

► Gemelos Digitales en la Práctica: Simulación Integral del Proceso de Deshidratación en una EDAR

Gabriel David Modia^a
Alberto García^b
Raúl Ivars^c
Miriam Carrasco^c
Silvia Doñate^b
María Deseada Esclapez^b
José Luis Sánchez-Asins^a

- a) Centro Tecnológico AIDIMME, C/Leonardo Da Vinci, 38. Parque Tecnológico. 46980 Paterna (Valencia)
- b) Depuración de Aguas del Mediterráneo S.L. Av. Benjamin Franklin, 21, Parque Tecnológico. 46980, Paterna (Valencia); innovacion@dam-aguas.es
- c) UTE S y D Aguas 108, Camino Pardines s/n, 46680 Algemés (Valencia)

Gemelos Digitales en la Práctica: Simulación Integral del Proceso de Deshidratación en una EDAR

Resumen

Los gemelos digitales emergen como eficaces herramientas para la optimización de los procesos del ciclo integral del agua. Muestra de ello, este estudio presenta un piloto demostrador que se enfoca en la creación de un gemelo digital para el proceso de deshidratación en una Estación Depuradora de Aguas Residuales (EDAR) ubicada en Algemesí - Albalat de la Ribera. Utilizando tecnología de realidad virtual, este enfoque permite la simulación completa del proceso, brindando una interfaz que se asemeja a un Sistema de Control y Adquisición de Datos (SCADA). El gemelo digital posibilita la observación en tiempo real y la realización de simulaciones, permitiendo la variación virtual de las condiciones de entrada y evaluando su impacto en los resultados. En su desarrollo, se aplicaron tecnologías de inteligencia artificial, con un enfoque especial en el aprendizaje automático (*Machine Learning*), sobre los datos operativos del proceso de deshidratación. Además, se utilizaron tecnologías de representación virtual de objetos basadas en representaciones gráficas.

Palabras clave

Gemelo digital, realidad virtual, simulación, deshidratación de lodos

1. Introducción

La optimización de la eficiencia en las plantas de tratamiento de aguas residuales se ha convertido en un objetivo prioritario en la actualidad. Una vanguardia en la resolución de este desafío se encuentra en la aplicación de avances en ciencia de datos y modelización. La introducción de sensores para la medición de diversos parámetros en estas instalaciones presenta oportunidades significativas para aprovechar la abundancia de datos generados. La inteligencia artificial (IA) emerge como una

solución clave para la automatización y digitalización en el sector de aguas residuales, facilitando la creación de gemelos digitales como herramientas prometedoras en la gestión y operación eficiente del ciclo del agua (Bahramian et al., 2023). En este contexto, el centrifugado de lodos en Estaciones Depuradoras de Aguas Residuales (EDARs) desempeña un papel fundamental para mitigar los costes asociados al transporte y disposición de residuos, al mismo tiempo que contribuye a la minimización del impacto ambiental y mejora la eficiencia global del proceso. La reducción de la carga de sólidos en el tratamiento global de aguas residuales no solo optimiza la sostenibilidad del proceso, sino que también promueve prácticas respetuosas con el medio ambiente. La monitorización y modelado adecuados del proceso son esenciales para alcanzar estos objetivos, al ayudar a reducir los costes asociados al transporte y disposición de lodos, minimizar el impacto ambiental y mejorar la eficiencia del tratamiento de aguas residuales.

En el marco actual, la tecnología de gemelos digitales se destaca por la creación de réplicas virtuales precisas de objetos o procesos, permitiendo simular su comportamiento y optimizar su eficiencia mediante la utilización de datos provenientes de sensores o sistemas automatizados (Singh et. Al, 2021). Esta capacidad facilita la anticipación de problemas, la prevención de interrupciones, la identificación de nuevas oportunidades comerciales, la realización de simulaciones para la planificación futura y la adaptación de la producción según las necesidades del proceso, entre otros beneficios (Grieves and Vickers, 2017). Sin embargo, la falta de una definición precisa de gemelos digitales y de una arquitectura de referencia ha llevado al desarrollo de soluciones utilizando diversas tecnologías, interfaces, protocolos de comunicación, modelos y datos. Es esencial trabajar en el establecimiento de soluciones estándar para gemelos digitales que proporcionen criterios y restricciones de diseño claros. Esto implica la definición clara de aspectos arquitectónicos, modelos de información de referencia y protocolos de comunicación para garantizar la coherencia y la interoperabilidad en el desarrollo y la implementación de gemelos digitales. (Semeraro et al., 2021). Las herramientas de aprendizaje automático se convierten en aliados valiosos al permitir no solo la monitorización, sino

también la predicción del comportamiento del dispositivo monitorizado bajo diversas circunstancias. Además, la disponibilidad de numerosos sistemas de monitorización, sensores y PLCs (Controlador Lógico Programable) en el mercado facilita la creación de una infraestructura hardware sólida capaz de gestionar el constante flujo de información alimentando el gemelo digital (Grieves, 2014). Finalmente, una infraestructura de software eficiente resulta indispensable para adquirir, almacenar, transformar y distribuir información, permitiendo a los sistemas de aprendizaje automático aprender y adaptarse de manera efectiva (Semeraro, 2020).

2. Materiales y métodos

En el presente trabajo se ha modelizado el centrifugado de lodos en la EDAR de Algemés - Albalat (Valencia). En el proceso, el lodo fluye desde una tubería principal hacia la centrífuga, donde se introduce en un tambor de la máquina. Para facilitar la deshidratación, se añade un compuesto en base a poliácridamida catiónica conocido como polielectrolito, que fomenta la aglomeración de los sólidos. Este compuesto también se inyecta en el tambor a través de un conducto. La centrífuga opera a una velocidad constante, mientras que la velocidad diferencial se refiere a la velocidad relativa entre el tambor y el tornillo que introduce el lodo. El resultado de este proceso es la obtención de lodo deshidratado, con un porcentaje variable de deshidratación.

El estudio se implementó sobre una máquina centrifugadora *Alfa Laval NX 4000* y para poder construir el gemelo digital del proceso se obtuvieron los datos de las variables principales. Las variables de entrada incluyen el contenido en sólidos del lodo de entrada (3 - 4%), el caudal de fango ($3 - 7 \text{ m}^3 \cdot \text{h}^{-1}$), y el caudal de poli ($800 - 900 \text{ l} \cdot \text{h}^{-1}$). Las variables de funcionamiento de la máquina abarcan el valor medio de corriente (0 - 50 A), las vibraciones de rodamientos (0 a 20 $\text{mm} \cdot \text{s}^{-1}$), la velocidad diferencial (0 - 100%), el par motor (0 - 8 Nm), y la temperatura de los rodamientos (0 - 70 °C). Las variables de salida, determinadas a través de análisis de laboratorio, comprenden el peso del residuo sólido (14 - 15%), la concentración de fango espesado (0 - 5%), y la calidad del escurrido ($100 - 800 \text{ mg} \cdot \text{l}^{-1}$).

Además, se equipó al sistema con nuevos sensores, que incluyen un sensor de vibración para el rodamiento del apoyo trasero (VKV022 de IFM), 2 sensores de temperatura para ambos rodamientos (TS 2759, Pt 100, de IFM), 2 analizadores de red (DIRIS B-30 RS485-MODBUS-230VAC), y una sonda de sólidos en suspensión (LXV424.99.00100), junto con su set de montaje y su controlador universal. Los nuevos sensores fueron conectados al PLC donde se recibe el resto de datos mostrados en el SCADA, de forma que para la lectura y extracción de los datos sólo se tuviese que realizar una conexión a un PLC. Esta conexión se realizó a través de protocolo *Modbus*.

Arquitectura del modelo digital implementado

Se ha desarrollado un modelo de gemelo digital que consta de seis capas interrelacionadas. La primera es la capa física, que engloba sensores y controladores, constituyendo la interfaz directa con el entorno físico. A continuación, la capa de extracción y almacenamiento de datos a nivel local procesa y almacena la información recopilada en la capa física.

La capa de comunicaciones facilita la transferencia eficiente de datos entre la capa física y la capa de almacenamiento de datos en la nube, donde se centraliza la información para un acceso remoto y una gestión más amplia. La siguiente capa, la de análisis, se encarga de procesar los datos almacenados y derivar conclusiones o patrones significativos.

Finalmente, la capa de visualización e interacción proporciona una interfaz intuitiva para la presentación de resultados y la interacción con el gemelo digital. Este diseño en seis capas permite una representación completa y eficaz del gemelo digital, asegurando una conexión integral entre el mundo físico y su contraparte digital (Redelinghuys et al. 2019).

En las instalaciones de la planta depuradora, ya se disponía de una arquitectura preexistente para la recopilación de datos, sobre la cual se trabajó realizando todas las adaptaciones necesarias:

(a) Instalación de sensores adicionales. Se llevó a cabo la instalación de sensores adicionales, que

incluyeron dispositivos específicos como sensores de vibración para los rodamientos, sensores de temperatura, analizadores de red para evaluar el consumo y una sonda de sólidos en suspensión, con su correspondiente set de montaje y controlador universal. Estos sensores se comunicaron directamente con los PLCs de la instalación, permitiendo la transferencia eficiente de datos a la red. Esta integración aseguró la accesibilidad de los datos desde cualquier punto con acceso a la red local interna, mejorando tanto la capacidad de monitoreo como la base para la integración de datos adicionales en el sistema de recopilación de información existente en la planta depuradora.

(b) Extracción de datos del PLC existente. Se llevó a cabo la extracción de datos del PLC existente en la planta depuradora. Los PLCs empleados pertenecen a la marca *Schneider*, utilizando para la comunicación el protocolo *Modbus TCP*. Este protocolo *Modbus* facilita el intercambio de datos entre los controladores lógicos programables y los ordenadores (PC), permitiendo que los dispositivos electrónicos compartan información a través de conexiones en serie mediante este protocolo. Dado que los valores proporcionados por los sensores dependen de la resolución del convertidor analógico digital, se llevó a cabo una conversión de variables para garantizar la coherencia y precisión de los datos extraídos del PLC. Este proceso de conversión asegura que los valores recopilados sean adecuadamente interpretados y utilizados en las etapas subsiguientes del proyecto.

(c) Recogida y almacenamiento de datos. Se procedió a la recogida y almacenamiento de datos, planificando su almacenamiento en un sistema en la nube para permitir el acceso a los resultados desde cualquier ubicación con conexión a internet. Para llevar a cabo esta tarea, se instaló un dispositivo *hardware* específico: un miniPC. Este tipo de ordenador, compacto en su diseño al incluir todo el *hardware* en un espacio reducido, fue la elección para desempeñar la función de recopilación y envío de datos al servidor en la nube. Dada la complejidad inherente de este miniPC, se decidió instalar en él un gestor de bases de datos *MySQL*. Esta adición al dispositivo se encargará de mantener una copia local de los datos en caso de que la conexión directa con la nube experimente

algún fallo, corte o interrupción. Este enfoque garantiza la redundancia y disponibilidad continua de los datos almacenados, incluso en situaciones adversas de conectividad.

(d) Conexión a internet. Se considera esencial contar con una conexión a internet estable para acceder a un servidor en la nube. Con este propósito, se optó por emplear un *router Huawei 4G*. Este dispositivo se distingue por disponer de una batería que posibilita su funcionamiento de manera inalámbrica. Permanece conectado de forma continua, permitiendo que la conexión siga operativa incluso si se producen fallos eléctricos. Este tipo de dispositivo presenta una ventaja adicional en comparación con los *routers* más tradicionales, ya que no requiere una conexión física, operando de forma inalámbrica mediante tecnologías 3G o 4G, similar a la conectividad de un *smartphone*. Esta característica proporciona independencia respecto a líneas de conexión físicas, ofreciendo una solución más flexible y adaptable a diversas condiciones de operación.

(e) Servidor cloud. En lo que respecta al servidor en la nube, se ha optado por emplear un servidor propio de AIDIMME, donde se desarrolló una *API* (Interfaz de Programación de Aplicaciones) de comunicación. Este servidor opera con *Linux Debian* y gestiona las solicitudes entrantes mediante *Slim*, un *microframework* de *PHP* (*Hypertext Preprocessor*) conocido por su versatilidad y bajo costo en recursos computacionales, en contraste con otros como *Laravel* o *Symphony*. El servidor en la nube asume la responsabilidad de llevar a cabo el tratamiento de datos e implementar los sistemas de analítica predictiva. La elección de estos sistemas ha influido en la configuración final de la arquitectura de datos. Para el análisis predictivo, se empleó *Tensorflow*, una biblioteca de *Python* que implementa diversos algoritmos de aprendizaje automático (Abady et al., 2016). Sin embargo, al ser esta librería incompatible con la versión de *Debian* utilizada por el servidor, se resolvió incluyendo el sistema de predicción en un segundo servidor en la nube externo. Para este propósito, se seleccionó la plataforma *Azure* de *Microsoft* para implementar este segundo servidor de análisis.

Azure, el servicio en la nube de *Microsoft*, ofrece una amplia variedad de aplicaciones y servicios bajo

demanda, con opciones diversas para alojar servicios de predicción. En este proyecto, se han utilizado dos de estas opciones: *Container Registry* y *App Service*. *Container Registry* es el sistema de almacenamiento de contenedores de Azure, permitiendo la carga, mantenimiento y actualización de contenedores de aplicaciones. Aunque *Docker* dispone de su propio sistema de almacenamiento de contenedores, *Docker Hub*, se optó por *Container Registry* debido a su mejor integración con el servicio de aplicaciones de Azure.

La segunda aplicación utilizada *Azure App Service*, es un servicio que facilita el despliegue de aplicaciones web utilizando la tecnología *serverless*. Este enfoque prescinde de un servidor dedicado para gestionar la aplicación, adaptándose dinámicamente a las necesidades de recursos de *Azure* en cada momento. Esto minimiza los costos y asegura la disponibilidad del servicio, independientemente del tráfico recibido en cada instante. Con la implementación de todos estos elementos, ahora es posible representar la arquitectura de datos definitiva que se despliega para modelar el proceso (Figura 1).

implementación de sistemas de analítica predictiva utilizando tecnologías como *Tensorflow* y servicios en la nube de *Azure*, garantizando así una gestión eficiente y avanzada de la información en todo el ciclo del proceso.

Tal como se aprecia en el diagrama, el flujo de datos debe comenzar en la centrífuga. Todos los datos que se obtengan deben ser gestionados en los PLCs de la EDAR, desde donde llegan hasta el SCADA y se pueden extraer directamente del PLC mediante *Modbus*. Se programó el miniPC utilizando *Python* para leer los datos del PLC. Después, estos se guardan en su base de datos *MySQL* local, a la vez que son enviados al servidor cloud de AIDIMME. Allí, son recibidos por *PHP* y *Slim*, los cuales se encargarán de gestionar la *API* y servir los datos en tiempo real. En caso de recibir una petición de simulación, este lanzará la petición a *Azure*, que utilizará el *App service* para ejecutar los sistemas de predicción. Estos resultados son devueltos al servidor de AIDIMME, responsable de comunicarse con el método de visualización, que en este caso son unas gafas de realidad virtual *Oculus Quest 2*.

Análisis de datos y aprendizaje

Se identificaron tres tipos principales de variables en el proceso de centrifugado. Las variables de entrada son conocidas antes del proceso y afectan los resultados. Incluyen el caudal de fango (ajustable manualmente), el caudal de polielectrolito (regulado por operarios), y la velocidad diferencial, que, al aumentar, reduce la exigencia a la centrifuga. La concentración del fango espesado varía según los residuos recibidos. Por otro lado, las variables de funcionamiento son datos generados durante el proceso. Incluyen vibraciones en los rodamientos, temperaturas de los rodamientos, consumo eléctrico de los motores y el par del motor. Finalmente, las variables de salida miden características de los productos resultantes: la materia seca del fango deshidratado (concentración de masa sólida) y la calidad del escurrido (partículas sólidas en el agua residual). Estas variables son fundamentales para el análisis del funcionamiento de la máquina y son datos valiosos para el sistema de simulación.

Una vez se dispuso de una cantidad considerable de datos almacenados, se llevó a cabo un

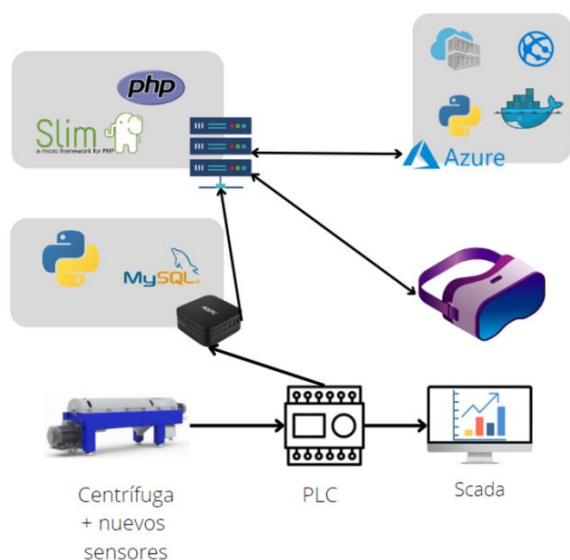


Figura 1. Diagrama general de la arquitectura de datos

Esta arquitectura integral abarca desde la recogida inicial de datos mediante sensores y PLCs, pasando por la consolidación y almacenamiento en el sistema en la nube utilizando el miniPC y el servidor propio de AIDIMME, hasta la

minucioso proceso de limpieza y ajuste para garantizar la calidad adecuada de los datos, permitiendo que los algoritmos de aprendizaje automático desempeñaran sus funciones de manera eficaz. Este procedimiento abarcó la identificación y eliminación de valores erróneos y nulos, así como la gestión de outliers, aquellos valores que se desviaban significativamente del conjunto principal. Se destacó la importancia del monitoreo de datos relacionados con las vibraciones, velocidad de rotación, corriente eléctrica y torque.

El último paso antes de la ejecución de los algoritmos implicó la división del conjunto de datos en dos: el conjunto de entrenamiento y el conjunto de prueba. El primero fue proporcionado a los algoritmos para facilitar el aprendizaje y el desarrollo de modelos necesarios para las simulaciones, mientras que el segundo se utilizó para evaluar el rendimiento del modelo en datos no utilizados previamente para el aprendizaje. Esta división se realizó para verificar que el modelo se comportara adecuadamente no solo en los datos conocidos sino también al realizar predicciones precisas en datos nuevos.

Se enfatizó que un modelo capaz de predecir correctamente los datos de entrenamiento pero con un desempeño mediocre en el conjunto de prueba podría haber caído en el fenómeno conocido como *overfitting*, donde el modelo se ajusta excesivamente a los patrones del conjunto de entrenamiento, resultando en un rendimiento subóptimo en datos no vistos previamente. Luego, se procedió a la ejecución de los algoritmos de aprendizaje automático seleccionados, aprovechando bibliotecas específicas disponibles en *Python*, como *TensorFlow*, *Keras*, *ScikitLearn* y *Pytorch*, que simplificaron significativamente el desarrollo e implementación de estas soluciones.

El gemelo digital empleó redes neuronales de regresión para reproducir los comportamientos de las variables relevantes del proceso. La estructura de las redes neuronales cumplió un doble propósito: garantizar la consistencia entre las diversas variables monitoreadas por el gemelo digital y permitir la simulación. Este último aspecto significó que, a partir de los datos recopilados y los modelos previamente entrenados, el gemelo pudo entrar en modo de simulación y prever qué datos

se obtendrían de una variable dada una configuración específica de las demás variables.

Uno de los desafíos comunes al utilizar sistemas de aprendizaje automático, especialmente redes neuronales, era la lentitud asociada al entrenamiento de modelos, que a menudo se extendía durante semanas o incluso meses al manejar conjuntos de datos extensos. Para abordar esta limitación, se implementó la transferencia de aprendizaje, una técnica que aprovecha modelos previamente entrenados para tareas similares, reduciendo significativamente el tiempo necesario para el aprendizaje. Se optó por la estrategia de bloqueo de neuronas entrenadas, reentrenando exclusivamente las capas finales de una red neuronal sobre datos específicos del proyecto. Este enfoque preservaba el conocimiento fundamental mientras se adaptaba a nuevas salidas. No obstante, se enfrentó al desafío común del sobreajuste, para lo cual se aplicaron técnicas como *Dropout* (Maier et al., 2023). En el proyecto, la transferencia de aprendizaje o *Transfer Learning* se utilizó como base para adaptar modelos de gemelos digitales a nuevas máquinas, permitiendo una rápida creación de simulaciones específicas mediante el bloqueo de neuronas entrenadas (Pisa et al. 2023).

Para llevar a cabo las simulaciones requeridas en el sistema de simulación, se desarrolló un sistema basado en *machine learning*, también conocido como aprendizaje automático. En este contexto, el problema se enmarca claramente como una regresión, lo que implica un aprendizaje supervisado. La elección de algoritmos se centró en las redes neuronales, y después de diversas pruebas, se optó por una arquitectura que demostró ofrecer resultados destacados: redes neuronales con dos capas ocultas, cada una con 10 neuronas. El diagrama de la Figura 2 ilustra esta configuración, con la capa de entrada en gris, las capas ocultas en salmón y la capa de salida en azul.

El número de neuronas y capas ocultas se determinó mediante experimentación, y esta configuración proporcionó buenos resultados para todas las variables. Además, el número de neuronas en la capa de entrada se ajusta según la cantidad de variables de entrada, en este caso, 4 variables.

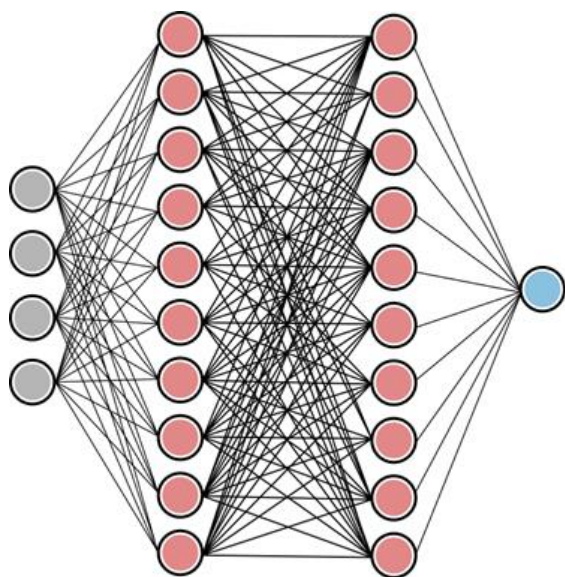


Figura 2. Representación de las redes neuronales utilizadas.

Por último, para la capa de salida, se eligió tener una red neuronal separada para cada variable, optimizando así el rendimiento de cada sistema independiente. El diseño final se resume en el esquema presentado en la Figura 3.

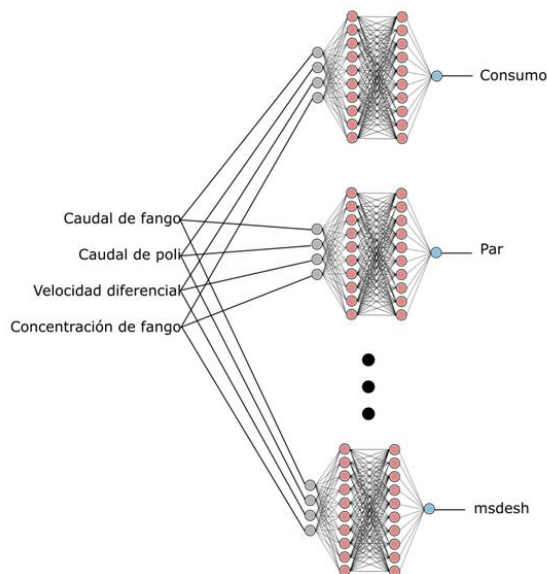


Figura 3. Diseño final de redes neuronales.

En cuanto al procesamiento, limpieza y preparación de datos, se emplearon las bibliotecas *Pandas* y *Sklearn* de *Python*. La implementación de las redes neuronales se llevó a cabo con la biblioteca *Keras* de *Python*, que utiliza *Tensorflow* como motor y se basa en la librería *Numpy*.

Visualización e interacción

Una vez completado el desarrollo del gemelo digital, capaz de generar simulaciones de funcionamiento, surge la necesidad de facilitar su accesibilidad a usuarios finales sin capacitación técnica especializada. Para lograr esto, se optó por la integración del sistema de simulación en una interfaz gráfica de usuario (GUI) utilizando aplicaciones de realidad virtual (RV). Esta elección ofrece la ventaja de crear interfaces más realistas e incluso inmersivas, permitiendo a los usuarios interactuar de manera más intuitiva.

Para crear el entorno gráfico del gemelo virtual, se llevó a cabo el modelado 3D tanto de la máquina a virtualizar, una centrífuga para el procesamiento de fangos (*Alfa Laval NX4000*), como del entorno en el cual se desea ubicar esta máquina virtual. El modelado 3D implica la representación tridimensional de objetos en un espacio virtual, y este proceso se realizó utilizando técnicas y herramientas disponibles en el software de modelado CAD 3D, específicamente *Solidworks Professional 2022* en este caso.

En la etapa inicial, se creó un espacio con características similares al entorno real donde se planeaba integrar la máquina virtual. Este espacio incorporó una puerta de entrada doble y una zona abierta destinada a albergar la máquina. Se diseñaron ventanas para facilitar la entrada de luz natural y se incluyeron vigas con forma de perfil IPN para conferirle un aspecto industrial distintivo. Un perfil IPN se refiere a un tipo específico de perfil de acero utilizado en la construcción. El modelado resultante brindó al usuario una experiencia inmersiva en un entorno claramente definido, lo que le permitió desplazarse e interactuar con los elementos diseñados con este propósito.

Al desarrollar el gemelo, el primer paso implica la creación de un modelado 3D de la centrífuga. Para llevar a cabo este proceso, se empleó *Blender*, una de las herramientas de modelado 3D más reconocidas y ampliamente utilizadas en todo el mundo. (Figura 4).

Con el objetivo de interactuar con los diversos controles de la máquina, ubicados en diferentes puntos de la nave en la instalación real, se diseñó

un panel de control que abarca todos los parámetros cruciales que los usuarios analizan para tomar decisiones y realizar ajustes. Este panel replicó de manera simplificada los paneles y elementos de control presentes en la realidad, facilitando la identificación visual para los usuarios mediante botones y mandos, tales como el botón de puesta en marcha, la seta de parada de emergencia, la palanca para seleccionar entre el estado real y el estado de simulación, la regulación del caudal de fango de entrada a la máquina, la regulación del caudal de polielectrolito de entrada a la máquina, la regulación de la velocidad diferencial y la regulación del contenido de sólidos en el fango de entrada.

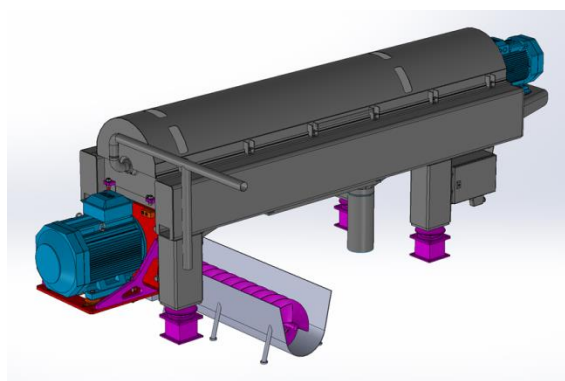


Figura 4. Imagen de la máquina modelada en CAD 3D.

Además de las pantallas que muestran los valores modificables con los reguladores, se incorporan pantallas de visualización que ofrecen información sobre diversos sensores y variables, permitiendo tomar decisiones informadas. Estas incluyen la concentración de fango espesado, el peso del residuo sólido, la concentración del escurrido, la temperatura del rodamiento de alimentación, la vibración del rodamiento de alimentación, la temperatura del rodamiento de freno, la vibración del rodamiento de freno y el valor medio de corriente.

La máquina y el panel de control fueron integrados en el entorno virtual creado, posibilitando la interacción del usuario en dicho espacio (Figura 5). Tras la generación del entorno en CAD 3D, se llevó a cabo el tratamiento necesario para definir el mallado de los elementos con un nivel de detalle suficiente para su correcta identificación. Simultáneamente, se aseguró de que el procesamiento en el entorno de realidad virtual

fuera ágil, garantizando así una experiencia óptima para el usuario.

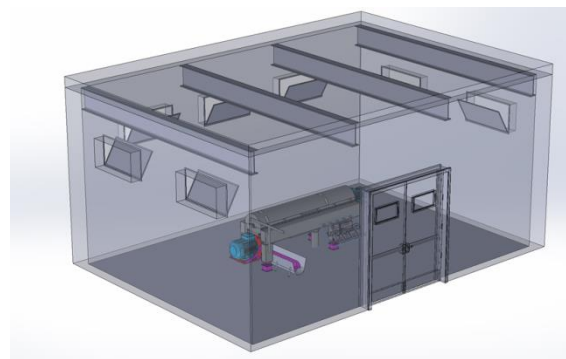


Figura 5. Imagen del espacio virtual en CAD 3D donde se aprecia en su interior la máquina virtual y el panel de control.

3. Resultados y discusión

Se implementaron diversas características para proporcionar una experiencia inmersiva y funcional. Se obtuvieron pantallas específicas para visualizar las variables de entrada, funcionamiento y salida, junto con pantallas dedicadas a los valores críticos, como temperaturas y vibraciones de los rodamientos. Este enfoque modular y visualmente claro facilita la interpretación de los datos, permitiendo a los usuarios acceder de manera eficiente a la información relevante según sus necesidades específicas. Además, se consiguió un panel de mando interactivo para controlar la simulación, incluyendo funciones como encendido y apagado, emergencia y cambio de modo de funcionamiento, junto con la visualización resumida de las variables. La capacidad de introducir datos en tiempo real a través de este panel proporciona una interfaz efectiva para controlar y observar el comportamiento del gemelo digital

Interfaz gráfica del Gemelo Digital

En la implementación del gemelo virtual, se logró una integración exitosa del modelo 3D de la centrífuga con un diseño de interacción intuitivo para los usuarios (Figura 6). Se habilitó un interruptor representado por una palanca para cambiar sin esfuerzo entre los modos de tiempo real y simulación, mejorando así la versatilidad y la experiencia del usuario. Una pantalla flotante fue incorporada para proporcionar información continua sobre el modo de funcionamiento actual

del gemelo digital, garantizando una comprensión clara para el usuario.

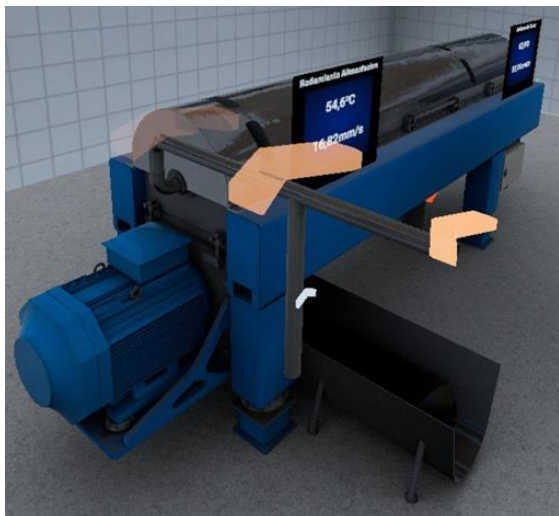


Figura 6. Modelo 3D de la centrífuga.

La interfaz de usuario del gemelo virtual fue diseñada con un enfoque modular y visualmente claro, priorizando la facilidad de interpretación de datos para los usuarios. Se implementaron pantallas específicas para visualizar las variables de entrada, funcionamiento y salida, así como para mostrar valores críticos como temperaturas y vibraciones de los rodamientos (Figura 7). Esta estructura modular permite a los usuarios acceder de manera eficiente a la información relevante, adaptándose a sus necesidades específicas.



Figura 7. Pantallas de visualización de variables y su posición.

La interacción con el gemelo virtual se planificó cuidadosamente para permitir su funcionamiento tanto en modo tiempo real, conectándose para observar los datos en directo, como en modo

simulación, solicitando entradas de valores y devolviendo predicciones. Se incorporó un interruptor dentro del modelo, accionado por una palanca, para cambiar entre ambos modos, y se implementó una pantalla flotante que proporciona información continua sobre el modo de funcionamiento actual del gemelo digital, garantizando una comprensión instantánea para el usuario. Adicionalmente, se incluyeron pantallas dedicadas para visualizar todas las variables clave, asegurando una representación completa del estado del sistema. Este diseño integral y orientado al usuario mejora la experiencia de interacción, permitiendo un monitoreo eficiente y una toma de decisiones informada.

En el ámbito del control de la simulación, se diseñó un panel de mando interactivo destinado a facilitar la introducción de las cuatro variables de entrada esenciales: caudal de fango, caudal de polielectrolito, concentración del fango y velocidad diferencial. Este panel de control interactivo permite encender y apagar la máquina, activar el apagado de emergencia, cambiar el modo de funcionamiento y visualizar de manera resumida las variables clave. En la Figura 8 se puede observar cómo en la parte inferior izquierda del panel de control se ubican el botón de encendido y apagado (en verde) y la seta de emergencia (en rojo).

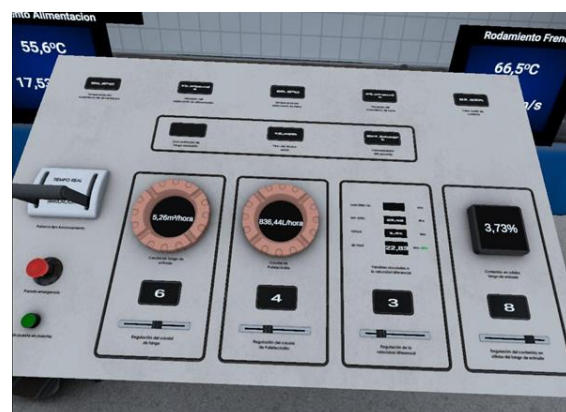


Figura 8. Panel de mando.

Justo encima de estos elementos, se encuentra la palanca para cambiar el modo de funcionamiento. A la derecha de esta sección, se disponen cuatro paneles destinados a la introducción de las variables de entrada mediante controles deslizantes, simplificando así el proceso de control y manteniendo una uniformidad en la manipulación

de todas las variables. Cada panel muestra el valor de la variable correspondiente en la parte superior, dentro de su recuadro individual. Este diseño busca mejorar la accesibilidad y eficiencia en la gestión de la simulación.

Se implementó otro elemento esencial para proporcionar una representación visual inmediata de las variaciones en la máquina. En situaciones donde, por ejemplo, hay cambios en el caudal del fango, el caudal del polielectrolito, la velocidad u otras salidas de materiales, resulta crucial reflejar estos cambios en el modelo. Sin embargo, la visualización directa de la mayoría de los fluidos es limitada debido a tuberías opacas. Para superar esta limitación, se incorporó un flujo continuo de flechas de diferentes colores en las entradas y salidas de la máquina (Figura 9).

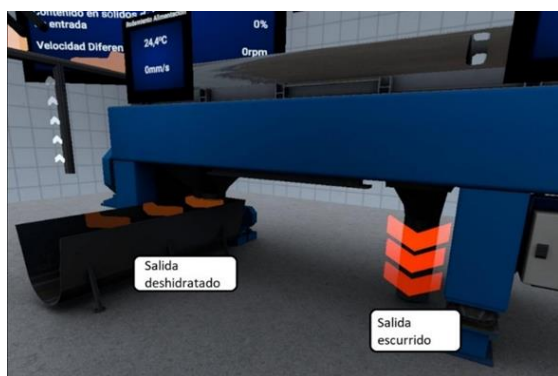


Figura 9. Flechas en los flujos de salida.

Estas flechas indican la presencia y la intensidad del flujo. En el caso de un aumento en el flujo, el número de flechas aumenta, mientras que una disminución en el flujo se representa con un número menor de flechas. Se han establecido cuatro flujos distintos: entrada de polielectrolito, entrada de fango, salida de escurrido y salida del deshidratado. Estas características visuales se ilustran en las imágenes siguientes, proporcionando así una representación gráfica intuitiva de los cambios en el sistema.

Finalmente, se implementó la capacidad de visualizar la variación diferencial de manera intuitiva mediante flechas en el motor de la centrifuga. En este contexto, las flechas no fluyen, sino que giran constantemente alrededor del motor, proporcionando así una representación gráfica dinámica de la variación en la velocidad diferencial del sistema. Esta visualización adicional

permite a los usuarios comprender rápidamente cualquier cambio en la velocidad diferencial, contribuyendo así a una experiencia más completa e informativa dentro del entorno del gemelo digital.

Sistemas de interacción con el gemelo digital

La implementación de la interfaz de usuario para el gemelo digital se llevó a cabo con éxito, destacando la integración eficiente de los componentes clave (Figura 10). Ubicados estratégicamente cerca de la centrifugadora, los sensores y dispositivo de captura de datos (E1) cumplieron con éxito su tarea de capturar datos operativos en tiempo real. La capacidad de procesar y transmitir estos datos al servidor en la nube mediante un API proporcionó una base sólida para la continuidad del flujo de información.

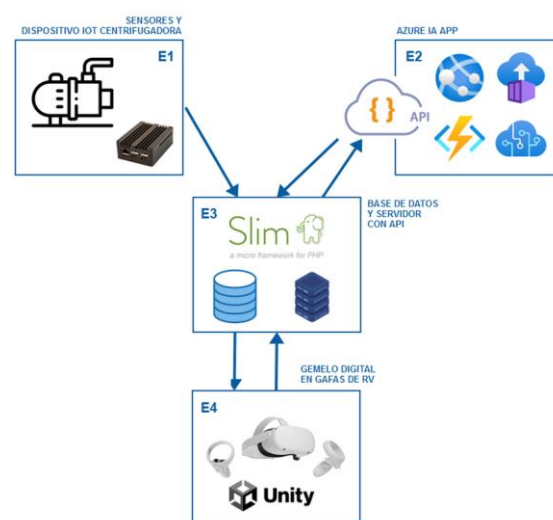


Figura 10. Esquema de interacciones entre los distintos elementos que componen el sistema.

La aplicación de inteligencia artificial (IA), desarrollada con enfoque en aprendizaje automático (E2), demostró una simulación precisa del comportamiento de la centrifugadora. La implementación de una API que permite la interrogación desde el servidor en la nube aseguró una comunicación efectiva y fluida entre los componentes.

El servidor en la nube (E3) desempeñó un papel central al lograr el almacenamiento eficiente de los datos en tiempo real del dispositivo IoT. Su capacidad para comunicarse sin inconvenientes con

la aplicación de simulación en Azure mediante un API, así como proporcionar acceso remoto a los datos, consolidó su importancia en el sistema. La aplicación de gemelo digital (E4), desarrollada en Unity para experiencias en realidad virtual, fue la interfaz de usuario final. Su capacidad para visualizar en tiempo real el comportamiento de la centrifugadora, simular cambios en variables de entrada y ofrecer una interfaz intuitiva contribuyó significativamente a la experiencia del usuario.

La columna vertebral de la eficiente interacción entre los componentes del sistema reside en la implementación cuidadosa de *endpoints* y API. Tanto el servidor de datos en la nube como la aplicación de IA almacenada en Azure desempeñan roles críticos y se comunican de manera efectiva mediante estos puntos de conexión. Los *endpoints* del servidor de datos en la nube juegan un papel multifacético. En primer lugar, permiten la recepción y almacenamiento de datos en tiempo real provenientes del dispositivo IoT ubicado en la centrifugadora. Además, facilitan la comunicación con la aplicación de IA para la simulación, alojada en Azure. Esto se logra a través de *endpoints* específicos que responden a las consultas y solicitudes de información. La aplicación de IA en Azure presenta *endpoints* diseñados para interactuar con el servidor de datos. Estos *endpoints* posibilitan la obtención de datos almacenados en el servidor en tiempo real, contribuyendo así al proceso de simulación y aprendizaje automático. La efectividad de estas comunicaciones se traduce en la capacidad del gemelo digital para acceder y representar fielmente el comportamiento de la centrifugadora.

Pruebas en la planta depuradora y desafíos identificados

La etapa de pruebas en la planta depuradora tras la implementación y entrenamiento de los algoritmos predictivos fue crucial para evaluar el rendimiento del gemelo digital en condiciones del mundo real. Durante este proceso, los técnicos de la EDAR tuvieron la oportunidad de examinar y validar los resultados, así como de sugerir posibles ajustes.

La interfaz de visualización, a través de gafas de realidad virtual, permitió una observación en tiempo real de las variables, imitando la funcionalidad de un sistema SCADA. Además, se exploraron escenarios simulados utilizando el

modo de simulación del gemelo para prever las variaciones en los resultados al modificar los parámetros de entrada.

Se implementó un programa específico para demostrar las respuestas del gemelo a cambios en las variables de entrada. Este programa generó datos numéricos y gráficos, proporcionando una representación clara de las predicciones en diferentes situaciones. En términos generales, las predicciones del gemelo digital se alinearon de manera consistente con los resultados experimentales, según el análisis y la retroalimentación del personal técnico de la depuradora. No obstante, se identificó un desafío específico relacionado con la calidad del escurrido al aumentar el caudal de polielectrolito. A pesar de la fiabilidad general del algoritmo, se reconoció la necesidad de una investigación más detallada para abordar cualquier posible error en el análisis, entrenamiento o falta de datos.

Además, se destacaron dos desafíos fundamentales. En primer lugar, la asimetría en la frecuencia de recopilación de datos de salida en comparación con los datos de entrada y funcionamiento. La escasa disponibilidad de datos de salida podría estar contribuyendo a las desviaciones observadas. En segundo lugar, la capacidad del gemelo para funcionar "en los bordes", es decir, fuera de los rangos normales de operación plantea desafíos significativos debido a la falta de datos para prever comportamientos extremos. Este inconveniente podría abordarse mediante experimentos adicionales, a pesar de la incertidumbre asociada a forzar condiciones fuera de los rangos normales. Estos desafíos identificados señalan áreas de trabajo para futuras mejoras y refinamientos en el gemelo digital.

Transfer Learning: aprovechamiento del conocimiento previo

Para evaluar la viabilidad del *transfer learning* y determinar en qué medida el conocimiento previamente adquirido por los algoritmos puede reutilizarse, se aplicó esta técnica a los desarrollos existentes implementados.

En esta técnica, se "bloquea" el conocimiento aprendido en las 2 capas ocultas de los algoritmos originales para la generación del gemelo digital de

la máquina centrífuga (Figura 11). Este conocimiento está representado por las matrices de pesos $W1$ y $W2$ en las capas ocultas de las redes neuronales. La última capa ($W3$) es la única que se vuelve a entrenar para adaptarse a un nuevo problema o variable. La red neuronal resultante se compone de dos partes: la primera parte representa el conocimiento previo de la máquina y la segunda parte implementa la capa de salida específica para la nueva variable. Esto permitió acelerar significativamente el proceso de entrenamiento.

En comparación con el enfoque tradicional, donde se entrenarían 9 variables, el tiempo de entrenamiento se redujo de aproximadamente 36 minutos a 29 minutos utilizando *transfer learning*. Esto representa una mejora del 15% en la velocidad del proceso, demostrando la eficacia de la transferencia de conocimiento previo en la optimización de tareas de entrenamiento de redes neuronales. Los resultados obtenidos en términos de precisión y error mínimo cuadrado fueron comparables, validando la utilidad de esta estrategia para desarrollos futuros.

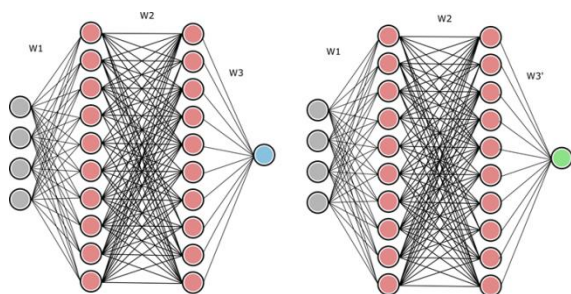


Figura 11. Redes neuronales y matrices de pesos.

4. Conclusiones

El gemelo digital emerge como una valiosa herramienta que aporta múltiples beneficios en función de sus modos de operación. El modo de trabajo en tiempo real ofrece la ventaja de proporcionar una representación inmediata y precisa de la maquinaria desde cualquier lugar del mundo con conexión a internet. Mientras tanto, el modo de simulación permite obtener resultados del proceso con variaciones específicas sin necesidad de ejecutarlo físicamente.

Para desarrollar sistemas de predicción efectivos, es imperativo considerar eventos más allá de los rangos de funcionamiento "normales". Esto

requiere la realización de experimentos que recojan datos de entrada, operación y resultados en condiciones extremas, permitiendo así predecir el comportamiento del sistema en situaciones atípicas. En el contexto de sistemas predictivos, los datos capturados a través de sensores son más propicios para predicciones precisas en comparación con mediciones manuales. Esto se debe a la frecuencia de toma de datos, lo que exige una extensa recopilación de información a lo largo del tiempo para lograr la precisión deseada.

Una vez que se establece la infraestructura de datos, las API habilitan la integración de la visualización con diversos sistemas, como la aplicación de realidad virtual en el caso del piloto, pero también la adaptación a otras plataformas, como aplicaciones web. El empleo de plataformas en la nube, como Azure, agiliza el desarrollo y despliegue de aplicaciones en entornos en línea, brindando un amplio conjunto de herramientas. Por otro lado, a pesar de la conectividad constante con servicios en la nube, es fundamental mantener una copia local de los datos, especialmente en ubicaciones industriales distantes de los centros urbanos, para garantizar la continuidad en la recopilación de datos en caso de interrupciones en el servicio de datos o de internet.

El uso de la tecnología de *transfer learning* puede acelerar el proceso de aprendizaje en operaciones similares. Su utilidad varía en función de la complejidad del proceso, siendo más beneficiosa en situaciones altamente complejas que requieren un extenso período de entrenamiento.

5. Agradecimientos

Los autores desean expresar su agradecimiento al Instituto Valenciano de Competitividad Empresarial (IVACE) por el apoyo financiero proporcionado a través de la subvención con número de expediente IMDEEA/2021/52, así como a la Entidad Pública de Saneamiento de Aguas Residuales de la Comunidad Valenciana (EPSAR) por hacer posible esta investigación.

5. Referencias

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al. (2016). "Tensorflow: a system for large-scale machine learning." 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), pp. 265-283.

Bahramian, M., Dereli, R. K., Zhao, W., Giberti, M., Casey, E. (2023). "Data to intelligence: The role of data-driven models in wastewater treatment." *Expert Systems with Applications*, 217, 119453.

Grieves, M. (2014). "Digital twin: Manufacturing excellence through virtual factory replication." White paper, 1-7.

Grieves, M., & Vickers, J. (2017). "Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems."

Maier, H. R., Galelli, S., Razavi, S., et al. (2023). "Exploding the myths: An introduction to artificial neural networks for prediction and forecasting." *Environmental Modelling & Software*, 167, 105776. DOI: 10.1016/j.envsoft.2023.105776.

Pisa, I., Morell, A., Lopez Vicario, J., Vilanova, R. (2023). "Transfer Learning in wastewater treatment plants control: Measuring the transfer suitability." *Journal of Process Control*, 124, 36-53. DOI: 10.1016/j.jprocont.2023.02.006.

Redelinghuys, A.J., Kruger, K., & Basson, A.H. (2019). "A Six-Layer Architecture for Digital Twins with Aggregation." *Service Orientation in Holonic and Multi-Agent Manufacturing*.

Semeraro, C. (2020). Contribution to the Formalization of Data-Driven Invariant Modelling Constructs of Cyber-Physical Systems. Doctoral dissertation, Université de Lorraine; Politecnico di Bari.

Semeraro, C., Lezoche, M., Panetto, H., & Dassisti, M. (2021). "Digital twin paradigm: A systematic literature review." *Computers in Industry*, 130, 103469.

Singh, M., Fuenmayor, E., Hinchey, E., Qiao, Y., Murray, N., & Devine, D. (2021). "Digital twin: Origin to future." *Applied System Innovation*.