

Multi-layer Adaptive Fuzzy Inference System for Predicting Student Performance in Online Higher Education

Rosa Leonor Ulloa-Cazarez, Noel García-Díaz y Leonel Soriano-Equigua

Abstract— Research on student performance prediction has evolved from the early application of statistical techniques to later use of computational techniques. Results in this field are varied, thus, we have to take advantage of previous research results. This study proposes a Multi-layer Adaptive Neuro-Fuzzy Inference System (MANFIS) for student performance prediction in online Higher Education settings. The MANFIS was trained and tested using a dataset integrated by the scores obtained by students in four online Higher Education courses. The MANFIS prediction accuracy was compared against the accuracies of Multilayer neural network, Radial Basis Function Neural Network, and General Regression Neural Network. The accuracy of the MANFIS prediction statistically outperformed at least one neural network (out of three possible) in each dataset. The Results indicate that MANFIS is an alternative model to predict student performance in online Higher Education settings.

Index Terms— Computational Intelligence, Fuzzy Systems, Hybrid Intelligent Systems, Neural Networks, Multi-layer Adaptive Fuzzy Inference System, student performance prediction, online education, e-learning

I. INTRODUCCIÓN

La predicción del desempeño del estudiante (DE) en programas de educación superior en línea (ESL), es una estrategia que ha sido utilizada para identificar tempranamente a los estudiantes en riesgo [1, 2] y abatir los índices de reprobación [3], ambos señalados como problemas importantes de la educación en línea [4].

A pesar de que se ha avanzado en la aplicación de técnicas de aprendizaje de máquina y minería de datos en la educación con propósitos predictivos [1, 5], el campo de estudio mantiene varias aristas de exploración [6, 7] sugiriéndose la construcción de modelos más explicativos del fenómeno, pero también con mayor precisión en la predicción [8].

Se han encontrado estudios que en su mayoría realizan predicciones en términos de clasificación o agrupamiento de los desempeños [5, 9, 10, 11, 12] y algunos pocos que realizan la predicción numérica de la calificación final [13, 14, 15, 16], ambos enfoques con una aplicación diferente.

En los trabajos donde la predicción es en términos numéricos, usualmente se comparan las precisiones de los modelos propuestos contra modelos estadísticos [2, 3, 8, 17] y también frente a modelos computacionales [2, 3, 17, 18, 19] sin embargo, la variedad de modelos y resultados encontrados, sugieren que es necesario continuar esta línea de indagación a fin de fortalecer los resultados logrados [20].

Algunas diferencias en los modelos predictivos son por ejemplo, el uso de diversas variables independientes, como las

relacionadas con la información demográfica del estudiante [14], la personalidad [21] y otras variables psicológicas [22], los reportes de navegación y de eventos obtenidos de las plataformas de gestión del aprendizaje [4], así como las calificaciones obtenidas por los estudiantes en diversos productos de aprendizaje [5, 23].

Dado que en cursos de ESL el DE generalmente es representado de manera cuantitativa a través de una calificación que sintetiza el esfuerzo y progreso del estudiante [24], la mayoría de los estudios con propósitos predictivos utiliza como variable dependiente la calificación final mesurada ya sea como el promedio de diversos productos de aprendizaje, o como la calificación en un examen final [4, 18, 25].

En un artículo reciente [26] se utilizaron las calificaciones obtenidas tempranamente por los estudiantes como variable independiente, y las calificaciones finales como variable dependiente, para crear tres redes neuronales (RN) cuya precisión predictiva superó de manera estadísticamente significativa a la de la regresión estadística.

Asimismo, tras un análisis estadístico se concluye que las calificaciones obtenidas por los estudiantes dentro del primer período lectivo (de cuatro) del calendario escolar, pueden ser utilizadas para generar un modelo de predicción que permite la identificación oportuna de los estudiantes con altas posibilidades de reprobación [26, 16].

El continuar replicando los análisis realizados para generar nuevos modelos y optimizarlos, es una práctica deseable y característica de la ciencia de datos [27] así como una estrategia que contribuye a la confiabilidad de los resultados con evidencias y criterios científicamente planteados [20], y dado que se mantiene la necesidad de mejorar la precisión de los modelos predictivos [8], el objetivo de este trabajo es el de aplicar un Sistema Adaptativo Multicapa de Inferencia Neurodifusa (SAMIND) para la predicción del DE en ESL, que mejore en precisión predictiva a modelos generados en un estudio previo [26] utilizando los mismos conjuntos de datos.

Nuestro experimento actual utilizando los mismos datos que Ulloa Cazarez y López Martín [26], contribuye con la ciencia al validar y aumentar la confiabilidad de los datos, también, al explorar el uso de modelos híbridos en el área de la gestión educativa, y en particular contribuimos con el área de la predicción al generar modelos que superan a los anteriores publicados en precisión.

En este estudio, un SAMIND es un sistema difuso que se acopla a una RN para mejorar sus características, por ejemplo, la flexibilidad y la adaptabilidad de la red [28]. Por tanto, un

SAMIND es una RN optimizada con lógica difusa.

Para el entrenamiento y evaluación del SAMIND y de los modelos a comparar, se utilizaron las calificaciones de cuatro materias del primer semestre de un programa de ESL relacionado con las tecnologías de la información [29]. Las calificaciones se colectaron en dos momentos, el primero, al finalizar el primer mes del semestre, colectando las calificaciones de la primera unidad de estudio (U1), el segundo momento, al finalizar el mismo semestre, colectando las calificaciones finales (CF). Las calificaciones U1 de cada curso fueron usadas como variables independientes y las CF de los mismos cursos fueron las variables dependientes.

La hipótesis a probar es la siguiente:

H_0 No hay diferencia estadísticamente significativa entre la precisión de la predicción del SAMIND y la de una red neuronal cuando ambos modelos son entrenados y probados para predecir la calificación final de un estudiante de educación superior en línea

De comprobarse que existe una diferencia estadísticamente significativa entre las precisiones de los modelos, indicaría que alguno de ellos ofrece una mayor precisión en la predicción por lo que como hipótesis alterna se formula la siguiente:

H_a Existe diferencia estadísticamente significativa entre la precisión de la predicción del SAMIND y la de una red neuronal cuando ambos modelos son entrenados y probados para predecir la calificación final de un estudiante de educación superior en línea

La estructura de este trabajo se distribuye en las siguientes 6 secciones: la sección II presenta los trabajos relacionados con la predicción del DE y el uso de los Sistemas Neurodifusos (SND) en este campo; la sección III introduce la teoría que sustenta el SAMIND que generamos y que describimos en la sección IV; la sección V presenta la metodología seguida para la comparación de los modelos; la sección VI presenta los resultados mientras que su discusión se desarrolla en la sección VII, junto con las limitaciones y el trabajo futuro.

II. TRABAJO RELACIONADO

A. La Predicción del DE

La predicción del DE en la literatura científica se ha documentado como una práctica prometedora [6] para abatir los altos índices de reprobación que observa la modalidad [2, 3] y se han aplicado diversos modelos computacionales que incluyen máquinas de soporte vectorial [3, 30], diversas RN [2, 15], RN optimizadas [6] y programación genética [16].

Las variables utilizadas en los trabajos revisados varían entre sí, e incluyen aspectos psicológicos, de personalidad [31], datos demográficos de los estudiantes [20], sus registros académicos previos [2], registros de actividad en las plataformas de gestión del aprendizaje [30, 32], la participación [16] y las calificaciones logradas en actividades de aprendizaje [15].

B. Uso de Sistemas Neurodifusos en la Predicción del DE

Los sistemas de inferencia difusa (SID) se han utilizado poco en educación, en su mayoría con buenos resultados [12, 33] donde se observa la creación de sistemas híbridos a partir de SID, con dos aristas principales [34]: una, generando un SID

equipado con algún algoritmo de optimización para mejorar sus capacidades de inferencia, y otra, utilizando el SID para mejorar otro sistema, como por ejemplo una RN [6].

Del primer tipo de hibridación encontramos un estudio entre cinco, que realiza predicciones con una salida numérica [17], y uno en términos de agrupamiento o clasificación [31], donde se aplicó un SID que fue optimizado por una RN para predecir el DE, aunque el propósito principal de los autores era desarrollar un sistema de evaluación del DE. Este tipo de sistemas presentan una oportunidad de aplicación en el campo educativo y prometen resultados en cuanto a predicción numérica. Del segundo tipo de hibridación se encontró un solo trabajo en el campo [6] con salida en términos de agrupamiento.

Las variables utilizadas en estos trabajos corresponden en su mayoría a distintos formatos de calificaciones, por ejemplo, Chen y Do [17] utilizaron un conjunto de datos compuesto por las calificaciones obtenidas por los estudiantes en diversos exámenes previos al curso de análisis, e información demográfica, constituyendo en total ocho variables de entrada. En Altaher y BaRukab [30] se utilizaron 5 variables de entrada correspondientes a las calificaciones en las materias cursadas un semestre anterior mientras que en Quang Hung y Jeng-Fung [10] se utilizaron las calificaciones de los estudiantes, obtenidas en exámenes y diversas actividades. En la Tabla I se describen éstos modelos.

TABLA I
DESCRIPCIÓN DE MODELOS SID

(Est: Estudio; SND: Tipo de sistema Neurodifuso creado; MO: Mecanismo de optimización; RE: Raíz del error cuadrático medio; EP: El error promedio absoluto; MS: El error absoluto medio; CP: Criterio de Precisión)

Est	SND	MO	CP
[19]	Sistema Jerárquico Adaptativo de Inferencia Neurodifusa	Algoritmo de búsqueda de Cuckoo Algoritmos genéticos	RE EP MS
[34]	Sistema de Inferencia Neurodifusa con múltiples entradas y una salida	Red Neuronal de Retro-propagación	MS
[33]	SAIND	Se Modificaron las funciones de membresía: gaussiana, triangular y de campana generalizada	RE
[12]	SAIND	Se Modificaron las funciones de membresía, concluyendo en una Gaussiana	MS
[8]	Red Neuronal	Algoritmo de optimización de Cuckoo (basado en)	RE EP

Los trabajos en la Tabla I, observan el uso de diferentes tipos de datos, notando que requirieron de registros históricos, y también, se diferencian en la variedad de métodos para obtener los parámetros del SID, encontrando métodos de exploración, empíricos o algoritmos de aprendizaje [35]. Estas diferencias contribuyen a una gran variedad de resultados y aplicaciones, así como a la dificultad de no poder llegar a conclusiones sobre uno u otro modelo aplicado, por lo que se justifica la continuación de esta línea de investigación para construir conocimiento a partir del ya logrado, mediante la replicación de estudios que generen impacto en el ámbito educativo y en el de los algoritmos computacionales para la predicción.

III. SISTEMA ADAPTATIVO MULTICAPA DE INFERENCIA NEURODIFUSA

A. Sistemas de Inferencia Difusos

La lógica difusa ha sido utilizada en diversos campos para la optimización de algoritmos [36] ya que proporciona un método conveniente para mapear los dominios de entrada y salida debido a la forma natural en que las reglas difusas son expresadas (Si X entonces Y) proporcionando así, un método [32].

Existen varios tipos de SID, uno de ellos, el Takagi-Sugeno [37], se utiliza para modelar cualquier sistema de inferencia en el que la salida sea continua o constante, en lugar de una proposición difusa [38].

Las reglas difusas tienen la forma siguiente:

$$R_i: \text{IF } x_1 \text{ is } A_1^i \text{ and } \dots \text{ and } x_n \text{ is } A_n^i \text{ THEN } y_1 = p_1 x_1 + q_1 x_2 + r_1 \quad (1)$$

Donde:

- x_1 y hasta x_n son las entradas al sistema;
- A_1^i y hasta A_n^i son las etiquetas lingüísticas;
- p_1 , q_1 y r_1 son los parámetros del consecuente [39] que son identificados en la etapa del entrenamiento;
- y_1 corresponde a la variable de salida.

B. Sistemas Neurodifusos

Una RN es un sistema paralelo, no lineal que consiste de un arreglo de unidades de procesamiento simple o neuronas artificiales que están interconectadas de forma masivamente paralela y distribuida para imitar una RN biológica [40]. Las RN artificiales reciben vectores de valores numéricos y tratan cada valor del vector de forma independiente con un peso particular denominado sináptico [41].

Se han propuesto varios algoritmos de entrenamiento para las RN entre ellos el de retropropagación [42] que en combinación con el algoritmo de mínimos cuadrados (MC) realiza una actualización continua de los pesos sinápticos de acuerdo a la contribución de cada neurona sobre el error total de la red.

Asimismo, se han acoplado SID a RN generando SND con la intención de mejorar las características de flexibilidad y adaptabilidad de la red [26, 31].

El SND tiene tres capas: la primera recibe los valores de la variable de entrada, la segunda capa es oculta y representa las reglas difusas, y la tercera capa representa los valores de la variable de salida [43]. La operación de cada capa es como sigue:

Capa 1 (*de entrada*): Cada nodo en esta capa corresponde a una variable de entrada y transmite los valores de entrada a la siguiente capa sin cálculo. Así, el peso de enlace para cada nodo en esta capa es la unidad.

Capa 2 (*de difuminación*): Cada nodo corresponde a una etiqueta lingüística a partir de una de las variables de entrada en la capa anterior. El enlace de salida para cada uno de los nodos representa el valor de membresía, que denota el grado en el que un valor de entrada pertenece a un conjunto difuso; el grado de membresía es calculado en esta capa.

Capa 3 (*antecedente de regla*): Cada nodo en esta capa corresponde a la parte del antecedente de una regla. Generalmente se utiliza un operador norma T en este nodo. La salida de un nodo de Capa 3 representa la fuerza de disparo de

la regla difusa correspondiente.

Capa 4 (*normalización de la fuerza de la regla*): Cada nodo en esta capa calcula la relación de la fuerza de disparo de la regla i -ésima y la suma de todas las reglas de la fuerza de disparo, como en (2):

$$\bar{w}_i = \frac{w_i}{w_1 + w_2 + \dots} \quad i = 1, 2, \dots \quad (2)$$

Donde \bar{w}_i es la fuerza de disparo normalizada y w_i es la fuerza de la regla i -ésima.

Capa 5 (*consecuencia de la regla*): Cada nodo i en esta capa contiene un nodo función:

$$\bar{w}_i f_i = \bar{w}_i (p_i x_1 + q_i x_2 + r_i) \quad (3)$$

Donde \bar{w}_i es la salida de la capa 4 y $\{p_i, q_i, r_i\}$ es el conjunto de parámetros. El algoritmo MC es normalmente utilizado para determinar los parámetros del consecuente representados por f_i [44].

Capa 6 (*inferencia de la regla*): El único nodo presente en esta capa calcula la salida general (*overall output*) como la suma de todas las señales entrantes ($\bar{w}_i f_i$), como se describe en (4):

$$\text{overalloutput} = \sum \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \quad (4)$$

Los componentes de un SND son similares a los de un SID, sin embargo, añadiendo una capa oculta de neuronas en cada etapa se aprovecha la capacidad de aprendizaje de las RN y se aumenta el conocimiento del sistema [45].

El procedimiento de aprendizaje de un SND tiene dos partes: en la primera, los patrones de entrada se propagan, y los parámetros óptimos del consecuente, es decir, los parámetros de la ecuación lineal de salida, son estimados mediante un procedimiento iterativo de MC. En la segunda parte de ejecución, los patrones son propagados nuevamente y en esta época el algoritmo de retropropagación modifica los parámetros del antecedente, en tanto que los del consecuente permanecen ajustados. Este procedimiento es iterado enseguida.

El primer SND híbrido fue ANFIS (Sistemas Adaptativos de Inferencia Neurodifusa —por sus siglas en inglés—, *Adaptive NeuroFuzzy Inference Systems*) que tiene el más bajo valor de la raíz del error cuadrático medio con respecto a otros SND [44].

IV. PREPARACIÓN DEL SISTEMA ADAPTATIVO MULTICAPA DE INFERENCIA NEURODIFUSA

En nuestro estudio utilizamos cuatro conjuntos de datos correspondientes a las calificaciones de cuatro materias del primer semestre de nivel licenciatura, mismas que hemos denominado: CFC (Fundamentos de Computación), DPL (Pensamiento Lógico Matemático), DPP (Planeación de Proyecto) y DS (Diagnóstico Situacional). La Tabla 2 tomada de [24] detalla el contenido de los cuatro conjuntos de datos.

TABLA 2.
DESCRIPCIÓN DE LOS CONJUNTOS DE DATOS
MAT: MATERIA; MTR: MATRÍCULA

MAT	MTR	APROBADOS	REPROBADOS	DESERCIÓN
CFC	169	42	69	58
DPL	180	51	99	30
DPP	165	45	56	64
DS	178	61	65	52
TOTALES	692	199	289	204

De la Tabla 2, eliminamos los valores atípicos identificados, así como lo estudiantes que desertaron. En total, para este estudio utilizamos 488 registros distribuidos entre los cuatro conjuntos de datos. El rango de entrada de U1 para cada conjunto de datos utilizado es:

- CFC: 8 a 100
- DPL: 14 a 100
- DPP: 5 a 100
- DS: 2 a 100

Identificamos que el problema a resolver se caracteriza por tener una salida de tipo continuo, representada por y_l en (1), que corresponde a las calificaciones finales de los estudiantes en cada materia, y que hemos denominado CF. Dado que el SID Takagi-Sugeno ha demostrado buen desempeño en este tipo de problemas [33], se justifica su aplicación.

Como variable de entrada tenemos las calificaciones obtenidas en la unidad de estudio 1, denominada variable U1 y representadas por x_l a x_n en (1).

El número inicial y tipo de FM asignado a cada una de las variables de entrada se decidió mediante el algoritmo *Grid Partitioning* (Partición de Rejilla o GP). Las formas de las FM se ajustaron mediante la fase de aprendizaje del algoritmo, llevando a cabo una búsqueda exhaustiva.

El algoritmo GP nos permite elegir el número y tipo de función de membresía (FM), que representan los parámetros del antecedente que han sido ajustados para el ciclo actual a través del conjunto de entrenamiento.

Nuestra elección de los tipos de FM usados, triangular, trapezoidal y gaussiana, se basó en el supuesto que coincide con Leondes [46], en cuanto a que los valores promedio del antecedente dentro de estos tres tipos de FM, pueden ser considerados para el cálculo del consecuente. Asimismo, estos tres tipos de FM son los más frecuentemente utilizados en aplicaciones de la lógica difusa [47, 48, 49].

Los parámetros ajustados de las FM se muestran en la Tabla 3.

TABLA 3
CONFIGURACIÓN POR SISTEMA

MAT: Conjunto de datos; TFM: Tipo de función de membresía; NFM: Número de funciones de membresía;

MAT	TFM	NFM	PARÁMETROS	
			ENTRADA	SALIDA
CFC	Trapezoidal	3	-24.20; -5.80; 24.19; 43.70;	0.50; -3.58;
			27.64; 42.81; 58.72; 83.92;	0.74; -18.68;
			56.33; 77.19; 113.80; 132.20	0.56; 19.48
DPL	Trapezoidal	2	-46.20; -11.80; 41.42; 75.68;	0.76; -10.71;
			39.04; 75.74; 125.8; 160.2	1.55; -76.01
DPP	Gaussiana	4	12.87; 4.63;	1.38; -5.21;
			7.13; 33.84	1.68; -40.38
			13.96; 69.69	2.18; -96.08
			16.23; 99.30	0.88; -23.01
DS	Gaussiana	4	13.78; 1.97	1.48; -2.59
			11.88; 33.87	1.96; -49.76
			7.93; 64.02	2.97; -158.7
			14.05; 98.29	2.88; -205.83

En el algoritmo GP la base de reglas difusas contiene una regla por cada FM de entrada. La Tabla 4 muestra las reglas obtenidas por cada materia.

TABLA 4
REGLAS DIFUSAS POR SAMIND
MAT: Conjunto de datos

MAT	REGLAS
CFC	If (input1 is in1mf1) then (output is out1mf1)
	If (input1 is in1mf2) then (output is out1mf2)
	If (input1 is in1mf3) then (output is out1mf3)
DPL	If (input1 is in1mf1) then (output is out1mf1)
	If (input1 is in1mf2) then (output is out1mf2)
DPP	If (input1 is in1mf1) then (output is out1mf1)
	If (input1 is in1mf2) then (output is out1mf2)
	If (input1 is in1mf3) then (output is out1mf3)
	If (input1 is in1mf4) then (output is out1mf4)
DS	If (input1 is in1mf1) then (output is out1mf1)
	If (input1 is in1mf2) then (output is out1mf2)
	If (input1 is in1mf3) then (output is out1mf3)
	If (input1 is in1mf4) then (output is out1mf4)

El método de defuzzificación utilizado fue *waver*. El valor de la mejor precisión (MAR), es decir, la más cercana a cero, fue obtenido a través de la iteración y la observación del comportamiento del algoritmo desde $x = 1$, donde $x =$ número de época, hasta aquella época en la que se mostrara la mejor precisión y que, a partir de ésta época, el valor de la precisión comenzara a empeorar sin mejorar en épocas posteriores. La época final de iteración, se estableció como hiperparámetro, correspondiendo a aquella donde ya no se observó mejora en la precisión.

El número de experimentos realizados fueron determinados en función del número de épocas por cada conjunto de datos, para las materias CFC, DPP y DS se llevaron a cabo de 1 a 200 épocas, mientras que para el conjunto DPL fue de 1 a 196 épocas.

Las épocas con mejor precisión de la predicción para cada conjunto de datos fueron 191, 88, 132 y 126 para CFC, DPL, DPP y DS, respectivamente.

La arquitectura del ANFIS que utilizamos en este estudio para implementar un SND Takagi-Sugeno, es mostrada en la Fig. 1 [40].

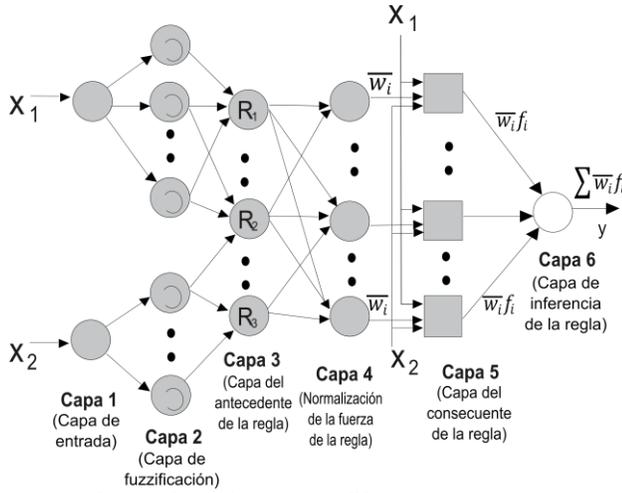


Fig. 1. Arquitectura de un Sistema Neurodifuso

V. METODOLOGÍA

A. Generación de Modelos

En este trabajo, se generó un SAMIND para ser comparado en su precisión contra tres RN, previamente entrenadas en el trabajo por Ulloa Cazarez y López Martín [24]. Para la generación de los modelos, primero se realizaron análisis estadísticos de los cuatro conjuntos de datos CFC, DPL, DPP y DS. Este análisis estadístico se presenta completo en el trabajo mencionado [24] y aquí, en la Tabla 5, se presentan las ecuaciones de regresión generadas para dichos conjuntos, de donde se rescata que existe una relación moderadamente fuerte entre U1 y CF para todos los conjuntos y es estadísticamente significativa con un 99% de confianza con base en los valores del p-valor calculado.

TABLA 5
ANÁLISIS DE REGRESIÓN
ERL: Ecuaciones de regresión lineal

Materia	R	R ²	ERL	P-valor ANOVA
CFC	0.82	0.67	CF = -15.15 + 0.90*U1	0.0000
DPL	0.78	0.61	CF = -19.36 + 0.91*U1	0.0000
DPP	0.85	0.72	CF = - 5.15 + 0.79*U1	0.0000
DS	0.72	0.52	CF = -14.76 + 0.87*U1	0.0000

Asimismo, en el trabajo previo mencionado [24], se generaron tres RN, una RN de Propagación Múltiple (RNPM), una RN de Base Radial (RNBR) y una RN de Regresión (RNRG), de las que describimos sus parámetros y resultados en la Tabla 6.

TABLA 6
CONFIGURACIÓN DE LAS RN Y SUS VALORES DE MRA Y MdRA
MAT: Materia; #NA: NÚMERO DE NEURONAS; SP: VALOR DE SPREAD

MAT	RNPM				RNBR			RNRG	
	#NA	MRA	MdRA	Sp	MRA	MdRA	Sp	MRA	MdRA
CFC	11	14.04	12.57	10	13.10	12.26	6	14.40	13.30
DPL	14	14.62	13.59	15	14.40	12.18	11	15.69	14.84
DPP	10	9.97	6.89	9	8.59	4.59	4	10.46	7.72
DS	12	15.59	12.92	13	14.66	12.76	7	16.49	14.70

B. Generación del Sistema Adaptativo Multicapa de Inferencia Neurodifusa

El SAMIND fue creado (incluyendo las FM y sus reglas difusas) utilizando como datos de entrada las U1 de las cuatro materias y sus CF respectivas como los datos de salida.

Asociando la regla difusa de la ecuación (1) al presente

estudio, x_1 representa a las calificaciones de la U1 y A_1^i simboliza a las etiquetas lingüísticas de la variable de entrada, *Baja*, *Media*, *Alta* o *Excelente*. La salida y_1 representa la ecuación lineal con sus respectivas etiquetas lingüísticas (para la variable de salida) *Inferior*, *Promedio*, *Superior* o *Sobresaliente*. El SAMIND generará una regla difusa por cada número de FM.

En la fase de optimización del SAMIND se hace uso de un método híbrido del algoritmo de retropropagación que se usa para aprender los parámetros del antecedente y el algoritmo de MC para determinar los parámetros del consecuente (coeficientes de la combinación lineal en las conclusiones de las reglas).

Los parámetros del SAMIND son obtenidos mediante el método de GP que divide el espacio de datos en secciones rectangulares utilizando una partición paralela al eje basada en un número predefinido de FM y sus tipos en cada dimensión [41]. Esta partición únicamente es adecuada para casos con un número pequeño de variables de entrada (menor que 6) [35].

C. Entrenamiento y Prueba de los Modelos

Para el entrenamiento y validación de los modelos, el método *leave-one-out* (LOO) es ampliamente utilizado pues es determinístico [42] y permite entrenar y probar el modelo con el total del conjunto de datos. El LOO consiste en el entrenamiento de cada modelo con los registros totales de los estudiantes, menos uno. El modelo entrenado se prueba con el estudiante no incluido en el conjunto de entrenamiento. Este procedimiento se ejecuta por cada registro en el conjunto de datos y se calculan los residuales absolutos (RA) por modelo, para luego calcular la media (MRA) y mediana (MdRA) por conjunto de datos.

La precisión de la predicción de los modelos se compara utilizando los RA, debido a que es una medida asimétrica que no presenta sesgo [42]. El RA se calcula según la ecuación 5.

$$RA_i = | \text{CalificaciónFinal}_i - \text{CalificaciónPredicha}_i | \quad (5)$$

El valor del RA se obtiene por cada estudiante i cuya calificación final es predicha. La agregación de RA sobre múltiples estudiantes (n) se puede calcular a través de su MRA o de su MdRA y la precisión del modelo es inversamente proporcional a la MRA o la MdRA.

Un SAMIND fue entrenado y probado para cada una de las cuatro materias con base en el método LOO para su validación. Los parámetros configurados por SAMIND se muestran en la Tabla 3 y la Tabla 7 incluye los mejores resultados obtenidos por el SAMIND para cada materia.

TABLA 7
MRA Y MdRA OBTENIDOS POR LAS SAMIND

MATERIA	MRA	MdRA
CFC	15.18	13.14
DPL	15.82	12.21
DPP	11.48	6.95
DS	16.48	12.60

Los tipos de FM para el SAMIND que fueron comparadas para encontrar la mejor precisión (menor MRA) por materia fueron Triangular, Trapezoidal y Gaussiana. Asimismo, el rango del número de FM usadas fue entre dos y cuatro para cada

uno de los tipos de FM (Tabla 3).

D. Selección de la Prueba Estadística para la Comparación de las Precisiones entre los Modelos

La selección de la prueba estadística adecuada para comparar grupos de datos depende de tres factores:

1. La dependencia de los datos. Dado que cada SAMIND así como cada RN fueron aplicados a cada uno de los estudiantes de cada materia, entonces los conjuntos son dependientes.
2. La distribución de los datos. Los conjuntos de datos de los RA de los tres tipos de redes neuronales mostraron no tener distribución normal [24], mientras que, con relación al conjunto de los RA de las SAMIND, la Tabla 8 muestra que tampoco ninguno de ellos observa una distribución normal con una confianza estadística del 99%.
- 3.

TABLA 8

P-VALORES DE PRUEBAS ESTADÍSTICAS DE DISTRIBUCIÓN NORMAL APLICADAS A LOS CONJUNTOS DE RA POR SAMIND

MATERIA	χ^2	SHAPIRO-WILK	ASIMETRÍA	CURTOSIS
CFC	0.0000	0.0000	0.0319	0.7189
DPL	0.0000	0.0000	0.0069	0.5237
DPP	0.0000	0.0000	0.0002	0.0021
DS	0.0000	0.0000	0.0021	0.0953

3. El número de conjuntos a ser comparados. Dado que los conjuntos de datos son dependientes, que ninguno de ellos muestra distribución normal, y que la comparación es entre más de dos conjuntos, la prueba estadística adecuada es Friedman [43]. Los resultados de los p-valores de esta prueba se presentan en la Tabla 9.

TABLA 9

PRUEBA DE FRIEDMAN POR MODELOS/MATERIA

MATERIA	P-VALOR
CFC	0.0000
DPL	0.0001
DPP	0.0000
DS	0.0030

En los cuatro conjuntos el p-valor es menor que 0.01, por lo que hay diferencia estadísticamente significativa en al menos un modelo con un 99% de confianza.

Para identificar qué modelo resultó diferente en términos estadísticos, procedimos a ejecutar comparaciones por pares de modelos por materia, como a continuación se describe.

1. Generamos un nuevo conjunto de datos correspondiente a las diferencias entre los RA de cada RN y de SAMIND.
2. Aplicamos las cuatro pruebas de distribución normal a cada conjunto de diferencias y ningún conjunto resultó con distribución normal, por lo que la comparación de la precisión para cada par de modelos deberá ser con base en las medianas reportadas en las Tablas 6 y 7, aplicando la prueba de Wilcoxon [50, 51]. La Tabla 10 muestra los p-valores resultantes.

TABLA 10

P-VALORES DE LA APLICACIÓN DE LAS PRUEBAS ESTADÍSTICAS (T-PAREADA O WILCOXON) A PARES DE MODELOS POR MATERIA (A: SAMIND, B: RNPM, C: RNBR, D: RNRG)

Materia	p-valor		
	A-B	A-C	A-D
CFC	0.0418	0.0000	0.0000
DPL	0.0034	0.0014	0.7214
DPP	0.0000	0.0000	0.0000
DS	0.0117	0.0015	0.6682

VI. RESULTADOS

Tomando como referencia los resultados reportados en las Tablas 6, 7 y 10, la precisión de la predicción del SAMIND superó estadísticamente al menos a una RN en cada caso:

- A la RNRG para CFC con un 99% de confianza
- A la RNPM para DPL con un 99% de confianza
- A la RNRG para DPP con un 99% de confianza
- A la RNBR con un 99% de confianza y a la RNPM con un 95% de confianza para DS

Con ello, las siguientes hipótesis derivadas de la formulada en la sección de Introducción de este estudio, pueden aceptarse en favor del SAMIND:

H_{a1} Existe diferencia estadísticamente significativa entre la precisión de la predicción de un SAMIND:

- y la de una RNRG, cuando los modelos son entrenados y probados para predecir la calificación final de los cursos de educación superior en línea CFC y DPP.
- y la de una RNPM cuando los modelos son entrenados y probados para predecir la calificación final de los cursos de educación superior en línea DPL y DS.
- y la de una RNBR cuando los modelos son entrenados y probados para predecir la calificación final del curso de educación superior en línea DS.

Asimismo, aun cuando el SAMIND obtuvo un mejor valor de MdRA, la diferencia resultó estadísticamente no significativa frente a las siguientes redes neuronales:

- RNRG para DPL con un 99% de confianza
- RNRG para DS con un 99% de confianza

Por lo que las siguientes hipótesis también son aceptadas:

H₀ No existe diferencia estadísticamente significativa entre la precisión de la predicción de un SAMIND y la de RNRG cuando los modelos son entrenados y probados para predecir la calificación final de los cursos de educación superior en línea DPL y DS.

VII. DISCUSIÓN

En este trabajo, al construir conocimiento a partir de un trabajo previo, mediante su replicación y comparación, hemos generado resultados en al menos dos sentidos: uno en función de la utilidad y confianza de la predicción; el otro en términos de la generación de modelos con mayor precisión predictiva.

Retomando un trabajo anteriormente publicado [24], se han generado modelos SAMIND de predicción del DE en ESL, que en términos de precisión superan a la regresión estadística y a algunos modelos anteriormente probados como alternativas para la predicción: SAMIND superó al menos a una RN en cada

conjunto de datos, sugiriendo la aplicación del SAMIND como modelo alternativo a una red neuronal para la predicción del DE en ESL.

Desde el punto de vista de la aplicación práctica de nuestro estudio respecto a otros mencionados aquí, una ventaja a señalar radica en el uso del método *leave-one-out* para el entrenamiento y validación del modelo lo que permite generar resultados más representativos del evento educativo, considerando que se utilizan todos los datos de los conjuntos y que sus resultados no llevan a la asimetría [42]. Esto, brinda mayor confianza en los resultados y en sus interpretaciones.

Asimismo, los modelos SAMIND de este trabajo brindan opciones computacionales validadas para fortalecer la toma de decisiones y la gestión educativa, proporcionando más herramientas a los profesores y administradores educativos. Por ejemplo, hemos utilizado una sola variable independiente (U1) generando un modelo sencillo y factible de replicación para predecir el DE.

Por otro lado, las variables U1 y CF fueron colectadas del sistema de calificaciones del LMS, U1 al finalizar la primera unidad de estudio y CF al finalizar el semestre. Ambas variables son comunes en el contexto educativo por lo que utilizando conjuntos de datos de otros programas y materias de ESL, SAMIND es generalizable siempre y cuando se garantice que el procedimiento de colección de las variables U1 y CF es el mismo que el reportado en el trabajo publicado y mencionado previamente [24].

La motivación principal de usar un modelo híbrido fue la de combinar las ventajas de una RN y lógica difusa con el objetivo de desarrollar un modelo SAMIND basado en un SND que encuentra los parámetros apropiados y un modelo con menor error y mayor precisión.

Las reglas difusas del SND permiten concluir que, si la calificación de la U1 es baja, la calificación final será inferior; mientras que si U1 es media, la CF será promedio; y así sucesivamente hasta que CF sea sobresaliente cuando U1 sea excelente.

Con relación a las limitaciones del trabajo, identificamos las siguientes:

- Dado el carácter exploratorio del estudio, los modelos SAMIND fueron configurados a través de una forma empírica en la mayoría de sus parámetros.
- Como es común en los modelos matemático-computacionales, SAMIND observa una dificultad en términos de su interpretación del fenómeno educativo.
- El conjunto de datos utilizado se limita a un solo programa educativo de ESL.
- El algoritmo GP se utilizó con una sola variable (U1) cuando permite el uso de más variables.

A partir de este estudio, como trabajo futuro identificamos las siguientes oportunidades:

- Probar el SAMIND utilizando el algoritmo de GP en una mayor capacidad, por ejemplo, usando más variables independientes para definir el número y tipo de FM.
- Utilizar una mayor cantidad de registros de estudiantes provenientes de otras materias y en distintos momentos del programa educativo.

AGRADECIMIENTOS

Los autores de este trabajo agradecen al Dr. Cuauhtémoc López Martín por su valioso apoyo en la realización de los análisis. Asimismo, agradecen al Sistema de Universidad Virtual de la Universidad de Guadalajara, México, como al Tecnológico Nacional de México/I.T. Colima, México, por el apoyo otorgado a los autores en la realización de este trabajo.

REFERENCIAS

- [1] S. K. Banihashem, K. Aliabadi, S. P. Ardakani, A. Delaver and M. N. Ahmadabadi, "Learning Analytics: A Systematic Literature Review," *Interdisciplinary Journal of Virtual Learning in Medical Sciences*, vol. 9, no. 2, p. e63024, 2018.
- [2] T.-Y. Yang, C. G. Brinton, C. Joe-Wong y M. Chiang, «Behavior-Based Grade Prediction for MOOCs Via Time Series Neural Networks,» *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no 5, pp. 716-728, 2017.
- [3] S. Huang y N. Fang, «Predicting student academic performance in an engineering dynamics course: a comparison of four types of predictive mathematical models,» *Computers & Education*, vol. 61, pp. 133-145, 2013.
- [4] Y. Chen, Q. Zheng, S. Ji, F. Tian, H. Zhu y M. Liu, «Identifying at-risk students based on the phased prediction model,» *Knowledge and Information Systems*, vol. 59, pp. 1-17, 2019.
- [5] A. Mueen, B. Zafar y U. Manzoor, «Modeling and predicting students' academic performance using data mining techniques,» *International Journal Modern Education and Computer Science*, vol. 11, pp. 36-42, 2016.
- [6] S. Helal, J. Li, L. Liu, E. Ebrahimi, S. Dawson, D. J. Murray y Q. Long, «Predicting academic performance by considering student heterogeneity,» *Knowledge-Based Systems*, vol. 161, no 1, pp. 134-146, 2018.
- [7] J. L. C. Ramos, R. E. D. Silva, R. L. Rodrigues, J. C. S. Silva and A. S. Gomes, "A comparative study between clustering methods in educational data mining," *IEEE Latin America Transactions*, vol. 14, no. 8, pp. 3755-, 2016.
- [8] J.-F. Chen, H.-N. Hsieh y Q. H. Do, «Predicting Student Academic Performance: A Comparison of Two Meta-Heuristic Algorithms Inspired by Cuckoo Birds for Training Neural Networks,» *Algorithms*, vol. 7, no 4, pp. 538-553, 2014.
- [9] S. Natek and M. Zwillling, "Student data mining solution-knowledge management system related to higher education institutions," *Expert Systems with Applications*, vol. 41, pp. 6400-6407, 2014.
- [10] Y.-H. Hu, C.-L. Lo y S.-P. Shih, «Developing early warning systems to predict students' online learning performance,» *Computers in Human Behavior*, vol. 36, pp. 469-478, 2014.
- [11] A. Peña-Ayala, «Educational data mining: a survey and a data mining-based analysis of recent works,» *Expert Systems with Applications*, vol. 41, pp. 1432-1462, 2014.
- [12] D. Quang Hung y C. Jeng-Fung, «A neuro-fuzzy approach in the classification of students' academic performance,» *Computational intelligence and neuroscience*, vol. 2013, pp. 1-7, 2013.
- [13] D. Xu and S. Smith Jaggars, "The impact of online learning on students' course outcomes: evidence from a large community and technical college system," *Economics of Education Review*, vol. 37, pp. 46-57, 2013.
- [14] B. Giesbers, B. Rienties, D. Tempelaar and W. Gijssels, "Investigating the relations between motivation, tool use, participation, and performance in an e-learning course using web-videoconferencing," *Computers in Human Behavior*, vol. 29, no. 1, pp. 285-292, 2013.
- [15] J. W. You, "Examining the effect of academic procrastination on achievement using LMS data in e-learning," *Educational Technology & Society*, vol. 18, no. 3, pp. 64-74, 2015.

- [16] R. L. Ulloa-Cazarez, C. López-Martín, A. Abran y C. Yáñez-Márquez, «Prediction of Online Students Performance by Means of Genetic Programming.» *Applied Artificial Intelligence*, vol. 32, no 9-10, pp. 858-881, 2018.
- [17] I. Lykourantzou, I. Giannoukos, G. Mpardis, V. Nikolopoulos y V. Loumos, «Early and dynamic student achievement prediction in E-Learning courses using neural networks.» *Journal of the American Society for Information Science and Technology*, vol. 60, no 2, pp. 372-380, 2009.
- [18] X. Wanli, G. Rui, P. Eva and G. Sean, "Participation-based student final performance prediction model through interpretable Genetic Programming: integratin learning analytics, educational data mining and theory," *Computers in Human Behavior*, vol. 47, pp. 168-181, 2015.
- [19] J.-F. Chen y Q. H. Do, «A cooperative Cuckoo Search – hierarchical adaptive neuro-fuzzy inference system approach for predicting student academic performance.» *Journal of Intelligent & Fuzzy Systems*, vol. 27, pp. 2551-2561, 2014.
- [20] B. Aczel, B. Palfi y B. Szaszi, «Estimating the evidential value of significant results in psychological science.» *PLOS one*, vol. 12, no 8, pp. 1-8, 2017.
- [21] K. Chen and S. Jang, "Motivation in online learning: Testing a model of self-determination theory," *Computers in Human Behavior*, vol. 26, pp. 741-752, 2010.
- [22] S. Ransdell, "Online activity, motivation, and reasoning among adult learners," *Computers in Human Behavior*, vol. 26, pp. 70-73, 2010.
- [23] C. Romero, P. G. Espejo, A. Zafra y J. R. Romero, «Web usage mining for predicting final marks of students that use Moodle courses.» 2010.
- [24] G. Adelfio, G. Boscaino y V. Capursi, «A new indicator for higher education student performance.» *Higher Education*, vol. 68, no 5, pp. 653-668, 2014.
- [25] M. Mirzaei y S. Sahebi, «Modeling Students' Behavior Using Sequential Patterns to Predict Their Performance.» de *International Conference on Artificial Intelligence in Education*, Chicago, 2019.
- [26] R. L. Ulloa Cazarez y C. López Martín, «Neural Networks for predicting student performance in online education.» *IEEE Latin America Transactions*, vol. 16, no 7, pp. 2053-2060, 2018.
- [27] Open Science Collaboration, "Estimating the reproducibility of psychological science," *Science*, vol. 349, no. 6251, pp. aac4716-1 - aac4716-8, 2015.
- [28] D. Nauck, «A fuzzy perceptron as a generic model for neuro-fuzzy approaches.» de *2nd German GI-Workshop Fuzzy-Systeme '94*, Munich, 1994.
- [29] UDGVirtual, «Ruta de formación. Licenciatura en Tecnologías e Información.» s/d 2013. [En línea]. Disponible: http://www.udgvirtual.udg.mx/sites/default/files/ruta_de_formacion_it_i_2006.pdf. [Último acceso: 24 marzo 2020].
- [30] C. Villagrà-Arnedo, F. J. Gallego-Durán, P. Compañ-Rosique, F. Llorens-Largo y R. Molina-Carmona, «Predicting academic performance from behavioural and learning data.» *International Journal of Design & Nature and Ecodynamics*, vol. 11, no 3, pp. 239-249, 2016.
- [31] O. E. Varela, J. J. Cater III and N. Michel, "Online learning in management education: an empirical study of the role personality traits," *Journal of Computing in Higher Education*, vol. 24, no. 3, pp. 209-255, 2012.
- [32] A. Zafra and S. Ventura, "Multi-instance genetic programming for predicting student performance in web based educational environments," *Applied Soft Computing*, vol. 12, pp. 2693-2706, 2012.
- [33] A. Altaher y O. BaRukab, «Prediction of Student's Academic Performance Based on Adaptive Neuro-Fuzzy Inference.» *International Journal of Computer Science and Network Security*, vol. 17, no 1, pp. 165-169, 2017.
- [34] O. Taylan y B. Karagözog'lu, «An adaptive neuro-fuzzy model for prediction of student's academic performance.» *Computers & Industrial Engineering*, vol. 57, no 3, pp. 732-741, 2009.
- [35] S. Kar, S. Das y P. K. Ghosh, «Applications of neuro fuzzy systems: A brief review and future outline.» *Applied Soft Computing*, vol. 15, no 1, pp. 243-259, 2014.
- [36] F. Valdez, P. Melin y O. Castillo, «A survey on nature-inspired optimization algorithms with fuzzy logic for dynamic parameter adaptation.» *Expert systems with Applications*, vol. 41, no 14, pp. 6459-6466, 2014.
- [37] T. Takagi y M. Sugeno, «Derivation of fuzzy control rules from human operator's control actions.» *IFAC Proceedings Volumes*, vol. 16, no 13, pp. 55-60, 1983.
- [38] T. Takagi y M. Sugeno, «Fuzzy Identification of Systems and its Application to Modelling and Control.» *IEEE Transactions on Systems, Man, and Cybernetics*, Vols. %1 de %2SMC-15, no 1, pp. 116-132, 1985.
- [39] J.-S. R. Jang, «ANFIS: adaptive-network-based fuzzy inference system.» *IEEE Transactions on Systems, Man and Cybernetics*, vol. 23, no. 3, pp. 665–685., 1993.
- [40] S. Haykin, *Neural Networks. A comprehensive foundation*, Ontario: Pearson Prentice Hall, 1999.
- [41] C. López-Martín y A. Abran, «Neural networks for predicting the duration of new software projects.» *The Journal of Systems and Software*, vol. 101, pp. 127-135, 2015.
- [42] C. López-Martín, C. Isaza y A. Chavoya, «Software development effort prediction of industrial projects applying a general regression neural network.» *Empirical Software Engineering*, vol. 17, no 6, pp. 738-756, 2012.
- [43] D. Nauck, F. Klawonn y R. Kruse, *Foundations of Neuro-Fuzzy Systems*, Nueva York: John Wiley & Sons, Inc, 1997.
- [44] A. Abraham, «Adaptation of Fuzzy Inference System Using Neural Learning.» de *Fuzzy Systems Engineering. Studies in Fuzziness and Soft Computing*, Berlin, Springer, 2005, pp. 53-83.
- [45] MathWorks, *Global Optimization Toolbox 3. User's Guide*, Massachusetts: The MathWorks Inc, 2010.
- [46] C. T. Leondes, *Fuzzy Logic and Expert Systems Applications*, San Diego: Academic Press, 1998.
- [47] J. J. Buckley y E. Eslami, «Chapter 4. Fuzzy Numbers.» de *An Introduction to Fuzzy Logic and Fuzzy Sets*, Berlin, Springer-Verlag, 2002, pp. 55-94.
- [48] B. Bede, «Studies in fuzziness and Soft Computing.» de *Mathematics of Fuzzy Sets and Fuzzy Logic*, vol. 295, Berlin, Springer, 2013, pp. 51-64.
- [49] R. Belohlavek y G. J. Klir, *Concepts and Fuzzy Logic*, Cambridge: MIT Press, 2011.
- [50] S. M. Ross, *Introduction to Probability and Statistics for Engineers and Scientists*, Burlington, MA: Elsevier Press, 2004.
- [51] M. W. J. Conover, *Practical nonparametric statistics*, Nueva York: Wiley, 1998.



Rosa Leonor Ulloa-Cazarez, nació en el estado de Sonora, México, en 1975. Es miembro del Sistema Nacional de Investigadores en México desde 2019, con nivel Candidato. Pertenece al Cuerpo Académico en Sistemas para la Gestión del Conocimiento en la Educación y cuenta con reconocimiento de Perfil

Deseable del Programa para el Desarrollo Profesional Docente, otorgado por la Secretaría de Educación Pública. Obtuvo el doctorado en Tecnologías de la Información en 2016, en 2011 el grado de Maestra en Gestión y Políticas de la Educación Superior, y en 2010, la licenciatura en Educación, todos por la Universidad de Guadalajara. La Dra. Ulloa se desempeña actualmente como profesora e investigadora de tiempo completo para el Sistema de Universidad Virtual de la Universidad de Guadalajara desde 2016, donde además ha desempeñado varios cargos relacionados con la gestión de proyectos tecnológicos. Ha publicado cuatro artículos científicos en revistas indexadas en el Journal Citation Reports, y desde el año 2018 ha sido considerada para el arbitraje de artículos científicos para la revista IEEE Latin America Transactions, IEEE Access, entre otras.

Ingeniería de Software y sistemas multiusuario MIMO-OFDM. Actualmente es miembro de la IEEE y de la Sociedad de Comunicaciones de la IEEE.



Noel García-Díaz Nació en el estado de Colima en 1975. Es profesor investigador adscrito al Tecnológico Nacional de México/I.T. Colima, en Colima México. Obtuvo el grado de doctor en Tecnologías de la Información en 2014 por la Universidad de Guadalajara. Sus áreas de interés son la estadística y las técnicas de

aprendizaje de máquina aplicadas a la Ingeniería de Software y los modelos de predicción para la Ingeniería de Software. El Dr. García pertenece al Cuerpo Académico ITCOL-CA-3 Sistemas Inteligentes y desde 2018 cuenta con el reconocimiento de Perfil Deseable PRODEP, otorgado por la Secretaría de Educación Pública.



Leonel Soriano-Equigua Nació en el estado de Colima en México, en 1974. Es profesor e investigador en la Facultad de Ingeniería Mecánica y Electrónica en la Universidad de Colima y miembro del Sistema Nacional de Investigadores de México, nivel I. Recibió sus grados de Maestría y Doctorado en el Centro de Investigación CICESE. El Dr. Soriano

pertenece al Cuerpo Académico UCOL-CA-83 Investigación de Sistemas Basados en el Conocimiento. Cuenta con reconocimiento de Perfil Deseable PRODEP, otorgado por la Secretaría de Educación Pública. Recibió sus grados de Maestría y Doctorado en el Centro de Investigación CICESE. Fue investigador visitante en el Departamento de Ingeniería Eléctrica y Computación en la Universidad de Texas en Austin en 2007. Su investigación actual incluye los campos de la