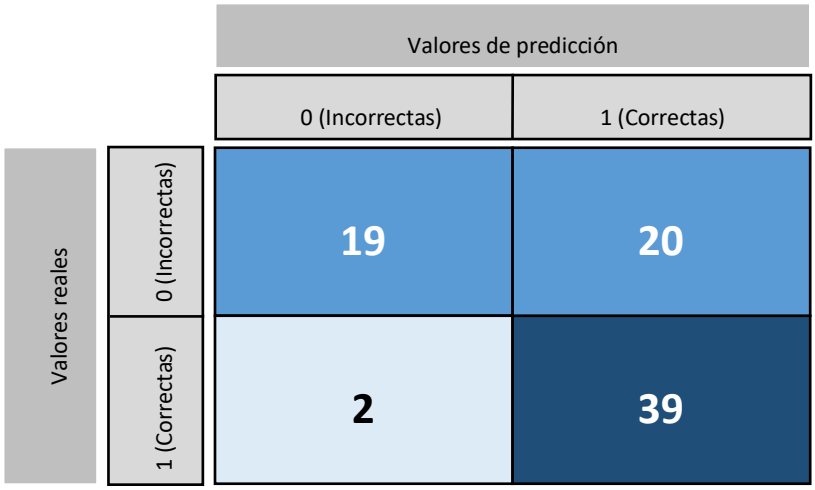


Figura 11. Matriz de confusión del modelo kNN.

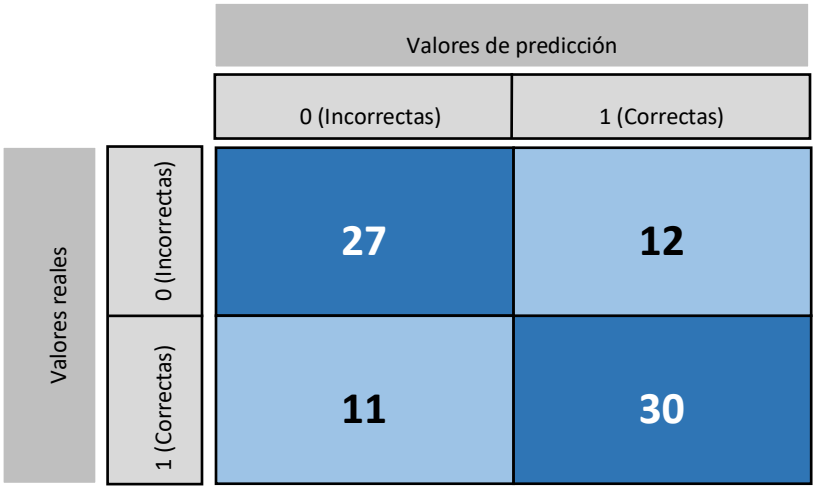


Figura 12. Matriz de confusión del modelo Random Forest.

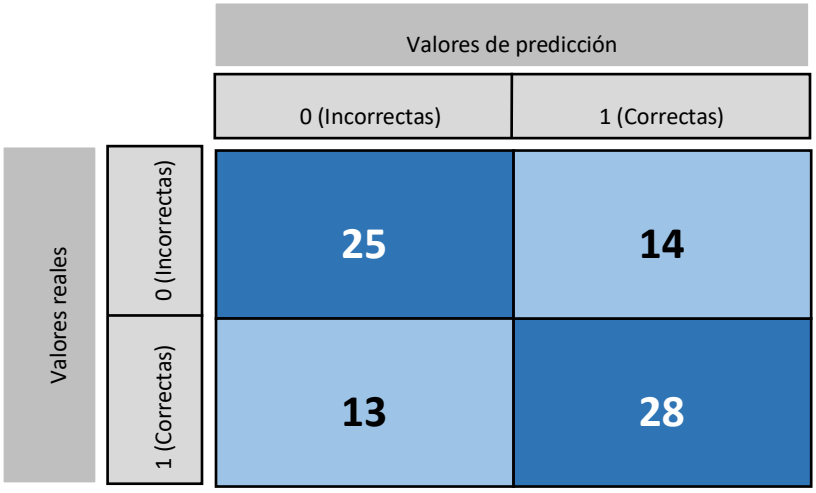


Figura 13. Matriz de confusión del modelo de una MLP.

5.1 Discusión de resultados de los modelos entrenados

Los resultados muestran que el SVM lineal obtuvo la mayor exactitud global, con 81.25 %, además de la puntuación F1 más alto, con 83.52 %. Esto indica que, bajo el criterio de desempeño general, el SVM lineal fue el modelo con mejor resultado global. Sin embargo, al analizar el tipo de error cometido, el SVM con kernel RBF presentó un comportamiento relevante, ya que redujo los falsos positivos de 12 a 7 en comparación con el SVM lineal.

En este trabajo, un falso positivo corresponde a una muestra de escritura incorrecta clasificada como correcta. Este tipo de error puede considerarse más delicado en un sistema de apoyo a la evaluación de escritura, ya que implicaría aceptar como adecuada una muestra que presenta características gráficas problemáticas. Por esta razón, aunque el SVM RBF tuvo una exactitud global ligeramente menor que el SVM lineal, puede considerarse preferible cuando el criterio principal es disminuir el riesgo de clasificar como correcta una escritura incorrecta.

El modelo kNN obtuvo el valor de sensibilidad más alto, con 95.12 %, lo que indica que identificó correctamente la mayoría de las muestras de escritura correcta. No obstante, también presentó la especificidad más baja, con 48.72 %, y el mayor número de falsos positivos, con 20. Esto significa que tendió a clasificar muchas escrituras incorrectas como correctas, lo cual puede ser problemático si se desea utilizar el sistema como apoyo para identificar dificultades gráficas.

Random Forest presentó un desempeño intermedio, con una exactitud de 71.25 % y una exactitud balanceada de 71.20 %. Por otro lado, el perceptrón multicapa obtuvo el desempeño más bajo, con 66.25 % de exactitud. Este resultado puede relacionarse con el tamaño reducido del conjunto de datos, ya que las redes neuronales suelen requerir mayor cantidad de muestras para generalizar adecuadamente.

En conjunto, la validación cuantitativa muestra que las características geométricas extraídas de la escritura manuscrita aportan información útil para diferenciar parcialmente entre las dos clases. Sin embargo, también evidencia que el desempeño del sistema depende del criterio de selección del modelo: si se prioriza la exactitud global, el SVM lineal fue el mejor; si se prioriza reducir falsos positivos, el SVM con kernel RBF ofrece un comportamiento más conveniente.

Resultados

Los resultados obtenidos muestran que las características geométricas extraídas de las imágenes contienen información útil para diferenciar parcialmente entre escritura correcta e incorrecta. Las variables relacionadas con la dimensión de la imagen, la pendiente, la separación entre componentes y la regularidad de alto y ancho, permitieron representar aspectos visuales de la escritura a mano, tales como la ocupación espacial, la alineación general, la distribución horizontal de los trazos y la variabilidad en el tamaño de los componentes.

En términos generales, el SVM lineal presentó mejor desempeño global, con una exactitud de 81.25% y el F1 más alto de todos los modelos evaluados. Sin embargo, el SVM con kernel RBF mostró un comportamiento relevante al reducir el número de falsos positivos, es decir, los casos en los que una escritura incorrecta fue clasificada como correcta.

Desde una perspectiva práctica para este problema, este último resultado es importante, ya que, aceptar como correcta una escritura incorrecta es un error más delicado que clasificar como incorrecta una escritura correcta, pues el objetivo es apoyar la detección de posibles dificultades o errores en la escritura. Por lo anterior, aunque el SVM lineal obtuvo mejor exactitud global, el SVM con kernel RBF puede considerarse la mejor alternativa para este caso.

Así mismo, el modelo kNN obtuvo un valor de sensibilidad elevado, pero presentó una baja especificidad, lo que indica que reconoció la mayoría de las escrituras correctas, aunque no acertó en una cantidad importante de escrituras incorrectas como correctas. Random Forest mostró un desempeño intermedio, mientras que el perceptrón multicapa obtuvo el rendimiento más bajo, lo cual, puede estar relacionado con el tamaño del conjunto de datos, ya que las redes neuronales requieren un mayor número de muestras para generalizar adecuadamente.

Limitaciones

Una de las limitaciones de este trabajo es el tamaño del conjunto de datos, compuesto por 80 imágenes de clases relativamente balanceadas. Una base de datos más amplia permitiría evaluar con mayor solidez la capacidad de generalización de los modelos. Además, todas las muestras correspondieron a una misma frase escrita en una sola línea, lo cual favoreció la comparación geométrica entre imágenes, pero al mismo tiempo, limita el alcance del sistema para generalizar ante frases distintas o con un mayor número de palabras.

Otra limitación está en el etiquetado de las muestras como escritura correcta e incorrecta. Aunque esta clasificación permitió entrenar modelos supervisados, en trabajos posteriores sería conveniente fortalecer el proceso de etiquetado mediante la participación de más de un evaluador, ya que esto permitiría reducir la subjetividad.

Por otro lado, las características extraídas se enfocaron principalmente en aspectos geométricos de forma, alineación, separación y tamaño, relacionadas con la evaluación textos conformados por palabras y frases completas. Sin embargo, el sistema no evalúa si las letras están escritas correcta o legiblemente. En consecuencia, este prototipo debe entenderse como una herramienta para evaluar la escritura en personas que ya dominan el trazo correcto de cada una de las grafías necesarias de la escritura.

Trabajo futuro

Como trabajo futuro se recomienda ampliar el conjunto de datos, incorporando más participantes de diferentes grados escolares y una mayor variedad de muestras. Además, se podrían incorporar nuevas características relacionadas con la densidad en el trazo para obtener detalles adicionales sobre la escritura.

Por otro lado, sería útil comparar los modelos supervisados clásicos con modelos de aprendizaje profundo, como las CNN, esto cuando se disponga de un conjunto de imágenes más amplio.

Finalmente, se recomienda realizar un trabajo donde se pueda evaluar el trazo de las diferentes grafías utilizadas en nuestra lengua escrita como un trabajo complementario a este, ya que, con esto, se podría obtener una evaluación completa de la escritura a mano de una persona.

Conclusiones

Se desarrolló un prototipo en MATLAB para clasificar imágenes de escritura a mano mediante procesamiento de imágenes, extracción de características geométricas y comparación de modelos supervisados. El proceso integró binarización mediante el método de Otsu, inversión de la imagen, detección de bordes con Canny, extracción de propiedades mediante *regionprops*, y evaluación cuantitativa de distintos clasificadores.

Los resultados muestran que es posible utilizar características geométricas simples para realizar una clasificación preliminar entre escritura correcta e incorrecta bajo condiciones controladas. El SVM lineal obtuvo mejor desempeño global; sin embargo, el SVM con kernel RBF redujo los falsos positivos, lo cual, para este problema, es útil debido a que se debe evitar clasificar una escritura incorrecta como correcta.

Finalmente, el trabajo evidencia que el procesamiento de imágenes y los modelos supervisados pueden servir como apoyo para analizar rasgos de la escritura a mano, aunque se requiere ampliar el dataset para mejorar la generalización del sistema.

Referencias

Aizerman, M. A., Braverman, E. M., y Rozonoer, L. I. (1964). Theoretical foundations of the potential function method in pattern recognition learning. Automation and Remote Control, 25, 821–837.

Amundson, S. J. (1995). Evaluation Tool of Children's Handwriting (ETCH): Examiner's manual. OT KIDS.

Bin Durayhim, A., Al-Ajlan, A., Al-Turaiki, I., y Altwaijry, N. (2023). Towards accurate children's Arabic handwriting recognition via deep learning. Applied Sciences, 13 (3), Article 1692. https://doi.org/10.3390/app13031692

Boser, B. E., Guyon, I. M., y Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory (pp. 144–152). Association for Computing Machinery. https://doi.org/10.1145/130385.130401

Breiman, L. (2001). Random forests. Machine Learning, 45, 5–32. https://doi.org/10.1023/A:1010933404324

Bublin, M., Santos, L., y Schaer, S. (2023). Online handwriting quality prediction using sensorized pens and recurrent neural networks. IEEE Transactions on Learning Technologies, 16(4), 512–523.

Cortes, C., y Vapnik, V. (1995). Support-vector networks. Machine Learning, 20, 273–297. https://doi.org/10.1007/BF00994018

Cover, T. M., y Hart, P. E. (1967). Nearest neighbor pattern classification. IEEE Transactions on Information Theory, 13(1), 21–27. https://doi.org/10.1109/TIT.1967.1053964

Feder, K. P., y Majnemer, A. (2003). Children's handwriting evaluation tools and their psychometric properties. Physical & Occupational Therapy in Pediatrics, 23(3), 65–84. https://doi.org/10.1080/J006v23n03_05

Hedmi, R., Ahmed, S., y Belaïd, A. (2022). Real-time handwriting evaluation using morphological descriptors and hybrid convolutional networks. Pattern Recognition Letters, 158, 114–121. https://doi.org/10.1016/j.patrec.2022.04.018

Otsu, N. (1979). A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man, and Cybernetics, 9(1), 62–66. https://doi.org/10.1109/TSMC.1979.4310076

Provenzale, A., Di Brina, C., y Rossi, S. (2022). A MATLAB-based semi-automatic framework for analyzing handwriting geometry and spacing in primary school children. Computers &Graphics, 105, 88–97.

Reisman, J. E. (1999). Minnesota Handwriting Assessment: Manual. Psychological Corporation / Harcourt Assessment.

Rumelhart, D. E., Hinton, G. E., y Williams, R. J. (1986). Learning representations by back-propagating errors. Nature, 323, 533–536. https://doi.org/10.1038/323533a0

Secretaría de Educación Pública (SEP) (2023). Acuerdo número 10/09/23 por el que se establecen las normas generales para la evaluación del aprendizaje, acreditación, promoción, regularización y certificación de las alumnas y los alumnos de educación preescolar, primaria y secundaria. Diario Oficial de la Federación.

Sokolova, M., y Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. Information Processing & Management, 45(4), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002

Tappert, C. C., Suen, C. Y., y Wakahara, T. (1990). The state of the art in on-line handwriting recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 12(8), 787–808. https://doi.org/10.1109/34.57669