

Recibido el 9 de diciembre de 2021, aceptado el 8 de enero de 2022, fecha de publicación 11 de enero de 2022, fecha de la versión actual 24 de enero de 2022.

Identificador de objeto digital 10.1109/ACCESS.2022.3142097

Predicción de la diabetes potenciada con aprendizaje automático fusionado

USAMA AHMED^{1,2}, GHASSAN F. ISSA³, MUHAMMAD ADNAN KHAN^{1,4}, SHABIB AFTAB^{2,5},
(Miembro, IEEE), MUHAMMAD FARHAN KHAN⁶, RAED EN DICHO⁷,
TAHER M. GHAZAL^{3,8}, (Miembro, IEEE) Y MUNIR AHMAD⁵, (Miembro, IEEE)

¹Escuela Riphah de Computación e Innovación, Facultad de Computación, Universidad Internacional Riphah, Lahore 54000, Pakistán

²Departamento de Ciencias de la Computación, Universidad Virtual de Pakistán, Lahore 54000, Pakistán

³Escuela de Tecnología de la Información, Skyline University College, Ciudad Universitaria, Al Sharjah, Emiratos Árabes Unidos

⁴Laboratorio de reconocimiento de patrones y aprendizaje automático, Departamento de Software, Universidad Gachon, Seongnam-si 13557, Corea del Sur

⁵Facultad de Ciencias de la Computación, Facultad Nacional de Administración de Empresas y Economía, Lahore 54000, Pakistán

⁶Departamento de Ciencias Forenses, Universidad de Ciencias de la Salud, Lahore 54000, Pakistán

⁷Facultad de Administración, Universidad Canadiense de Dubai, Dubai, Emiratos Árabes Unidos

⁸Centro de Seguridad Cibernética, Facultad de Ciencia y Tecnología de la Información, Universiti Kebangsaan Malaysia, Bangi, Selangor 43600, Malasia

Autores correspondientes: Munir Ahmad (munir@ncbae.edu.pk) y Muhammad Adnan Khan (adnan@gachon.ac.kr)

ABSTRACTO En el ámbito médico, es fundamental predecir las enfermedades de forma temprana para prevenirlas. La diabetes es una de las enfermedades más peligrosas en todo el mundo. En los estilos de vida modernos, el azúcar y las grasas suelen estar presentes en nuestros hábitos alimentarios, lo que ha aumentado el riesgo de diabetes. Para predecir una enfermedad, es extremadamente importante comprender sus síntomas. Actualmente, los algoritmos de aprendizaje automático (ML) son valiosos para la detección de enfermedades. Este artículo presenta un modelo que utiliza un enfoque de aprendizaje automático fusionado para la predicción de la diabetes. El marco conceptual consta de dos tipos de modelos: modelos de máquina de vectores de soporte (SVM) y de redes neuronales artificiales (ANN). Estos modelos analizan el conjunto de datos para determinar si un diagnóstico de diabetes es positivo o negativo. El conjunto de datos utilizado en esta investigación se divide en datos de entrenamiento y datos de prueba con una proporción de 70:30 respectivamente. La salida de estos modelos se convierte en la función de pertenencia de entrada para el modelo difuso, mientras que la lógica difusa finalmente determina si un diagnóstico de diabetes es positivo o negativo. Un sistema de almacenamiento en la nube almacena los modelos fusionados para su uso futuro. Basado en el historial médico del paciente en tiempo real, el modelo fusionado predice si el paciente es diabético o no. El modelo ML fusionado propuesto tiene una precisión de predicción de 94,87, que es mayor que los métodos publicados anteriormente.

TÉRMINOS DEL ÍNDICE Predicción diabética, sistema difuso, modelo de aprendizaje automático fusionado, síntomas diabéticos, predicción de enfermedades.

I. INTRODUCCIÓN

La diabetes es uno de los trastornos metabólicos crónicos más importantes del mundo. Hay dos tipos de diabetes, tipo 1 y tipo 2. Cuando el sistema inmunológico daña las células Beta pancreáticas (b-células), la diabetes tipo 1 se produce dentro del cuerpo, lo que provoca la liberación de una pequeña cantidad de insulina o ninguna. La diabetes tipo 2 es una enfermedad autoinmune en la que las células del cuerpo no interactúan con la insulina o las células pancreáticas no producen suficiente insulina para regular los niveles de glucosa en sangre. Una cantidad insuficiente de insulina hace que los niveles de glucosa en sangre aumenten y el metabolismo de los carbohidratos, grasas y proteínas se debilite, lo que resulta en diabetes tipo 1. Los síntomas de la diabetes incluyen (i) poliuria

El editor asociado que coordinó la revisión de este manuscrito y aprobó su publicación fue Baozhen Yao.

(ii) Polidipsia (iii) Debilidad (iv) Polifagia (v) Obesidad (vi) Pérdida repentina de peso (vii) Candidiasis genital (viii) Visual borroso (ix) Picazón (x) Irritabilidad (xi) Retraso en la curación (xii) Paresia parcial (xiii) Rigidez muscular (xiv) Alopecia, etc. [1]. La diabetes es una enfermedad metabólica que causa millones de muertes en todo el mundo cada año debido a diversas complicaciones de salud. Entre 2000 y 2019 se ha observado un aumento del 70% en la tasa de mortalidad por diabetes en todo el mundo [2]. Se requiere un sistema de diagnóstico inteligente basado en ML para detectar este tipo de enfermedades mortales. Un sistema de decisión experto basado en ML puede diagnosticar con éxito a los pacientes con diabetes en una etapa temprana. Los investigadores utilizaron varios tipos diferentes de conjuntos de datos para predecir la diabetes. El marco basado en ML necesita un conjunto de datos apropiado que tenga las características necesarias para el entrenamiento y la validación. La selección de características apropiadas y relevantes del

El conjunto de datos aumenta las capacidades del modelo ML para predecir con precisión. El conjunto de datos utilizado en el sistema propuesto proviene del repositorio de aprendizaje automático UCI (Universidad de California Irvine) [3], compilado por el hospital de Sylhet, Bangladesh.

La diabetes mellitus (DM) se produce debido a una mala absorción de alimentos que altera el nivel de glucosa en el cuerpo. Las medidas preventivas contra la desnutrición o la obesidad, que a veces son causas principales de la diabetes, incluyen una dieta saludable y un cambio de estilo de vida. Además, estas medidas ayudan a controlar la presión arterial y reducen el riesgo de complicaciones de salud. El chequeo médico facilita el diagnóstico de la enfermedad de la diabetes. También se realizan algunas pruebas de laboratorio para detectar la enfermedad. Los pacientes con DM tipo 2 necesitan insulina que les salve la vida mientras permanezcan con vida. Por lo tanto, si no se aborda, esta condición insalubre agota a los individuos, las familias y los recursos nacionales. La detección precoz y el tratamiento sintomático son fundamentales para garantizar la vida sana y el bienestar de los pacientes prediabéticos. Un sistema inteligente de diagnóstico médico basado en síntomas, signos, pruebas de laboratorio y observaciones será útil en la detección y prevención de enfermedades. La Inteligencia Artificial (IA) también se ha aplicado a los sistemas de diagnóstico médico de varias formas interesantes para la detección de enfermedades. Esta investigación propone un marco para la detección temprana de pacientes diabéticos mediante la fusión de aprendizaje automático.

II. REVISIÓN DE LITERATURA

La literatura reciente ha producido una cantidad significativa de investigaciones para reconocer a los pacientes diabéticos en función de los síntomas mediante la aplicación de técnicas de aprendizaje automático. Basado en aprendizaje supervisado, aprendizaje híbrido o aprendizaje en conjunto, Pradhan *et al.* [4] aplicaron varios algoritmos para diagnosticar la diabetes mellitus para obtener una mayor tasa de precisión, pero el enfoque conjunto funciona mejor que los otros dos enfoques. En un enfoque conjunto, Kumari *et al.* [5] mejoró la precisión de la clasificación mediante la aplicación de un clasificador de votación suave al conjunto de datos Pima-Diabetes y al conjunto de datos Breast-Cancer. Según los resultados, el clasificador de votación suave logró una precisión del 79,08 % en comparación con otros algoritmos de aprendizaje automático.

Sarwar *et al.* [6] utilizaron algoritmos de aprendizaje automático para la detección de diabetes en una etapa temprana utilizando el conjunto de datos Pima Diabetes. Sus tasas de precisión logradas con KNN y SVM fueron del 77%, que es más alta que las de los otros cuatro algoritmos. Una limitación de este artículo es el tamaño del conjunto de datos y los valores faltantes. Dey *et al.* [7] utilizó algoritmos de aprendizaje automático supervisados: SVM, KNN, Naive Bayes y ANN con escalado Min-Max (MMS) en el conjunto de datos Pima. La precisión del modelo ANN con MMS es del 82,35 %, superior a la de los otros cuatro algoritmos. En [8], los investigadores utilizaron algoritmos de aprendizaje automático, incluidos Naive Bayes, Random Forest y Simple CART, y utilizaron la herramienta Weka para predecir la diabetes. El clasificador SVM funciona mejor y logró una precisión del 79,13%, que es superior a la de los otros tres algoritmos. Saru *et al.* [9] predijo la diabetes utilizando un modelo basado en Regresión Logística, SVM, Árbol de Decisión,

y KNN. También compararon sus tasas de precisión con y sin Bootstrapping. La tasa de precisión del árbol de decisión con bootstrapping es del 94,4%, que es más alta que la de los otros dos algoritmos.

Utilizando algoritmos de aprendizaje automático como Decision Tree, ANN, Naive Bayes y SVM, Sonar y Jaya Malini [10] construyeron un modelo para predecir pacientes diabéticos. La tasa de precisión del árbol de decisión es del 85%, que es más alta que la de los otros dos algoritmos. Wei *et al.* [11], en su artículo, diseñó un modelo utilizando algoritmos de ML como Naive Bayes, Deep Neural Network (DNN), Regresión logística y árboles de decisión. La tasa de precisión de DNN es del 77,86%, que es más alta que la de los otros cuatro algoritmos. Faruque *et al.* [12] propuso un modelo que utiliza cuatro algoritmos de ML: Support Vector Machine (SVM), C4.5 Decision Tree, K-Nearest Neighbor (KNN) y Naive Bayes para predecir la diabetes. La tasa de precisión del árbol de decisión C4.5 es del 73,5%, que es más alta que la de los otros tres algoritmos. Jain *et al.* [13] predijo la diabetes, utiliza varios algoritmos de ML como Neural Network (NN), Fisher Linear Discriminant Analysis (FLDA), Random Forest, Chi-square Automatic Interaction Detección (CHAID) y SVM. La tasa de precisión de NN es del 87,88%, que es más alta que la de los otros cuatro algoritmos.

Los algoritmos de aprendizaje automático son actualmente útiles para la detección de enfermedades, pero los modelos de investigación anteriores son menos precisos porque generalmente se centran en técnicas de preprocesamiento, equilibrio de datos y varios tipos de modelos de aprendizaje supervisados y semisupervisados. Por lo tanto, es necesario encontrar una nueva técnica con fusión de niveles de decisión que pueda integrar la precisión de múltiples algoritmos de aprendizaje automático con una alta precisión en la detección de enfermedades. Para este propósito, se propone un modelo de ML fusionado que utiliza dos enfoques de aprendizaje automático supervisado, incluidos ANN y SVM [14]-[16], seguidos de lógica difusa para la fusión a nivel de decisión.

III. MATERIALES Y MÉTODOS

Este artículo propone un modelo fusionado para la predicción de la diabetes (FMDP). El modelo FMDP propuesto consta de dos fases principales. La primera fase consta de la capa de entrenamiento, mientras que la segunda fase consta de la capa de prueba. La capa de capacitación se divide en diferentes etapas, que incluyen adquisición de datos, preprocesamiento, clasificación, evaluación del desempeño y fusión de aprendizaje automático. El conjunto de datos utilizado en esta investigación se tomó del Repositorio de Aprendizaje Automático de la UCI [3]. En la etapa de Adquisición de datos, se puede utilizar un conjunto de datos que tenga suficientes características para predecir la diabetes. Los datos se limpian, normalizan y dividen en conjuntos de datos de entrenamiento y prueba durante la etapa de preprocesamiento. Los datos preprocesados se pueden utilizar para entrenar máquinas de vectores de soporte (SVM) y redes neuronales artificiales (ANN) para la predicción. Podemos seleccionar varios algoritmos de aprendizaje automático para que la clasificación alcance la precisión requerida. Sin embargo, en el modelo propuesto, utilizamos solo dos algoritmos de ML ampliamente utilizados (SVM y ANN) [14], [16], [19]. Estos algoritmos se seleccionan en este

Investigamos después de algunos experimentos iniciales donde hemos encontrado que estas técnicas son más efectivas para este problema. Utilizamos varias medidas de precisión, que incluyen: exactitud, especificidad, sensibilidad, precisión y puntuación F1 en la etapa de Evaluación del desempeño. Si el modelo propuesto no cumple con los requisitos de aprendizaje, será reentrenado. Cuando se cumplen los requisitos de aprendizaje, las salidas ANN y SVM se utilizan como entradas en la fusión del aprendizaje automático. En la etapa de fusión de aprendizaje automático, se aplican reglas difusas a la salida real de los resultados de SVM y ANN para la predicción final. Luego, el modelo entrenado fusionado se almacena en la nube.

La segunda fase del marco propuesto se refleja en la capa de prueba. La capa de prueba adquiere un conjunto de datos de la base de datos médica y carga el modelo de entrenamiento preprocesado desde la nube. Se utiliza un modelo fusionado para predecir si un diagnóstico de diabetes es positivo o negativo. La precisión de la predicción se calcula comparando la salida requerida con la salida real.

El modelo ANN se entrena con el conjunto de datos de entrenamiento preprocesado. Hemos dividido los datos preprocesados en datos de entrenamiento y prueba con una proporción de 70:30 según la división de la base de clases. Para entrenar los datos, hemos utilizado la función de regularización bayesiana: el 5% se usa para pruebas y el 5% para validación, y el 90% restante se usa para entrenamiento.

Hay 16 capas ocultas entre las neuronas de entrada y salida. Donde 1, 2, 3, ..., 16 y 1, 2, 3, ..., 16 representan las neuronas de la capa de entrada y las neuronas de la capa oculta, respectivamente. La salida se representa como "salida". Un sesgo se puede representar como h_1 y h_2 . Esto produce una salida y fuera $\%$ basado en las siguientes ecuaciones 1 y 2.

$$afuera_{\theta} = \frac{1}{1 + \sum_{i=1}^n \frac{mi}{h_1^{2 \sum_{r=1}^n (tu_{r, \theta} * \omega)}}} \quad (1)$$

dónde, $\theta = 1, 2, \dots, \text{norte}$

$$afuera_{\%} = \frac{1}{1 + \sum_{i=1}^n \frac{mi}{h_2^{2 \sum_{\theta=1}^n (pag_{\theta, \%} * afuera_{\theta})}}} \quad (2)$$

dónde, $\% = 1, 2, \dots, r$.

Usando la función de error al cuadrado, el error de cada neurona de salida se puede calcular y sumar para encontrar el error total (E).

$$mi = \frac{1}{2} \sum_{\%} (\tau_{\%} - afuera_{\%})^2$$

Los pesos se pueden cambiar según el error usando la fórmula de la Ecuación.4

$$1\omega_{\alpha} = \frac{\partial mi}{\partial \omega} \quad (4)$$

La ecuación 5 actualiza el peso entre una capa oculta y una capa de salida.

$$1\text{pag}_{\theta, \%} = \epsilon \frac{\partial mi}{\partial \theta, \%} \quad (5)$$

Como $\theta, \%$ no se puede calcular directamente, así que utilice la ecuación. 6 fórmulas.

$$1\text{pag}_{\theta, \%} = -\epsilon \frac{\partial mi}{\partial afuera_{\%}} \times \frac{\partial afuera_{\%}}{\partial \text{neto}_{\%}} \times \frac{\partial \text{neto}_{\%}}{\partial \text{pag}_{\theta, \%}} \quad (6)$$

dónde $\tau_{\%}$ representa el peso real de $\%$ como se describe en la ecuación. 7.

$$1\text{pag}_{\theta, \%} = \epsilon \left(\tau_{\%} - afuera_{\%} \right) \times \left(1 - afuera_{\%} \right) \quad (7)$$

Las ecuaciones 8 y 9 describen cómo se actualizan los pesos en blanco y negro de las neuronas de la capa oculta y de las neuronas de la capa de entrada.

$$1\tilde{u}_{i, \theta} = \left[\sum_{\%} \frac{\partial mi}{\partial afuera_{\%}} \times \frac{\partial afuera_{\%}}{\partial \text{neto}_{\%}} \times \frac{\partial \text{neto}_{\%}}{\partial \tilde{u}_{i, \theta}} \right] \times \left[\frac{\partial afuera_{\theta}}{\partial \text{neto}_{\theta}} \times \frac{\partial \text{neto}_{\theta}}{\partial \tilde{u}_{i, \theta}} \right] \quad (8)$$

$$1\tilde{u}_{i, \theta} = \xi \left[\sum_{\%} \left(\tau_{\%} - afuera_{\%} \right) \times \left(1 - afuera_{\%} \right) \right] \times \text{pag}_{i, \theta} \times \left(1 - afuera_{\%} \right) \times \left(1 - afuera_{\%} \right) \quad (9)$$

La fórmula de actualización de pesos entre las neuronas de la capa oculta y de salida se describe en la Ecuación.10.

$$1\tilde{u}_{i, \theta}(t+1) = i, \theta(t) + \lambda 1\tilde{u}_{i, \theta} \quad (10)$$

Una vez que el modelo de entrenamiento se haya entrenado exitosamente, se debe guardar y validar con el 30% de los conjuntos de datos restantes. Cuando se guardan los resultados, la salida de los datos de validación se compara con la salida real y se encuentra que la predicción es del 92,31%.

SVM genera un hiperplano que clasifica los datos según clases. SVM clasifica los síntomas de la diabetes en Positivos y Negativos [15], [17], [18]. La separación de clases en un hiperplano comienza trazando una línea. La ecuación lineal se puede expresar en la Ecuación.11.

$$\dot{x} = \dot{x} + b \quad (11)$$

donde \dot{x} indica la pendiente y b representa un punto de intersección. Por lo tanto, se escribe de la siguiente manera:

$$aX_1 - x_2 + \text{segundo} = 0 \quad (12)$$

Si $\ddot{x} = (X_1, X_2)$ y $\ddot{\omega} = (-a, \text{segundo})$, luego usando lo anterior expresión, podemos formular una ecuación. 13.

$$\ddot{\omega} \cdot \ddot{x} + \text{segundo} = 0 \quad (13)$$

(3) La ecuación del hiperplano se puede utilizar para analizar una ecuación de tres vector dimensional. En la ecuación. 14, el vector de $\ddot{x} = (X_1, X_2)$ está representado por $\ddot{\omega}$.

$$\ddot{\omega} = \frac{X_1}{\|x\|} + \frac{X_2}{\|x\|} \quad (14)$$

En la Ecuación.15, se muestra cómo se pueden escribir vectores n -dimensionales.

$$\ddot{\omega} \cdot \ddot{x} = \sum_{i=1}^n \omega_i x_i \quad (15)$$

dónde, $i = 1, 2, \dots, \text{norte}$

La ecuación.15 permite comprobar si los datos se han clasificado correctamente.

$$-D_i = i \ddot{y} \cdot \ddot{\omega} \cdot \ddot{x} + b$$

Los márgenes funcionales de los conjuntos de datos se denominan d y se expresan como

$$d = \min_{i=1, \dots, metro} D_i$$

El margen geométrico d del conjunto de datos proporciona el hiperplano que será el hiperplano óptimo con la función lagrangiana

$$\begin{aligned} Y(\omega, b, B) &= \frac{1}{2} \omega \cdot \omega - \sum_{i=1}^{metro} B_i \left[\left(\omega \cdot \dot{x} + b \right) - y_i \right] \\ \nabla_{\omega} Y(\omega, b, B) &= \omega - \sum_{i=1}^{metro} B_i y_i \dot{x}_i = 0 \\ \nabla_b Y(\omega, b, B) &= - \sum_{i=1}^{metro} B_i y_i = 0 \end{aligned}$$

Después de la simplificación, se puede escribir como

$$\omega = \sum_{i=1}^{metro} B_i y_i \dot{x}_i \quad \& \quad \sum_{i=1}^{metro} B_i y_i = 0$$

La función lagrangiana Y es sustituido.

$$\omega(B, b) = \sum_{i=1}^{metro} B_i - \frac{1}{2} \sum_{i=1}^{metro} \sum_{j=1}^{metro} B_i B_j \dot{x}_i \cdot \dot{x}_j$$

Por lo tanto, la ecuación 20 también se puede utilizar para definir la ecuación anterior.

$$\sum_{i=1}^{metro} B_i - \frac{1}{2} \sum_{i=1}^{metro} \sum_{j=1}^{metro} B_i B_j \dot{x}_i \cdot \dot{x}_j = 0 \quad (20)$$

donde $i = 1, 2, 3, \dots, metro$

Para evitar desigualdades de contención, aplique la condición KKT (Karush-Kuhn-Tucker) al procedimiento del multiplicador lagrangiano.

$$B_i y_i \left[\left(\omega \cdot \dot{x} + b \right) - y_i \right] = 0$$

donde \dot{x} representa un punto óptimo y el valor positivo; mientras \dot{x} es que para otros puntos es casi nulo. En consecuencia, la ecuación se puede escribir como Ecuación.22.

$$[y_i (\omega \cdot \dot{x} + b - 1)] = 0 \quad (22)$$

Los puntos más cercanos al hiperplano también se conocen como vectores de soporte. Basado en la Ecuación 23,

$$\omega - \sum_{i=1}^{metro} B_i y_i \dot{x}_i = 0 \quad (23)$$

En otras palabras, se puede escribir como

$$\omega = \sum_{i=1}^{metro} B_i y_i \dot{x}_i$$

Ecuación. 24 obtiene el valor de b cuando lo calculamos.

$$y_i (\omega \cdot \dot{x}_* + b - 1) = 0 \quad (25)$$

Ambos lados de la ecuación se multiplican por y_i

$$y_i (\omega \cdot \dot{x}_* + b - 1) = 0 \quad (26)$$

Se sabe que y_i^2 es igual a 1.

$$\text{segundo} = y_i - [\omega \cdot \dot{x}_*] \quad (27)$$

$$\text{segundo} = \left\{ \sum_{i=1}^S (y_i - [\omega \cdot \dot{x}_*]) \right\} \quad (28)$$

La ecuación 27 determina no. de vectores de apoyo S , y predicciones que se realizan en base al hiperplano.

En la Ecuación.28, la función de hipótesis se describe donde ω representan el peso óptimo.

$$h(\omega_i) = \begin{cases} +1 & \text{si } (\omega_i \cdot \dot{x} + b) \geq 0 \\ -1 & \text{si } (\omega_i \cdot \dot{x} + b) < 0 \end{cases} \quad (29)$$

- (18) Cuando los puntos están por encima del hiperplano, es decir $+1$, representa diabetes positiva y los puntos están debajo del hiperplano, es decir -1 , representa diabetes negativa. Utilizamos el mismo conjunto de datos con SVM y con ANN. Los datos se entrenan usando (19) y optimizando todos los parámetros disponibles de SVM en Matlab R2020a. El proceso de validación cruzada quíntuple divide los datos en cinco niveles y los valida en consecuencia.

La lógica difusa utiliza funciones de membresía. El sistema difuso utiliza salidas SVM y ANN como variables de entrada. Las funciones de membresía definen el conjunto de reglas que se aplican tanto a las entradas como a las salidas. ANN y SVM utilizaron lógica difusa para determinar si los síntomas de un paciente coinciden o no con un diagnóstico de diabetes. Una decisión matemáticamente difusa podría describirse de la siguiente manera:

$$\begin{aligned} \zeta_{ANA} \cap \zeta_{SVM}(ANA, SVM) \\ = \min [\zeta_{ANA}(ANA), \zeta_{SVM}(SVM)] \quad (30) \end{aligned}$$

La función de membresía de ANN se define como ζ_{ANA} el de SVM como ζ_{SVM} . Según los resultados, los parámetros de resultado para ANN y SVM son 0 o 1. Dos resultados posibles de cada modelo producen cuatro conjuntos de reglas, que son

(21) que se proporciona a continuación.

- ❖ Si el resultado del modelo ANN es Positivo (0) y el resultado del modelo SVM es Positivo (0), entonces la diabetes es Positiva (0).
- ❖ Si el resultado del modelo ANN es Positivo (0) y el resultado del modelo SVM es Negativo (1), entonces la diabetes es Positiva (0).
- ❖ Si el resultado del modelo ANN es Negativo (1) y el resultado del modelo SVM es Positivo (0), entonces la diabetes es Negativa (1).
- ❖ Si el resultado del modelo ANN es Negativo (1) y el modelo SVM El resultado es Negativo (1), entonces la diabetes es Negativa (1).

La predicción individual de ANN y SVM en términos de positivo o negativo se envía al módulo de lógica difusa, que consta de cuatro reglas (discutidas anteriormente) reflejadas en la función de membresía difusa en la Tabla 1, donde se refleja (24) Función de membresía de SVM y refleja la función de membresía de ANN. La lógica difusa incorpora la fusión de niveles de decisión para la predicción final de si el paciente es diabético.

O no. Un algoritmo para inferencia difusa se puede expresar mediante μ_{mi} tu. que puede describirse como

$$\mu_{mi} = \frac{\mu_{mi} \cdot \mu_{mi}}{\mu_{mi} + \mu_{mi}} \quad (31)$$

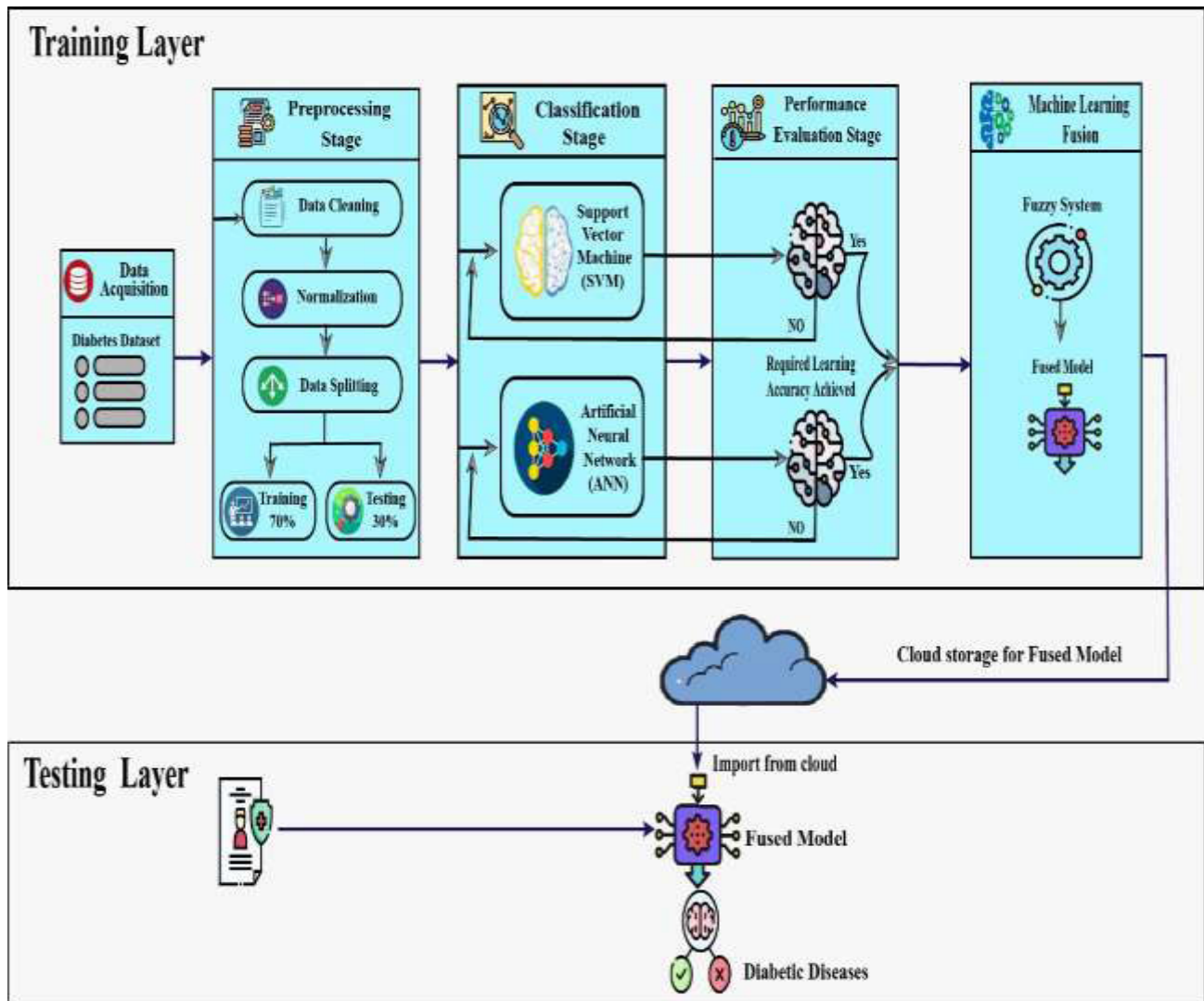


FIGURA 1. Modelo fusionado propuesto para la predicción de la diabetes (FMDP).

$$\zeta_{ANASVM} = \zeta_{SVM} \cap \zeta_{ANA} \quad (32)$$

Los puntos nítidos se analizan en la Ecuación 35.

relación difusa μ_{ζ} se define usando las reglas.

$$\mu_{\zeta} = \bigcup_{mi=1}^4 \mu_{\zeta}^{mi} \quad (33)$$

$$\mu_{\zeta}(D_{\text{decisión}}) = \max_{i=1}^4 \mu_{\zeta}^{mi}(\zeta_{ANASVM})$$

Se puede aplicar un desenfoque utilizando varios tipos de métodos, como el método del centro de área (COA), el método del promedio ponderado; el método de la media de máximos (MOM) y el principio de membresía máxima, pero el modelo propuesto aplica el método del centroide de-fuzzier. El motor de interfaz produce resultados difusos que se transforman utilizando funcionalidades similares a las del más difuso para generar resultados frangibles.

$$F = \frac{\mu_{\zeta}(\zeta_{SVM}) \cap \mu_{\zeta}(\zeta_{ANA})}{\mu_{\zeta}(\zeta_{SVM}) \cup \mu_{\zeta}(\zeta_{ANA})} \quad (35)$$

El gráfico de la figura 2 describe que los ejes x y corresponden a SVM y ANN, mientras que z indica el sistema FMDP. La superficie de reglas de diabetes del sistema FMDP se puede ver en comparación con los resultados de ANN y SVM. El resultante

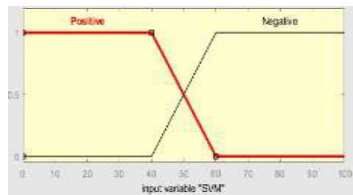
(34) El sistema FMDP no predice diabetes si ambas soluciones no predicen diabetes; Y si ambos modelos predicen que la diabetes es sí, entonces El FMDP también predice que la diabetes es sí.

La Figura 3 muestra que si ANN diagnostica sí (0) diabetes y SVM diagnostica no (1) diabetes, entonces el modelo fusionado también diagnostica sí (0) diabetes.

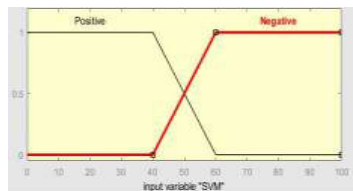
La Figura 4 muestra que si ANN no diagnostica (1) diabetes y SVM no diagnostica (1) diabetes, entonces el modelo fusionado tampoco diagnostica (1) diabetes.

TABLA 1. Funciones de membresía difusas del sistema FMDF.

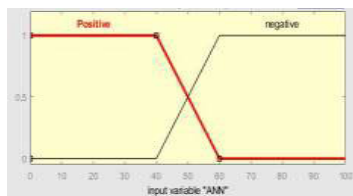
$$SVM_{(\zeta_{SVM(3)})} \quad \zeta_{(Positive)}(\tilde{z}) = \left\{ \max \left(\min \left(1, \frac{60 - \tilde{z}}{20} \right), 0 \right) \right\}$$



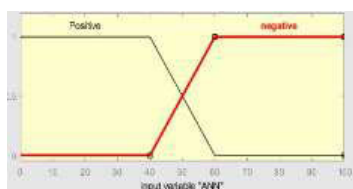
$$SVM_{(\zeta_{SVM(3)})} \quad \zeta_{(Negative)}(\tilde{z}) = \left\{ \max \left(\min \left(\frac{\tilde{z} - 40}{20}, 1 \right), 0 \right) \right\}$$



$$ANN_{(\zeta_{ANN(5)})} \quad \zeta_{(Positive)}(\tilde{b}) = \left\{ \max \left(\min \left(1, \frac{60 - \tilde{b}}{20} \right), 0 \right) \right\}$$



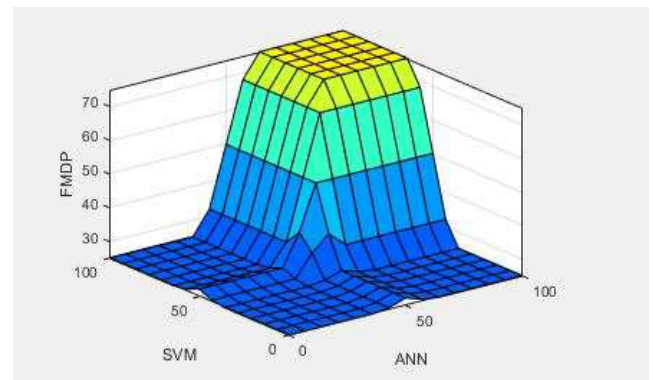
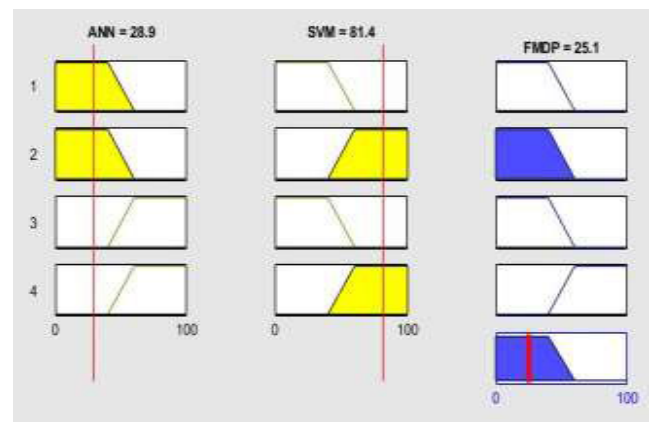
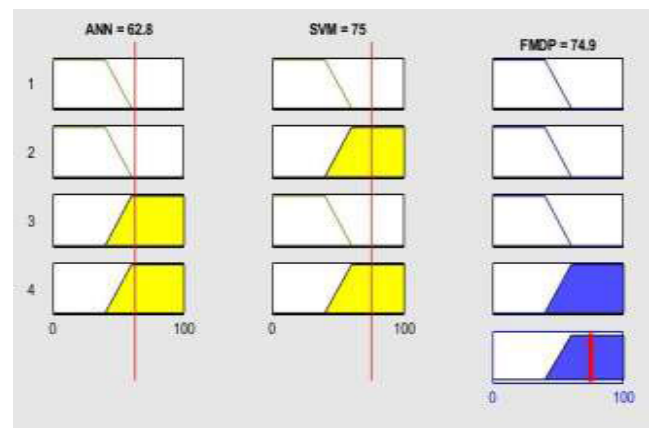
$$ANN_{(\zeta_{ANN(5)})} \quad \zeta_{(Negative)}(\tilde{b}) = \left\{ \max \left(\min \left(\frac{\tilde{b} - 40}{20}, 1 \right), 0 \right) \right\}$$



En el marco propuesto, la capa de validación se relaciona con el diagnóstico y la clasificación en tiempo real de un diabético. El modelo de ML fusionado propuesto puede utilizar datos de pacientes en tiempo real como entrada y mejorar el sistema de detección de enfermedades.

IV. RESULTADOS Y DISCUSIÓN

Para implementar el marco propuesto, utilizamos un conjunto de datos [3] donde el número total de casos es 520 y tiene 17 atributos basados en los síntomas de la diabetes. Dieciséis características son independientes y una es la característica objetivo (dependiente). La característica objetivo está etiquetada como clase, que tiene dos valores 0 o 1. La clase 0 representa que la persona tiene síntomas diabéticos (Positivos) y la clase 1 representa que

**FIGURA 2.** Superficie de reglas del sistema FMDF propuesta.**FIGURA 3.** Resultado del sistema FMDF propuesto con diabetes (sí).**FIGURA 4.** Resultado del sistema FMDF propuesto con diabetes (no).

la persona no presenta síntomas de diabetes (Negativo). La primera característica del conjunto de datos es la edad en la que 93 personas tienen entre 20 y 35 años, 138 personas tienen entre 36 y 45 años, 149 personas tienen entre 46 y 55 años, 89 personas tienen entre 56 años a 65 años, y 51 personas tienen más de 65 años. La segunda característica es Sexo en la que 382 son hombres y 192 son mujeres. Masculino se refleja con "0" y Femenino se refleja con "1". La tercera característica es

TABLA 2.(Training) matriz de confusión para RNA.

Input Values	Output Results	
	Positive $\hat{O}R_0$	Negative $\hat{O}R_1$
Positive ($\hat{E}R_0 = 246$)	236	10
Negative ($\hat{E}R_1 = 118$)	11	107

el síntoma Poliuria, que tiene 258 valores como "sí" y reflejados por "0" y 262 valores como "no", reflejados por "1". La cuarta característica es el síntoma de polidipsia, que tiene 233 "sí" como "0" y 287 "no" como "1". La quinta característica es el síntoma de pérdida repentina de peso, que tiene 217 "sí" como "0" y 303 "no" configurados como "1". La sexta característica es el síntoma de debilidad, que tiene 305 "sí" establecidos en "0" y 215 "no" establecidos en "1". La séptima característica es el síntoma de polifagia, que tiene 237 "sí" establecidos en "0" y 283 "no" establecidos en "1". La octava característica es el síntoma de candidiasis genital, que tiene 116 "sí" establecidos en valores "0" y 404 "no" establecidos en valores "1". La novena característica es el síntoma Visual Blurring que tiene 233 valores como "sí" y establecidos en "0" y 287 como "no" establecidos en "1". La décima característica es el síntoma de picazón, que tiene 253 como "sí" establecido en "0" y 267 tiene "no" establecido en valores "1". La undécima característica es el síntoma de irritabilidad, que tiene 126 valores como "sí" y establecidos en "0" y 394 valores como "no" establecidos en "1". La duodécima característica es el síntoma de curación retardada que tiene 239 valores "sí" y establecidos en "0" y 281 valores como "no" establecidos en "1". La decimotercera característica es el síntoma de paresia parcial que tiene 224 valores como "sí" establecidos en "0" y 296 valores como "no" establecidos en "1" v. La decimocuarta característica es la rigidez muscular. síntoma que tiene 195 valores como "sí" establecidos en "0" y 325 valores "no" establecidos en "1". La decimoquinta característica es el síntoma de alopecia, que tiene 179 valores como "sí" establecidos en "0" y 341 "no" establecidos en "1". La decimosexta característica es el síntoma de obesidad, que tiene 88 "sí" establecidos en valores "0" y 432 "no" establecidos en valores "1".

En este artículo, se utiliza MATLAB R2020a con fines de simulación. ANN y SVM se utilizan para la predicción, mientras que la lógica difusa se utiliza para la toma de decisiones. El conjunto de datos se divide en conjuntos de datos de entrenamiento y prueba con una proporción de 70:30. Hay 364 instancias que se utilizan en el entrenamiento de ANN y su matriz de confusión se muestra en la Tabla 2.

Mesa. 2 describe los 246 casos positivos, de los cuales 236 casos se predijeron con precisión, mientras que 10 casos se predijeron incorrectamente. Sin embargo, hay 118 casos negativos, de los cuales 107 casos se predicen con precisión, mientras que 11 casos se predicen incorrectamente.

Hay 156 casos en los datos de prueba. La matriz de confusión de las pruebas de ANN se muestra en la Tabla 3.

Mesa. 3 describe los 69 casos con diabetes positiva, de los cuales 61 casos se predijeron con precisión, mientras que 8 los casos son predichos incorrectamente. Sin embargo, el total de 87 casos de Negativo de los cuales 83 casos se predicen con precisión, mientras que 4 casos se predicen incorrectamente.

TABLA 3.(Testing) matriz de confusión para RNA.

Input Values	Output Results	
	Positive $\hat{O}R_0$	Negative $\hat{O}R_1$
Positive ($\hat{E}R_0$)	61	8
Negative ($\hat{E}R_1 = 87$)	4	83

TABLA 4. (Capacitación) matriz de confusión para SVM.

Input Values	Output Results		
	Total= 364	Positive \hat{OR}_0	Negative \hat{OR}_1
	Positive ($\hat{ER}_0 = 246$)	227	19
	Negative ($\hat{ER}_1 = 118$)	13	105

TABLA 5. (Prueba) matriz de confusión para SVM.

Input Values	Output Results		
	Total= 156	Positive	Negative
		\hat{OR}_0	\hat{OR}_1
Positive ($\hat{ER}_0 = 69$)	59	10	
Negative ($\hat{ER}_1 = 87$)	7	80	

TABLA 6. (Prueba) matriz de confusión para FMDP.

Input Values	Output Results	
	Positive $\hat{O}R_0$	Negative $\hat{O}R_1$
Positive ($\hat{E}R_0$)	64	5
Negative ($\hat{E}R_1 = 87$)	3	84

Hemos utilizado cinco validaciones cruzadas para SVM. La matriz de confusión del entrenamiento SVM se muestra en la Tabla 4.

Mesa. 4 describe los 246 casos de diabetes positiva, de los cuales 227 casos se predijeron con precisión, mientras que 19 casos se predijeron incorrectamente. Sin embargo, hay 118 casos de diabetes negativa, de los cuales 105 casos se predicen con precisión, mientras que 13 casos se predicen incorrectamente. Hay 156 instancias que se utilizan en las pruebas de SVM y su matriz de confusión se muestra en la Tabla 5.

Mesa. 5 describe los 69 casos de diabetes positiva, de los cuales 59 casos se predijeron con precisión, mientras que 10 casos se predijeron incorrectamente. Sin embargo, hay 87 casos de diabetes negativa, de los cuales 80 casos se predijeron con precisión, mientras que 7 casos se predijeron incorrectamente.

TABLA 7. Resultados de ANN, SVM y FMDP propuesto.

	SVMs Training	SVMs Testing	ANNs Training	ANNs Testing	FMDP Testing
Accuracy	0.9121	0.8910	0.9423	0.9231	0.9487
Miss Rate	0.0879	0.109	0.0577	0.0769	0.0513
Sensitivity	0.9458	0.8939	0.9555	0.9385	0.9552
Specificity	0.8468	0.8889	0.9145	0.9121	0.9438
Positive Prediction Vlaue	0.9228	0.8551	0.9593	0.8841	0.9275
Negative Prediction Vlaue	0.8898	0.9195	0.9068	0.9540	0.9655
False Positive Rate	0.1532	0.1111	0.0855	0.0879	0.0562
False Negative Rate	0.0542	0.1061	0.0445	0.0615	0.0448
F1 Score	0.9342	0.8741	0.9574	0.9104	0.9412

TABLA 8. Comparación del FMDP con técnicas de última generación.

Authors/Papers	Approach	Accuracy (%)	Miss Rate (%)
S. Kumari et al [5]	Ensemble soft voting classifier	79.08 %	20.92%
M. A. Sarwar et al [6]	KNN & SVM	77%	23%
S. K. Dey et al [7]	ANN with Min-Max scaling	82.35 %	17.65%
A. Mir et al [8]	SVM	79.13%	20.87%
S. Saru et al [9]	Decision Tree after bootstrapping	94.4%	5.6%
P. Sonar et al [10]	Decision Tree	85%	15%
S. Wei et al [11]	Deep Neural Network	77.86%	22.14%
M. F. Faruque et al [12]		73.5%	26.5%
B. Jain et al [13]	Neural Network	87.88%	
Proposed Model	Fused ML Decision	94.87%	5.13%

Mesa. 6 reflexiona Cts thmi Confusi oh. atriode tepicar con
proponer d fusionado metmodelo. Él reflejar es el 69 casos de positivo
diabetes, desde W que 64 eran pre dictar d acuratel y, mientras que
5 fueron p re dictado l incorrecto yo En th y otros en a y, el somos totales

87 casos de diabetes negativa, de los cuales 84 casos se predijeron con precisión y 3 casos se predijeron incorrectamente.

En las fórmulas que se dan a continuación, $\hat{O}R_0$, $\hat{O}R_1$, $\hat{E}R_0$ y $\hat{E}R_1$ reflejan la producción positiva prevista, la producción negativa prevista, los resultados positivos esperados y los resultados negativos esperados, respectivamente.

$$\hat{A}precisión = \frac{(\hat{O}R + \hat{E}R_1 / \hat{E}R_1)}{(\hat{E}R_0 + \hat{E}R_1)} \quad (36)$$

$$\text{METRO es (comió)} = \frac{(\hat{O}R_1 / \hat{E}R_0 + \hat{O}R_0 / \hat{E}R_1)}{(\hat{E}R_0 + \hat{E}R_1)} \quad (37)$$

$$\text{Valor de predicción 'positivo'} = \frac{(\hat{O}R + \hat{E}R_1)}{(\hat{O}R + \hat{E}R + \hat{O}R_0 / \hat{E}R)} \quad (38)$$

$$\text{Valor de predicción negativo} = \left(\frac{\hat{O}R_0 / \hat{E}R_0}{\hat{O}R_0 / \hat{E}R_0 + \hat{O}R_1 / \hat{E}R_0} \right) \quad (39)$$

$$\text{especificidad} = \frac{(\hat{O}R_0 / \hat{E}R_0)}{(\hat{O}R + \hat{E}R_1 + \hat{O}R_0 / \hat{E}R_1)} \quad (40)$$

$$\text{Sensibilidad} = \frac{(\hat{O}R_1 / \hat{E}R_1)}{(\hat{O}R + \hat{E}R_0 + \hat{O}R_1 / \hat{E}R_1)} \quad (41)$$

$$\text{Tasa de descubrimiento falso} = \frac{(\hat{O}R_0 + \hat{E}R)}{(\hat{O}R + \hat{E}R + \hat{O}R_0 / \hat{E}R_0)} \quad (42)$$

$$\text{False 'Tasa positiva} = 1 - \hat{S}especificidad \quad (43)$$

$$\text{False 'Tasa negativa} = 1 - \text{sensibilidad} \quad (44)$$

$$\hat{F}1\hat{S}núcleo = 2 * \frac{\text{Valor de predicción 'positivo' * } \hat{S}sensibilidad}{\text{'Positiva' Predicción Vlaue} + \hat{S}sensibilidad} \quad (45)$$

El rendimiento de ambos modelos (ANN y SVM) junto con el modelo fusionado propuesto se evalúa utilizando varias medidas de precisión como se analizó anteriormente y se refleja en la Tabla 7. Se puede ver que el modelo fusionado propuesto tuvo un mejor rendimiento en las pruebas de datos en comparación con los dos modelos utilizados (ANN, SVM).

El modelo fusionado propuesto también se compara con modelos y técnicas publicados anteriormente en la Tabla 8. Se puede observar que la técnica fusionada propuesta superó a todas las otras técnicas publicadas y logró una precisión del 94,87% y una tasa de fallos del 5,13%.

CONCLUSIÓN V

Aunque se han utilizado diferentes modelos para la predicción de la diabetes, la precisión de los modelos propuestos en la predicción de enfermedades siempre ha sido la principal preocupación de los investigadores. Por lo tanto, se requiere un nuevo modelo para lograr una mayor precisión en la predicción de la diabetes. Esta investigación

propuso un sistema de apoyo a la toma de decisiones sobre diabetes basado en aprendizaje automático mediante el uso de fusión de niveles de decisión. En el modelo propuesto se integran dos técnicas de aprendizaje automático ampliamente utilizadas mediante el uso de lógica difusa. El sistema de decisión difusa propuesto ha logrado una precisión de 94,87, que es superior a la de otros sistemas existentes. A través de este modelo de diagnóstico podemos salvar varias vidas. Además, la tasa de mortalidad por diabetes se puede controlar si la enfermedad se diagnostica y se toman medidas preventivas en una etapa temprana.

REFERENCIAS

- [1] F. Islam, R. Ferdousi, S. Rahman y HY Bushra, *Visión por computadora e inteligencia artificial en el análisis de imágenes médicas*. Londres, Reino Unido: Springer, 2019.
- [2] Organización Mundial de la Salud (OMS). (2020). *La OMS revela las principales causas de muerte y discapacidad en todo el mundo: 2000-2019*. Consultado: 22 de octubre de 2021. [En línea]. Disponible: <https://www.who.int/news/item/09-12-2020-whoreveals-leading-causes-of-death-and-disability-worldwide-2000-2019>
- [3] A. Frank y A. Asunción. (2010). *Repositorio de aprendizaje automático de la UCI*. Consultado: 22 de octubre de 2021. [En línea]. Disponible: <http://archive.ics.uci.edu/ml>
- [4] G. Pradhan, R. Pradhan y B. Khandelwal, "Un estudio sobre varios algoritmos de aprendizaje automático utilizados para la predicción de la diabetes mellitus", en *Técnicas y aplicaciones de informática blanda* (Avances en sistemas inteligentes y computación), vol. 1248. Londres, Reino Unido: Springer, 2021, págs. 553–561, doi:10.1007/978-981-15-7394-1_50.
- [5] S. Kumari, D. Kumar y M. Mittal, "Un enfoque conjunto para la clasificación y predicción de la diabetes mellitus utilizando un clasificador de votación suave". En *t. J. Cogn. Computadora. Ing.*, vol. 2, págs. 40–46, junio de 2021, doi: 10.1016/j.ijcce.2021.01.001.
- [6] MA Sarwar, N. Kamal, W. Hamid y MA Shah, "Predicción de la diabetes mediante algoritmos de aprendizaje automático en la atención sanitaria", en *Proc. 24^o Int. Conf. Automático. Computadora. (ICAC)*, septiembre de 2018, págs. 6 y 7, doi: 10.23919/ICAC.2018.8748992.
- [7] SK Dey, A. Hossain y MM Rahman, "Implementación de una aplicación web para predecir la enfermedad de diabetes: un enfoque que utiliza un algoritmo de aprendizaje automático", en *Proc. 21^o Int. Conf. Computadora. inf. Tecnología. (TICCI)*, diciembre de 2018, págs. 21–23, doi:10.1109/ICCICTECHN.2018.8631968.
- [8] A. Mir y SN Dhage, "Predicción de la enfermedad de diabetes mediante aprendizaje automático en big data de atención médica", en *Proc. 4^{to} Int. Conf. Computadora. Comunitario. Control automático. (ICCUBE)*, agosto de 2018, págs. 1 a 6, doi: 10.1109/ICCUBE.2018.8697439.
- [9] S. Saru y S. Subashree. *Análisis y predicción de la diabetes mediante aprendizaje automático*. Consultado: 22 de octubre de 2022. [En línea]. Disponible: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3368308
- [10] P. Sonar y K. JayaMalini, "Predicción de la diabetes utilizando diferentes enfoques de aprendizaje automático", en *Proc. 3^o Int. Conf. Computadora. Metodologías Comun. (ICCMC)*, marzo de 2019, págs. 367–371, doi: 10.1109/ICCMC.2019.8819841.
- [11] S. Wei, X. Zhao y C. Miao, "Una exploración integral de las técnicas de aprendizaje automático para la identificación de la diabetes", en *Proc. IEEE 4^{to} Foro Mundial Internet Things (WF-IoT)*, febrero de 2018, págs. 291–295, doi: 10.1109/WF-IoT.2018.8355130.
- [12] MF Faruque e IH Sarker, "Análisis de rendimiento de técnicas de aprendizaje automático para predecir la diabetes mellitus", en *Proc. En t. Conf. Electricidad, Informática. Comunitario. Ing. (ECCE)*, febrero de 2019, págs. 7 a 9, doi: 10.1109/ECACE.2019.8679365.
- [13] B. Jain, N. Ranawat, P. Chittora, P. Chakrabarti y S. Poddar, "Una perspectiva del aprendizaje automático: para analizar la diabetes". *Madre. Hoy: Proc.*, págs. 1 a 5, febrero de 2021, doi:10.1016/j.MATPR.2020.12.445.
- [14] NB Padmavathi, "Estudio comparativo de clasificadores kernel SVM y ANN para la clasificación de neoplasias cerebrales", en *Proc. En t. Conf. Intel. Comput., Instrumento. Tecnología de control. (ICICT)*, julio de 2017, págs. 469–473, doi: 10.1109/ICICT1.2017.8342608.
- [15] J. Liu, J. Feng y X. Gao, "Diagnóstico de fallas de pozos de bombeo con varilla basado en una máquina de vectores de soporte optimizada mediante una optimización mejorada del enjambre de pollos". *Acceso IEEE*, vol. 7, págs. 171598–171608, 2019, doi: 10.1109/ACCESO.2019.2956221.
- [16] P. Chinas, I. López, JA Vázquez, R. Osorio y G. Lefranc, "Aplicación de SVM y ANN al reconocimiento de patrones multivariados utilizando datos dispersos". *IEEE Latinoamérica. Trans.*, vol. 13, núm. 5, págs. 1633–1639, mayo de 2015, doi: 10.1109/TLA.2015.7112025.
- [17] Y. Yang, J. Wang e Y. Yang, "Mejora del clasificador SVM con conocimientos previos en la detección de microcalcificaciones", en *Proc. 19^o Instituto Internacional del IEEE. Conf. Proceso de imagen.*, septiembre de 2012, págs. 2837–2840, doi: 10.1109/ICIP.2012.6467490.
- [18] J. Liu, J. Feng, Q. Xiao, S. Liu, F. Yang y S. Lu, "Diagnóstico de fallas de un pozo de petróleo con bomba de varilla basado en una máquina de vectores de soporte utilizando un diagrama indicador preprocesado", en *Proc. IEEE décimo aprendizaje sobre control basado en datos. Sistema. Conf. (DDCL)*, mayo de 2021, págs. 120–126, doi: 10.1109/DDCLS52934.2021.9455702.
- [19] P. Patil, N. Yaligar y SM Meena, "Comparación del rendimiento de los clasificadores: SVM, RF y ANN en la detección de la enfermedad del tizón de la papa mediante imágenes de hojas", en *Proc. IEEE Internacional. Conf. Computadora. Intel. Computadora. Res. (ICCIC)*, diciembre de 2017, págs. 1 a 5, doi:10.1109/ICCIC.2017.8524301.



USAMA AHMED Recibió el título de BSCS de la Universidad de Sargodha. Actualmente está cursando una maestría en ciencias de la computación en la Escuela de Computación e Innovación de Riphah, Facultad de Computación, Universidad Internacional de Riphah, Lahore, Pakistán. Actualmente, se desempeña como Instructor/Tutor del Departamento de Ciencias de la Computación de la Universidad Virtual de Pakistán. Sus intereses de investigación incluyen el aprendizaje automático, el diagnóstico médico, la minería de datos y el procesamiento de imágenes.



GHASSAN F. ISSA Recibió la maestría y el doctorado. Licenciatura en informática/inteligencia artificial de la Universidad Old Dominion, VA, EE. UU., en 1987 y 1992, respectivamente. Fue miembro de la facultad y director del Departamento de Ciencias de la Computación en la Facultad de Tecnología de Pensilvania (Penn State), EE. UU., de 1992 a 1995. También se desempeñó como Decano de Ciencias de la Computación en la Universidad de Ciencias Aplicadas, Ammán, Jordania, desde 2003. a 2005, y al Decano para información tecnología en la Universidad de Petra, Amman, de 2008 a 2018. Es profesor de informática. Actualmente, es profesor y decano de la Escuela de Tecnología de la Información del Skyline University College, Al Sharjah, Emiratos Árabes Unidos. Sus intereses de investigación incluyen la IA y el aprendizaje automático con trabajo en el ajuste de redes neuronales profundas, el aprendizaje por analogía y algoritmos de clasificación asociativa.



MUHAMMAD ADNAN KHAN Recibió la licenciatura y la maestría en filosofía. títulos de la Universidad Islámica Internacional, Islamabad, Pakistán, y el Ph.D. Licenciatura de la Universidad ISRA, Pakistán. Actualmente trabaja como profesor asociado en la Escuela de Computación e Innovación de Riphah, Facultad de Computación, Universidad Internacional de Riphah, Lahore, Pakistán, y profesor asistente en el Laboratorio de Reconocimiento de Patrones y Aprendizaje Automático del Departamento.

de Software, Universidad Gachon, Corea del Sur. Antes de unirse a la Universidad Gachon y la Universidad Internacional Riphah, trabajó en diversos puestos académicos e industriales en Pakistán. Ha estado enseñando a estudiantes de posgrado y pregrado en ciencias de la computación e ingeniería durante los últimos 13 años. Actualmente, dirige cinco Ph.D. Académicos y siete M.Phil. eruditos. Ha publicado más de 200 artículos de investigación con un JCR-IF acumulativo de más de 370 en revistas internacionales, así como en congresos internacionales de renombre. Sus intereses de investigación incluyen el aprendizaje automático, MUD, el procesamiento de imágenes y el diagnóstico médico, y la estimación de canales en sistemas de comunicación multiportadora que utilizan informática informática. Recibió el Premio de Beca de la Junta de Tecnología e Información de Punjab, Gobierno de Punjab, Pakistán, por su licenciatura y maestría en filosofía. grados y el Premio de Beca de la Comisión de Educación Superior, Islamabad, en 2016, por su doctorado. grado.



SHABIB AFTAB (Miembro, IEEE) recibió la maestría en ciencias de la computación del Instituto COMSATS de Tecnología de la Información, Lahore, Pakistán, y la maestría. Licenciado en tecnología de la información de la Facultad de Tecnología de la Información de la Universidad de Punjab (PUCIT) Lahore. Actualmente está cursando el Ph.D. Licenciado en Ciencias de la Computación por la Facultad de Ciencias de la Computación de la Facultad Nacional de Administración de Empresas y Economía de Lahore. Actualmente, él es

Se desempeña como profesor de informática en la Universidad Virtual de Pakistán. Sus intereses de investigación incluyen la minería de datos y la mejora de procesos de software.



MUHAMMAD FARHAN KHAN recibió el M.Phil. Licenciado en ciencias forenses por la Universidad de Ciencias Veterinarias y Animales (UVAS), Lahore. Actualmente está cursando el Ph.D. Licenciada en ciencias forenses por la Universidad de Ciencias de la Salud de Lahore, Lahore, Pakistán. Actualmente, se desempeña como profesor en el Departamento de Ciencias Forenses de la Universidad de Ciencias de la Salud de Lahore. Sus intereses de investigación incluyen ciencias médicas, ciencias del comportamiento, medicina

diagnóstico, ciencias forenses, biología molecular, minería de datos y procesamiento de imágenes.



RAED EN DICH recibió el B.Sc. Licenciada en Estadística por la Universidad de Yarmouk, Irbid, Jordania, M.Sc. Licenciatura en estudios de población de la Universidad de Jordania y Ph.D. Licenciada en Estadística por la Universidad de Leeds, Reino Unido.

Ha trabajado en varias universidades de Emiratos Árabes Unidos durante más de 20 años. Antes de unirse a la Canadian University Dubai (CUD), en 2018, fue profesor asistente en la Facultad de Administración de Empresas de la Universidad Al Ain.

ciudad, donde también se desempeñó como Director del Centro de Investigación Institucional y Garantía de Calidad durante tres años. Antes de eso, enseñó metodologías de investigación y diferentes cursos de estadística y matemáticas para estudiantes de negocios, ingeniería y educación en la Universidad de Ajman. Está particularmente interesado en desarrollar métodos mejorados de extracción de datos para descubrir patrones en datos espaciotemporales. Ha publicado en revistas arbitradas y presentado artículos en numerosas conferencias internacionales sobre temas relacionados con aplicaciones de minería de datos para la sostenibilidad global y ha contribuido a los avances en ciencia, tecnología e innovación. También participa en actividades de investigación sobre tecnologías de la información y sus aplicaciones para mejorar la calidad de la educación y mejorar los métodos de evaluación y la impartición de cursos en instituciones de educación superior.



TAHER M. GHAZAL (Miembro, IEEE) recibió el B.Sc. Licenciado en ingeniería de software por la Universidad de Al Ain, en 2011, el M.Sc. Licenciatura en gestión de tecnologías de la información de la Universidad Británica de Dubai, asociada con la Universidad de Manchester y la Universidad de Edimburgo, en 2013, y el Ph.D. Licenciado en TI/ingeniería de software de la Universidad de Damasco, en 2019. Actualmente está cursando el doctorado. Licenciatura en ciencias y tecnología de la información.

con la Universiti Kebangsaan Malaysia. Tiene más de diez años de experiencia amplia y diversa como instructor, tutor, investigador, profesor, ingeniero especialista/soporte de TI y analista de sistemas/negocios. Se desempeñó en los departamentos de ingeniería, informática, TIC, jefe de STEM e innovación. También participó en control de calidad, acreditación y análisis de datos en varias instituciones educativas gubernamentales y privadas dependientes de KHDA, el Ministerio de Educación y el Ministerio de Educación Superior e Investigación Científica de los Emiratos Árabes Unidos. Sus intereses de investigación incluyen IoT, TI, inteligencia artificial, sistemas de información, ingeniería de software, desarrollo web, modelado de información de construcción, calidad de la educación, gestión, big data, calidad del software y gestión de proyectos. Participa activamente en servicios comunitarios en el campo de proyectos e investigación.



MUNIR AHMAD (Miembro, IEEE) recibió la Maestría en Ciencias de la Computación de la Universidad Virtual de Pakistán, en 2018. Actualmente está cursando el Ph.D. Licenciado en Ciencias de la Computación por la Facultad de Ciencias de la Computación, Colegio Nacional de Administración de Empresas y Economía. Ha pasado varios años en la industria. Actualmente trabaja como Director Ejecutivo/Jefe del Departamento de TI, United International Group, Lahore, Pakistán. Tiene una amplia experiencia en la gestión de datos.

ment y utilización eficiente de los recursos en las organizaciones multinacionales. Ha realizado muchos estudios de investigación sobre análisis de sentimientos y utilización de IA para predecir diversos temas de atención médica. Sus intereses de investigación incluyen minería de datos, big data e inteligencia artificial.

...