

## FACIAL EXPRESSIONS RECOGNITION AT THE TEST SCHOOL, BASED ON DEEP LEARNING

Laura-Cleofas Sánchez Autor A., Juan Carlos-Lara Gordillo Autor B., Aldair Brayam-Rivera Castro Autor C.

<sup>A</sup> Tecnológico de Estudios Superiores de Tianguistenco, Carretera Tenango, Santiago-La Marquesa 22, 52650 Santiago Tilapa, laura\_cs@test.edu.mx, Tianguistenco, Estado de México, México Autor A.

<sup>B</sup> Tecnológico de Estudios Superiores de Tianguistenco, Carretera Tenango, Santiago-La Marquesa 22, 52650 Santiago Tilapa, juan\_201823055@test.edu.mx, Autor B.

<sup>C</sup> Tecnológico de Estudios Superiores de Tianguistenco, Carretera Tenango, Santiago-La Marquesa 22, 52650 Santiago Tilapa, aldair\_201826098@test.edu.mx, Autor C.

### Resumen

La identificación temprana de las expresiones faciales que representan las emociones desfavorables de los alumnos puede ser un medio para ayudarlos en situaciones emocionales complejas que pueden afectar su rendimiento académico. Sin embargo, reconocer de forma no automatizada sus emociones en las aulas con 20 o 40 alumnos, puede ser un trabajo complicado para el tutor, quien tiene que identificar factores de riesgo de deserción estudiantil. Una forma eficiente de identificar esas expresiones faciales de las emociones es a través de métodos de reconocimiento de patrones como es el detector de objetos *YOLO v3*, el cual se ha probado ampliamente en datos artificiales. Sin embargo, su rendimiento en la identificación de las expresiones faciales en tiempo real puede cambiar en su eficacia y eficiencia. En el presente trabajo se identifican las expresiones faciales desfavorables en tiempo real de los alumnos del *TEST* que pueden influir en el momento de su aprendizaje en las aulas. Los resultados experimentales mostraron un rendimiento adecuado de la red *YOLO v3* en la curva *ROC* de las expresiones faciales del *TEST*, tales como *afraid*, *disgust*, *happy*, *sad* y *surprise*. No obstante, ésta situación no se presenta en los experimentos de prueba de las imágenes de internet. Para ambos casos, la expresión facial *angry* no se reconoce de manera adecuada en la curva *ROC*.

**Palabras clave** —Aprendizaje Profundo, Expresiones Faciales, *YOLO v3*.

### Abstract

Early identification of facial expressions that represent the unfavorable emotions of the students can be a means to help in complex emotional situations that can affect their academic performance. However, recognizing their emotions in a non-automated way in classrooms with 20 or 40 students can be a complicated job for the tutor, who must determine risk factors for student dropout. An efficient way to recognize those facial expressions of emotions is through pattern recognition methods such as the *YOLO v3* object detector, which has been tested on artificial data. However, its effectiveness and efficiency may change as it identifies facial expressions in real-time. The present work identifies in real-time the

unfavorable facial expressions of *TEST* students that can influence their learning in the classroom. The experimental results showed an adequate performance of the *YOLO v3* network in the *ROC* curve of the facial expressions of the *TEST*. This included fear, disgust, happiness, sadness, and surprise. However, this situation does not occur in test experiments on internet images. In both cases, the angry facial expression is not adequately recognized by the *ROC* curve.

**Keywords**— *Deep Learning*, *Expression Facial*, *YOLO v3*.

## 1. INTRODUCCIÓN

La importancia de las expresiones no verbales ha tenido gran relevancia en los seres humanos, ya que es una de las acciones para transmitir o recibir gran cantidad de información no verbal que es expresada en un 55% de la comunicación realizada por el ser humano, quien es plenamente sociable al intercambiar información en su vida diaria. Dentro de la comunicación no verbal se pueden considerar las expresiones faciales, las gesticulares realizadas mediante las manos y los brazos, las del lenguaje corporal que se llevan a cabo a través de movimientos del cuerpo, entre otras expresiones no verbales. Asimismo, las expresiones faciales exhiben el estado emocional de los seres humanos, las cuales se realizan mediante 12 músculos faciales superiores y 18 músculos faciales inferiores [1].

En el ámbito escolar, las expresiones faciales de los alumnos pueden cambiar debido a la presión y al estrés en el aula. Desde hace algunas décadas se ha estudiado el comportamiento del alumnado en las aulas, descubriendo que la variable más importante para que un alumno pueda aprender mejor es la autoestima o la confianza al aprender sus materias [17].

Interpretar las expresiones faciales de un estudiante manualmente es cansado y tardado, sin mencionar que en una aula se pueden tener aproximadamente de 20 a 40 alumnos, por ello la implementación de una red neuronal convolucional resulta de gran ayuda, ya que puede reconocer rápidamente las expresiones faciales que tiene un estudiante, desde 45 a 150 fotogramas por segundo (*FPS*) lo que significa que se puede procesar la transmisión de vídeo en tiempo real con menos de 25 milisegundos.

Las redes neuronales artificiales se inspiran en la manera en cómo el ser humano aprende, analiza y reconoce las tareas cotidianas mediante el cerebro humano quien se encarga de realizar las interconexiones entre las neuronas, lo cual ayuda a los seres humanos a identificar diferentes tareas y comportamientos de su entorno. De igual manera, las redes neuronales convolucionales (por sus siglas en inglés, *CNN*) imitan el comportamiento del ser humano mediante modelos matemáticos que discriminan entre diversas clases, a partir de un gran cúmulo de información.

En los estudios realizados por [13], [14], [15] y [16], se muestra que *You only live once (YOLO)* es una de las *CNN* que realiza su detección de objetos de manera rápida considerando para ello 22, 29 y 51.9 segundos dependiendo de la estructura de la red, sin perder la esencia competitiva de su predicción (51.5, 55.3 y 57.9) de los objetos en comparación con otros modelos de detección de objetos tales como *SSD321*, *DSSD321*, *R-FCN*, *SSD513*, *DSSD513*, *FPN FRCN*, *RetinaNet-50-500*, *RetinaNet-101-500*, *RetinaNet-101-800*, *YOLOv3-320*, *YOLOv3-416* y *YOLOv3-608*.

Tabla 1. Comparación de detectores de objetos [13], [14], [15], [16], [18] y [19].

Autores, año	Detector	mAP-50	ms
W. Liu et al., 2016	SSD321	45.4	61
C.-Y. Fu et al., 2016	DSSD321	46.1	85
J. Dai et al., 2016	R-FCN	51.9	85
W. Liu et al., 2016	SSD513	50.4	125
C.-Y. Fu et al., 2016	DSSD513	53.3	156
S. Ren et al., 2016	FPN FRCN	59.1	172
Tsung-Yi Lin., 2018	RetinaNet-50-500	50.9	73
Tsung-Yi Lin., 2018	RetinaNet-101-500	53.1	90
Tsung-Yi Lin., 2018	RetinaNet-101-800	57.5	198
J. Redmon et al., 2018	YOLOv3-320	51.5	22
J. Redmon et al., 2018	YOLOv3-416	55.3	29
J. Redmon et al., 2018	YOLOv3-608	57.9	51

Algunos de los trabajos realizados utilizando algoritmos de inteligencia artificial, en específico en visión por computadora, se han enfocado en estudiar las expresiones de las emociones con herramientas de aprendizaje profundo, tales como las que corresponden a las *CNN* y a la red llamada *belief network* [2].

La necesidad de identificar un gran cúmulo de información no verbal procedente de la sociedad nos ha llevado a estudiar las expresiones faciales mediante herramientas de inteligencia artificial con el objetivo de identificar las emociones desfavorables que puedan afectar a los alumnos del Tecnológico de Estudios Superiores de Tianguistengo (*TEST*) en el rendimiento de sus estudios en las aulas.

El presente trabajo propone la identificación automática en tiempo real de las expresiones faciales que determinan el estado de ánimo de los alumnos del *TEST* cuando ellos están aprendiendo sus materias en las aulas. La identificación de las expresiones faciales puede ayudar a los expertos de psicología en canalizar a los estudiantes en el área correspondiente.,

donde se les puede dar atención adecuada de posibles problemas emocionales que afecten el desempeño académico de los estudiantes. Los resultados experimentales mostraron un reconocimiento adecuado de las expresiones faciales con las imágenes de prueba del *TEST* en comparación con las imágenes de prueba de internet, menos en la expresión facial que representa la emoción *angry*.

El presente trabajo está organizado de la siguiente manera., en la sección 2 se describen los trabajos relacionados, en la sección 3 la metodología propuesta, en la sección 4 los experimentos realizados, en la sección 5 las conclusiones y recomendaciones, en la sección 6 los agradecimientos, y finalmente en la sección 7 las referencias bibliográficas.

## 2. TRABAJOS RELACIONADOS

Las expresiones faciales son un aspecto importante de la comunicación diaria del ser humano, dentro de las expresiones no verbales, cabe mencionar que el 55% de ellas corresponden a las expresiones faciales que pueden determinar las emociones del ser humano mediante diversos músculos faciales que son de gran utilidad en el reconocimiento de patrones. Algunas de las aplicaciones de las expresiones faciales se pueden mencionar de forma general, por ejemplo, la detección de tipos de comportamientos tales como la depresión, la ansiedad, los trastornos mentales, entre otros [3], [4].

Gallardo *et al.*, realizaron un estudio de la eficiencia terminal de la escuela de nivel medio superior de Celaya para determinar las causas de deserción educativa, en donde se encontraron como factores que influyen en el desempeño académico, el índice de reprobación, la tutoría, los criterios de evaluar y la forma de impartir clases por parte del docente, así como las unidades de aprendizaje. Los resultados obtenidos de 227 encuestas realizadas a los alumnos, exhiben que fue posible determinar soluciones por parte de los expertos para menguar el retraso de la eficiencia terminal académica [3].

Singh *et al.*, expresan que el reconocimiento facial ha llamado mucho la atención en el ámbito de visión por computadora., el cual ha sido ampliamente implementado en varios ámbitos como por ejemplo la diversión, entretenimiento, negocios en línea, seguridad, comprensión de sentimientos para diagnósticos médicos, entre otros. Ellos proponen el reconocimiento de emociones mediante información facial de la web, considerando aspectos como alegría, asombro, entre otros. Para lo cual hicieron uso de la herramienta de *Tensor Flow* y *OpenCv* [5].

En el trabajo de Refat *et al.*, se mencionan que dentro de los métodos para realizar el reconocimiento de las expresiones faciales existen las *CNN* tales como *deep belief network* y *deep learning*, los autores se enfocaron en los algoritmos de *DL* para realizar el reconocimiento de expresión es faciales de personas del sexo femenino del país de Japón. Es relevante

mencionar que con las *CNN* alcanzaron un reconocimiento de las expresiones faciales del 97.01 % [2].

Ng *et. al.*, realizaron el reconocimiento supervisado de emociones en base de datos pequeños mediante un aprendizaje llamado *transfer* basado en *CNN* [4]. Los resultados mostraron que mediante el método llamado *cascading fine-tuning* obtuvieron los mejores resultados en comparación con el método *single fine-tuning*. El mejor resultado exhibe un porcentaje del 55.6 % [4].

En el trabajo realizado por Al-Darraj *et. al.*, muestra la identificación de cambios sutiles en expresiones faciales, el trabajo fue realizado mediante herramientas de aprendizaje profundo, los experimentos realizados mostraron un rendimiento del 90 % [6].

En el trabajo realizado por Saeed *et. al.*, se propone el reconocimiento automatizado de las expresiones faciales (miedo, disgusto, neutral, feliz, cansado, ira y sorpresa) en hombres y mujeres utilizando una red neuronal convolucional, considerando para ello cuatro capas convolucionales y dos capas ocultas. El rendimiento del reconocimiento de expresiones faciales fue del 94 %, no obstante, para validar el algoritmo propuesto se consideró el método de *10-fold cross-validation*, donde se obtuvieron las métricas de sensibilidad y especificidad, obteniendo valores en cada una de ellas del 94.02 % y 99.14 %, respectivamente. Adicionalmente, se obtuvieron otros resultados como la medida *f1*, y la precisión para validar la calidad del modelo, obteniendo de cada una de ellas el 84.07 %, 78.22 % y 94.09 %, respectivamente [7].

En el artículo realizado por Akhand, *et al.*, proponen el reconocimiento de emociones faciales realizando un aprendizaje *Transfer* en ocho redes neuronales profundas de la base de datos *KDEF* y *JAFFE* que contienen imágenes de datos faciales. Los resultados se realizaron con *10-fold cross validation*, la red que obtuvo mejor rendimiento fue la *DenseNet-161* en las bases de datos *KDEF* y *JAFFE*, con un rendimiento del 96.51 % y 99.52 %, respectivamente [8].

En el trabajo realizado por Aswathy, se menciona que existen varias técnicas para el reconocimiento de expresiones tales como *Linear Discriminant Analysis*, *Independent Component Analysis*, entre otros. Sin embargo, el autor menciona que es posible obtener un mejor reconocimiento de las expresiones faciales, tomando en cuenta algunas propiedades de las imágenes tales como su intensidad del color, el ángulo en el que se tome la imagen y la luz con la que se proyecta [9], entre otros aspectos.

Shehu *et. al.*, proponen utilizar el clasificador *Haar Cascade* y la red *ResNet* para analizar las expresiones faciales de tres bases de datos llamadas *CK+*, *KDEF* y *RAF*. Adicionalmente ellos utilizaron como método el preprocesamiento *Label Smoothing* para mejorar la calidad de las imágenes. Con el preprocesamiento de los datos, se mostró una mejora en el rendimiento considerando las bases de datos *CK+* y *KDEF*. Por otra parte, el algoritmo *Haar Cascade* muestra un

decremento en su rendimiento con la base de datos *KDEF* [10].

Barros *et. al.*, exhibe un modelo *neuro-computacional* que reconoce las expresiones emocionales no etiquetadas tales como expresiones faciales, de movimientos de cuerpo y neutrales. Los filtros del modelo convolucional propuesto por los autores, ayudan a localizar las expresiones emocionales, considerando para ello métodos de probabilidades de los filtros de la red que indican el lugar donde se encuentra la expresión dentro de la imagen [11].

### 3. METODOLOGÍA PROPUESTA

Dentro de las expresiones no verbales del ser humano se encuentran las expresiones faciales que pueden determinar las emociones de los seres humanos. Las expresiones faciales son los movimientos gesticulares de la cara, dichos movimientos son indicadores de las emociones que tiene cada persona. En el presente trabajo se reconocieron las expresiones faciales en tiempo real de los alumnos del *TEST* con la red neuronal *YOLO* para identificar emociones desfavorables que puedan afectar en el aprendizaje de los alumnos. Para la presente investigación se utilizaron 6 expresiones faciales tales como *afraid*, *disgust*, *happy*, *sad*, *surprize* y *angry.*, dichas expresiones son las más comunes que pueden reflejar las personas, esto es explicado por varios psicólogos en diferentes artículos, en especial Paul Ekman [12], quien fue un pionero de la investigación del reconocimiento de las expresiones faciales.

Para los experimentos, se obtuvieron 2500 imágenes de las expresiones faciales a través de la web, considerando para el reentrenamiento el 80 % de los datos y el 20 % para la evaluación de la red. El reentrenamiento de la red se realizó una sola vez. Las imágenes se etiquetaron con *Labelimg*, para lo cual se instaló *anaconda.*, dentro de *anaconda* se realizó un entorno de trabajo donde se instaló *Python* para poder etiquetar las imágenes con la herramienta *Labelimg*.

Para probar el modelo *YOLO*, se consideraron 500 imágenes de las expresiones faciales obtenidas de los estudiantes del *TEST* y 500 imágenes adquiridas de internet., cada una de ellas se tomaron para obtener el correspondiente video. La resolución de las imágenes fue de 416 \* 416 pixeles, considerando para ellas un formato \*.jpeg.

Para la creación del modelo de las expresiones faciales que reflejan las emociones, *YOLO* obtuvo su primer aprendizaje con los pesos predeterminados, en donde se aprendieron aproximadamente 80 clases (por ejemplo, *person*, *bicycle*, *motorbike*, entre otras), posteriormente se realiza el reentrenamiento de la red con 2500 imágenes de las emociones obtenidas de internet donde las clases se encuentra equilibradas, el modelo de *YOLO* se entrenó en google colab basado en las librerías tales como *TensorFlow*, *Keras*, *opencv*, *numpy*, *tensorboard*, *matplotlib*, *terminaltables*, *pillow*, *torchvision*, *cuda toolkit pytorch*, entre otras. Dichas librerías están escritas en el lenguaje Python. El

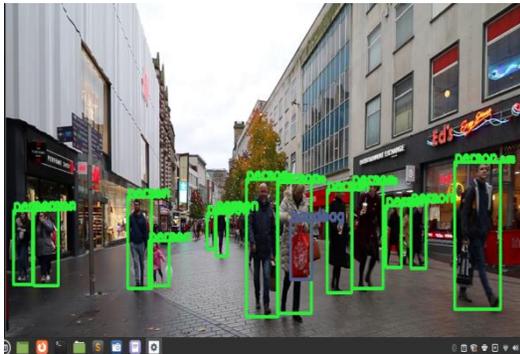
entrenamiento de la red *YOLO* se realizó en *google colab*, y el tiempo de entrenamiento fue de dos horas y 25 minutos. Las épocas constaron de 100 para el reentrenamiento, en cada una de ellas la función de pérdida disminuye hasta alcanzar un valor de 1.25, lo cual indica que el aprendizaje de la red *YOLO* incrementa. Asimismo, la identificación de los objetos fue del 98.74 %. Aunque el número de *max\_batches* se puede calcular con el número de las clases por 2000., por defecto la red *YOLO* consideró el número de *max\_batches* igual a 500,200.

El aprendizaje del modelo *YOLO* de cada expresión facial se puede observar en las figuras de las curvas *ROC* en la sección de experimentos., mediante las curvas *ROC* es posible analizar el reconocimiento de las expresiones faciales en tiempo real que representan las emociones de los estudiantes del *TEST* en el aula de aprendizaje., el reconocimiento de las expresiones faciales en tiempo real de los alumnos del *TEST* se compara con los resultados de las expresiones obtenidas de internet.

#### 4. EXPERIMENTOS

En la figura 1 se puede observar que en el primer experimento, con el aprendizaje de los pesos predeterminados de red, se reconocen 2 clases dentro de las 80 clases., tales como la clase persona y perro, de un video que se escogió de manera aleatoria, la primera clase se puede observar en recuadros de color verde, y la segunda clase se muestra en recuadros de color morado.

Figura 1.- Prueba del aprendizaje de los pesos predeterminados de *YOLO* de un video aleatorio



En la figura 2 se puede observar la identificación de la clase *angry*, la cual se encuentra dentro de un recuadro de color coral. Para ello se consideró el aprendizaje obtenido en el reentrenamiento de la red *YOLO*.

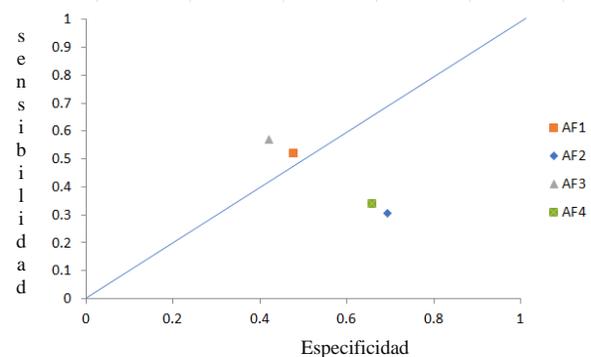
Figura 2.- Experimento en tiempo real del reconocimiento de la clase *angry* de los alumnos del *TEST*, considerando el reentrenamiento de la red *YOLO*.



Para los resultados experimentales se consideró el aprendizaje del reentrenamiento de la red *YOLO*, donde se pueden observar seis gráficas, desde la Figura 3 hasta la Figura 8, de la curva *ROC* del rendimiento del modelo de las expresiones faciales *afraid* (*AF*), *angry* (*AN*), *disgust* (*DI*), *happy* (*HA*), *sad* (*SA*) y *surprise* (*SU*)., a cada una de las siglas de las emociones se adjuntan números desde el 1 hasta el 4., los números 1 (primer conjunto de 250 imágenes) y 2 (segundo conjunto de 250 imágenes) son para representar los experimentos con las imágenes obtenidas de los alumnos del *TEST*, y los números 3 (tercer conjunto de 250 imágenes) y 4 (cuarto conjunto de 250 imágenes) indican los experimentos de las imágenes obtenidas de internet., considerando un total de 1000 imágenes de para probar el modelo *YOLO*.

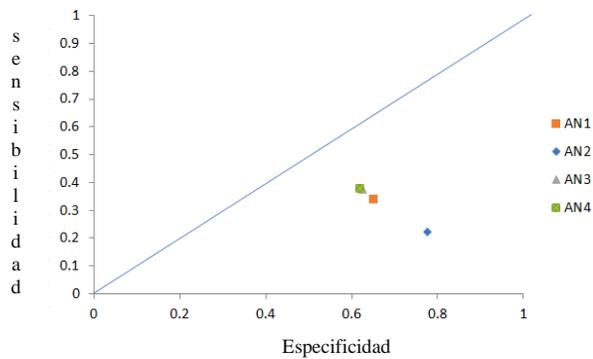
En la Figura 3 podemos observar un reconocimiento aceptable de primer y tercer conjuntos de datos de la clase *afraid*., el resto de ellos tiene un rendimiento pobre.

Figura 3.- Curva *ROC* de los resultados del reconocimiento de la clase *afraid*.



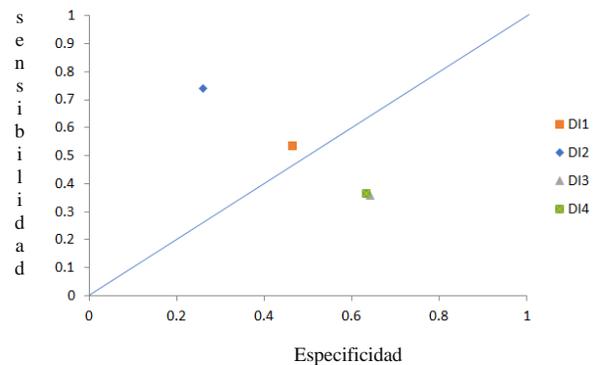
En la Figura 4 podemos observar un reconocimiento pobre en todos los conjuntos de datos de la clase *angry*.

Figura 4.- Curva *ROC* de los resultados del reentrenamiento de la red *YOLO* de la clase *angry*.



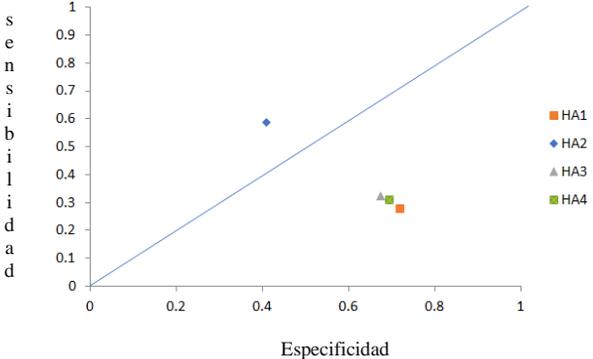
En la Figura 5 podemos observar un reconocimiento aceptable en el primer y segundo conjunto de datos de la clase *disgust*. Sin embargo, en el resto de los conjuntos la identificación de la clase *disgust* es pobre.

Figura 5.- Curva ROC de los resultados del reentrenamiento de la red YOLO de la clase *disgust*.



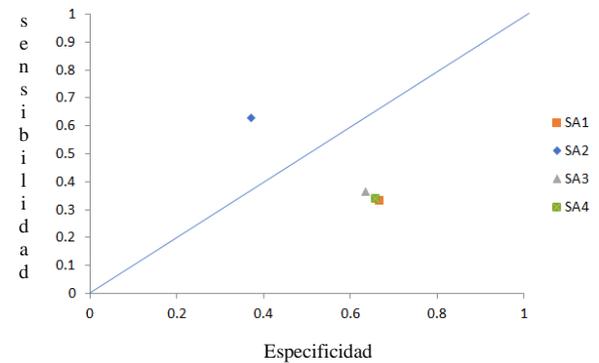
En la Figura 6 podemos observar un reconocimiento aceptable de la clase *happy*, del segundo conjunto de datos. Sin embargo, en el resto de los conjuntos la identificación de la clase *happy* es pobre.

Figura 6.- Curva ROC de los resultados del reentrenamiento de la red YOLO de la clase *happy*.



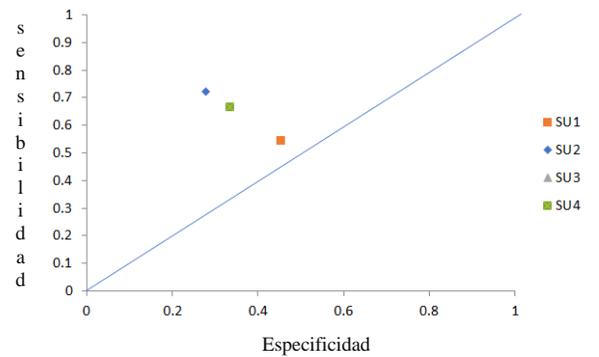
En la Figura 7 podemos observar un reconocimiento aceptable de la clase *sad*, del segundo conjunto de datos. Sin embargo, la identificación de la clase *sad* es pobre, en el resto de los conjuntos.

Figura 7.- Curva ROC de los resultados del reentrenamiento de la red YOLO de la clase *sad*.



En la Figura 8 podemos observar que los conjuntos de prueba obtienen un reconocimiento aceptable en todos los conjuntos de datos de la identificación de la clase *surprise*., es notable observar que en el tercer y cuarto conjuntos de datos se obtiene la misma identificación de la clase *surprise*.

Figura 8.- Curva ROC de los resultados del reentrenamiento de la red YOLO de la clase *surprise*.



## 5. CONCLUSIONES Y RECOMENDACIONES

En el presente trabajo se identificaron en tiempo real con YOLO v3 las expresiones faciales que reflejan las emociones de los alumnos del TEST que pueden influir en el momento de su aprendizaje., se realizó una comparativa con conjuntos de prueba obtenidos de internet. En los resultados experimentales de la expresión *afraid*, se puede observar un reconocimiento aceptable de los conjuntos de datos de prueba de las imágenes de las expresiones faciales del TEST, no obstante los conjuntos de prueba de las expresiones *afraid* de internet tuvieron un reconocimiento pobre (Figura 3). Con respecto a la expresión facial de la emoción *angry*, no se observó un reconocimiento aceptable en la curva ROC, ya que los resultados se muestran debajo de la línea de no discriminación (Figura 4). Del reconocimiento de la expresión *disgust* se puede observar favorable en los conjuntos de datos de las imágenes del TEST, sin embargo esta situación no es posible obtenerla en los conjuntos de prueba de las imágenes de internet (Figura 5).

Se observó un reconocimiento aceptable de la expresión *happy* y *sad* del conjunto de datos 2 de las imágenes de TEST, no obstante el reconocimiento del resto de los conjuntos de datos es pobre, ya que se encuentran debajo de la línea de no

discriminación de la curva *ROC* (Figura 6). En el reconocimiento de la expresión *suprise* se observó que en los conjuntos de datos del *TEST* y de internet, se reconoció la clase de manera aceptable, ya que los resultados del reconocimiento se encuentran por encima de la línea de no discriminación de la curva *ROC*. Es necesario realizar más experimentos y probar con otros modelos para determinar estadísticamente el reconocimiento predictivo del modelo *YOLO v3* en la problemática del reconocimiento de expresiones faciales de las emociones que emiten los alumnos del *TEST* en las aulas de aprendizaje.

## 6. AGRADECIMIENTOS

En el presente trabajo se agradece el apoyo otorgado por el Tecnológico de Estudios Superiores de Tianguistenco y de COMECYT.

## 7. REFERENCIAS

- [1] J. A. Hall, T. G. Horgan, N. A. Murphy. “Nonverbal communication”. *Annual review of psychology*. Vol. 70, pp. 271–294, 2019.
- [2] C. M. M. Refat, N. Z. Azlan. “Deep learning methods for facial expression recognition”. *7th International Conference on Mechatronics Engineering (ICOM), IEEE*, pp. 1–6. 2019.
- [3] F. A. Gallardo León, M. S. Gutiérrez Chávez, S. Espino Mandujano, E. A. Arreola Ramírez. “Factores que inciden en la eficiencia terminal del alumno en la escuela de nivel medio superior de Celaya”. *Revista de Divulgación en la Ciencia ISSN 2395-9797, Jóvenes en la ciencia*. 2019.
- [4] H. W. Ng, V. D. Nguyen, V. Vonikakis, S. Winkler. “Deep learning for emotion recognition on small datasets using transfer learning”. In *Proceedings of the 2015 ACM on international conference on multimodal interaction*, Association for computing Machinery, pp. 443–449. 2015.
- [5] A. Singh, A. P. Srivastav, P. Choudhary, S. Raj. “Facial emotion recognition using convolutional neural network”. *2nd International Conference on Intelligent Engineering and Management (ICIEM), IEEE Xplore*, pp. 486–490. 2021.
- [6] S. Al-Darraj, K. Berns, A. Rodić. “Action Unit Based Facial Expression Recognition Using Deep Learning”. In: Rodić, A., Borangiu, T. (eds) *Advances in Robot Design and Intelligent Control*. RAAD 2016. *Advances in Intelligent Systems and Computing*, vol 540. Springer, Cham. pp. 413–420. 2017.
- [7] S. Saeed, A. A. Shah, M. K. Ehsan, M. R. Amirzada, A. Mahmood, T. Mezgebo. “Automated facial expression recognition framework using deep learning”. *Journal of Healthcare Engineering*, Vol. 2022. 2022.
- [8] M. A. H. Akhand, S. Roy, N. Siddique, A. S. Kamal, T. Shimamura. “Emotion recognition using transfer learning in the deep cnn”. *Electronics*, Vol. 10, pp. 2–19. 2021.
- [9] R. Ms. Aswathy. “A literature review on facial expression recognition techniques”. *IOSR Journal of Computer Engineering*, Vol. 11, pp. 61–64. 2016.
- [10] H. A. Shehu, M. H. Sharif, S. Uyaver, “Facial expression recognition using deep learning”. In *AIP Conference Proceedings*, AIP Publishing. 2021.
- [11] P. Barros, G. I. Parisi, C. Weber, S. Wermter. “Emotion-modulated attention improves expression recognition: A deep learning model”. *Neurocomputing*, Vol. 253, pp. 104–114. 2017.
- [12] P. Ekman. “El rostro de las emociones”. RBA. ISBN: 9788492966981. 2017.
- [13] S. Ren, K. He, R. Girshick, J. Sun. “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks”. arXiv: 1506.01497. *Computer Science, Computer Vision and Pattern Recognition*. 2016.
- [14] J. Dai, Y. Li, K. He, J. Sun. “R-FCN: Object Detection via Region-based Fully Convolutional Networks”. *Advances in Neural Information Processing Systems*. Vol. 29. 2016.
- [15] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, Cheng-Yang Fu, A. C. Berg. “SSD: Single Shot MultiBox Detector”. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds) *Computer Vision – ECCV 2016*. ECCV 2016. *Lecture Notes in Computer Science*, vol 9905. Springer, Cham. 2016.
- [16] J. Redmon, A. Farhadi. “YOLOv3: An Incremental Improvement”. arXiv:1804.02767. *Computer Science, Computer Vision and Pattern Recognition*. 2018.
- [17] M. L. Ariza-Hernández. “Influencia de la inteligencia emocional y los afectos en la relación maestro-alumno, en el rendimiento académico de estudiantes de educación superior”. *Educación y educadores*. 20(2). 2017.
- [18] Tsung-Yi Lin, P. Goyal, R. Girshick, K. He, P. Dollar. “Focal Loss for Dense Object Detection”. arXiv:1708.02002v2. 2018.
- [19] C.-Y. Fu, W. Liu, A. Ranga, A. Tyagi, and A. C. Berg. “DSSD: Deconvolutional single shot detector”. arXiv:1701.06659, 2016.