

Predicción de la diabetes potenciada con aprendizaje automático fusionado

Zambrano Garcia Gina Yadira

19 de octubre del 2023

Resumen

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.

1. 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 poliur en 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

En este artículo, discutiremos algunas técnicas clave de aprendizaje automático y sus aplicaciones.

2. Algoritmos de Aprendizaje Automático

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 Aprendizaje en conjunto, Pradhanet 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. 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% 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 Detection (CHAID) y SVM. La tasa de precisión de NN es del 87,88%, que es más alta que la de los otros cuatro algoritmos.

$$[w_0 + w_1x + w_2x^2] = 0 \quad (22)$$

Existen diversos algoritmos de aprendizaje automático, incluyendo:

$$\begin{aligned}
\text{donde } \theta &= 1, 2, \dots, n_{\text{inter}} \quad (1) \\
\text{donde } \theta &= 1, 2, \dots, n_{\text{inter}} \quad (2) \\
\text{donde } \theta &= 1, 2, \dots, n_{\text{inter}} \quad (3) \\
\text{donde } \theta &= 1, 2, \dots, n_{\text{inter}} \quad (4) \\
\text{donde } \theta &= 1, 2, \dots, n_{\text{inter}} \quad (5) \\
\text{donde } \theta &= 1, 2, \dots, n_{\text{inter}} \quad (6)
\end{aligned}$$

Figura 1: Enter Caption

3. Resultados Experimentales

4. El Resultado DEL Modelo

Si el resultado del modelo ANN es Positivo (0) y el resultado del modelo donde **representa un punto óptimo y el valor positivo; mientras es SVM es Positivo (0), entonces la diabetes es Negativa (0). que para otros puntos es casi nulo.

$$Biyi ix +b -1=0$$

En consecuencia, la ecuación se si el resultado del modelo ANN es Positivo (0) y el resultado del modelo puede escribir como Ecuación.22 en .SVM es Negativo (1), entonces la diabetes es positiva (0). 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 nesecita de 27 ecuacuines que deterinen el vector de la prediccion de la doabetes 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 5validación, y el 90Hay 16 capas ocultas entre las neuronas de entrada y salida. Donde 1, 2, 3. . 16 y 1, 2, 3 . . . 16 representa 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 comoh1 yh2. Esto produce una salida fuera basado en las siguientes ecuaciones 1 y 2

4, el algoritmo de Regresión Lineal logra una precisión del 90

5. El modelo de ML fusionado

propuesto puede utilizar datos de pacientes en tiempo real como entrada y mejorar el sistema de detección de enfermedades. FIGURA Resultado del sistema FMDP propuesto con diabetes

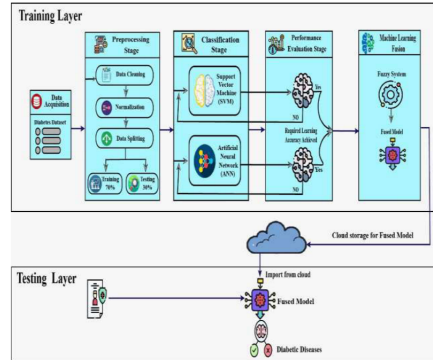


Figura 2: Enter Caption

$$y_i(\omega^T X_i + b - 1) = 0$$

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

$$\omega - \sum_{i=1}^{n_{\text{metro}}} B_i y_i X_i = 0$$

En otras palabras, se puede escribir como

$$\omega = \sum_{i=1}^{n_{\text{metro}}} B_i y_i X_i$$

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

$$y_i(\omega^T X_i + b - 1) = 0$$

Ambos lados de la ecuación se multiplican por y_i

$$y_i^2(\omega^T X_i + b - 1) = 0$$

Figura 3: Enter Caption

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

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

Input Values	Output Results		
	Total= 364	Positive OR ₀	Negative OR ₁
	Positive (ER ₀ = 246)	227	19
	Negative (ER ₁ = 118)	13	105

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

Input Values	Output Results		
	Total= 156	Positive	Negative
		$\hat{O}R_0$	$\hat{O}R_1$
Positive ($\hat{E}R_0 = 69$)	59	10	
Negative ($\hat{E}R_1 = 87$)	7	80	

ABLA 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

Figura 4: Enter Caption

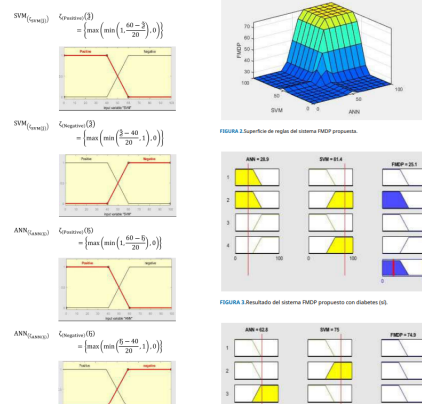


Figura 5: Enter Caption

RESULTADOS Y DISCUSIÓN 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 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

6. Materiales Y Metodos

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 al-

$$\begin{aligned} \text{Precisión} &= \frac{(\hat{R}_1/\hat{R}_0 + \hat{R}_1/\hat{R}_1)}{(\hat{R}_0 + \hat{R}_1)} \end{aligned} \quad (36)$$

$$\begin{aligned} \text{METRO estornó} &= \frac{(\hat{R}_1/\hat{R}_0 + \hat{R}_0/\hat{R}_1)}{(\hat{R}_0 + \hat{R}_1)} \end{aligned} \quad (37)$$

$$\begin{aligned} \text{Valor de predicción 'positivo'} &= \frac{(\hat{R}_1/\hat{R}_1)}{(\hat{R}_0 + \hat{R}_1 + \hat{R}_0/\hat{R}_1)} \end{aligned} \quad (38)$$

$$\begin{aligned} \text{Valor de predicción negativo} &= \frac{(\hat{R}_0/\hat{R}_0)}{(\hat{R}_0/\hat{R}_0 + \hat{R}_1/\hat{R}_0)} \end{aligned} \quad (39)$$

$$\begin{aligned} \text{especificidad} &= \frac{(\hat{R}_0/\hat{R}_0)}{(\hat{R}_0 + \hat{R}_1 + \hat{R}_0/\hat{R}_1)} \end{aligned} \quad (40)$$

$$\begin{aligned} \text{Sensibilidad} &= \frac{(\hat{R}_1/\hat{R}_1)}{(\hat{R}_0 + \hat{R}_1 + \hat{R}_1/\hat{R}_1)} \end{aligned} \quad (41)$$

$$\begin{aligned} \text{Tasa de descubrimiento falso} &= \frac{(\hat{R}_0/\hat{R}_0)}{(\hat{R}_0 + \hat{R}_1 + \hat{R}_0/\hat{R}_1)} \end{aligned} \quad (42)$$

$$\begin{aligned} \text{False 'Tasa positiva} &= 1 - \text{especificidad} \end{aligned} \quad (43)$$

$$\begin{aligned} \text{False Ntasa negativa} &= 1 - \text{sensibilidad} \end{aligned} \quad (44)$$

$$\begin{aligned} \text{f1Snúcleo} &= 2 * \frac{\text{Valor de predicción 'positivo'} * \text{Sensibilidad}}{\text{'Positiva' Predicción Vlau} + \text{Sensitividad}} \end{aligned} \quad (45)$$

rendimiento de ambos modelos (ANN u SVM) junto con el modelo

Figura 6: Enter Caption

goritmos 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 nvestigamos 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.

7. Conclusiones

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. 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.

8. 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-whorevealsleading-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:
- [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
- [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° Int. Conf. Automático. Computadora. (ICAC), septiembre de 2018,
- [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 Int. Conf. Computadora. inf. Tecnología. (TICCI), diciembre de 2018
- [8] A. Mir y SN Dhage, "Predicción de la enfermedad de diabetes mediante aprendizaje automático en big data de atención médica", en Proc. 4to Int. Conf. Computadora. Comunitario. Control automático. (ICCUBEA), agosto de 2018
- [9] S. Saru y S. Subashree. Análisis y predicción de la diabetes mediante aprendizaje automático. Consultado: 22 de octubre de 2022.
- [10] P. Sonar y K. JayaMalini, "Predicción de la diabetes utilizando diferentes enfoques de aprendizaje automático", en Proc. 3° Int. Conf. Computadora. Metodologías Comun. (ICCMC), marzo de 2019,