Universidad de Alcalá

Programa de Doctorado en Ingeniería de la Información y del Conocimiento (Plan: D442)

DESARROLLO DE SERVICIOS AVANZADOS PARA LA TOMA DE DECISIONES INTELIGENTES EN DIMENSIONES CLAVES DE LA SOCIEDAD DIGITAL

Tesis Doctoral presentada por

LUIS AGUSTIN MENDOZA PITTÍ

Directores: Dr. MIGUEL VARGAS LOMBARDO y Dr. JOSÉ MANUEL GÓMEZ PULIDO

Alcalá de Henares, Madrid, 2023

De parvis grandis acervus erit
Libro de Poemas, XVI

Dedicatoria

Esta tesis doctoral está dedicada a todas las personas que de alguna u otra manera estuvieron ahí en cada momento brindándome el apoyo para salir adelante en esta fase de mi carrera profesional. Y que, a pesar de la diferencia horaria y distancia, siempre permanecieron ahí para darme ánimos y motivación para salir adelante en este desafío.

Agradecimientos

A Dios por guiarme y bendecirme a lo largo de la realización de esta nueva faceta en mi vida profesional y personal.

A mi familia por ser ese apoyo incondicional en cada momento de mi carrera doctoral.

A mi madre, Dioris Pittí por su amor y apoyo en todos los instantes de mis estudios, así como por los valores y principios que me ha brindado a lo largo de mi vida. A mi hermana Daileen Mendoza, por estar ahí en cada instante que lo necesitaba y apoyarme en las decisiones que tomaba.

A mi pareja Sandra Gutiérrez, por su apoyo incondicional en los momentos más difíciles que se me presentaban, así como por creer siempre en mí y mis capacidades para salir hacia adelante.

A mis asesores de tesis, el PhD. Miguel Vargas-Lombardo y el PhD. José Manuel Gómez-Pulido por todo el conocimiento y orientación científica que me han dado durante mis estudios doctorales. Así como por sus sabios consejos en pro de mi carrera como investigador.

Al doctorando Huriviades Calderón, por su disponibilidad para tratar y discutir los conocimientos adquiridos durante el doctorado.

Al la Secretaría Nacional de Ciencia, Tecnología e Innovación (SENACYT) y al Instituto para la Formación y Aprovechamiento de Recursos Humanos (IFARHU), por ser ese apoyo económico durante toda mi estancia doctoral a través de la beca del programa doctoral 2018, ya que sin este recurso no sería posible desarrollar mi estudio doctoral.

A la Universidad Tecnológica de Panamá (UTP), por ofrecerme la oportunidad de crecimiento y desarrollo profesional, a fin de brindar todos mis conocimientos adquiridos durante esta tesis doctoral a esta alta casa de estudios.

Resumen

La presente tesis doctoral tiene como meta facilitar la toma de decisiones inteligentes dirigidas a los principales actores dentro de las dimensiones claves de la sociedad digital mediante el desarrollo de servicios avanzados. Para ello, los conceptos de Smart City, Soft-Computing y Arquitecturas de Software Orientada a Servicios favorecen la introducción de las tecnologías necesarias para el desarrollo de servicios flexibles, escalables y de calidad, con la finalidad de mejorar el bienestar social y la sostenibilidad ambiental.

Los resultados obtenidos a lo largo del trabajo de investigación realizado y que da lugar a esta tesis, han sido difundidos en el ámbito científico internacional mediante cinco contribuciones significativas publicadas tanto en revistas como congresos científicos de gran prestigio, los cuales conforman la memoria por compendio de la tesis doctoral que se expone. Esta investigación se centró en dos dimensiones claves de la Ciudad Inteligente (Smart City) como lo son la Energía Inteligente (Smart Energy) y la Salud Inteligente (Smart Health).

En este sentido, en la dimensión de la Smart Energy, se han incorporado tres contribuciones significativas. La primera es una revisión exhaustiva de la literatura científica, en la cual se identifican y analizan avanzadas arquitecturas orientada a servicios inteligentes en el contexto de la mejora de la eficiencia energética para apoyar la toma de decisiones respecto al funcionamiento y/o comportamiento óptimo de los sistemas energéticos de los edificios. La segunda contribución propone la simulación del comportamiento y operaciones de diferentes sistemas de calefacción, ventilación y aire acondicionado (HVAC) de un edificio concreto mediante el uso de modelos físicos adecuados a estos sistemas lo que permite desarrollar algoritmos de Soft-Computing, como sistemas de control efectivos con optimización multiobjetivo, que mejoren la eficiencia energética de los edificios. Por último, la tercera contribución propone un modelo de Deep Learning, basado en redes Long Short-Term Memory (LSTM) que abordan series temporales, para el pronóstico del consumo energético diario de los sistemas de climatización de los edificios, lo que permitiría mejorar la eficiencia y sostenibilidad del edificio. Este modelo incorpora factores como las horas de trabajo, las actividades del edificio, el clima, temperatura interior del edificio, entre otros.

Por otro lado, en la dimensión de la Smart Health, se han adicionado dos contribuciones científicas que muestran los resultados obtenidos de la investigación realizada. En la primera contribución se presentan dos plataformas de eHealth diseñadas y desarrolladas para dotarlas de una funcionalidad de ubicuidad e incorporando servicios avanzados mediante técnicas de aprendizaje automático e interoperabilidad, con la finalidad de mejorar la calidad de vida de los pacientes en cuidados paliativos y de sus familiares, así como ayudar a la toma de decisiones inteligentes mediante la estimación del riesgo de síndrome de Down. Mientras que, la segunda contribución en esta dimensión de la Smart Health propone un modelo de Machine Learning con capacidad para descubrir y aprender los patrones presentes en los datos para la predicción de la hipotensión en pacientes que reciben tratamiento de hemodiálisis. Para lograr esto, se utilizó un novedoso conjunto de datos en el que se unifican las variables de dos bases de datos bien diferenciadas (datos clínicos y analíticos). Todo ello proviene de la gestión de una cantidad ingente de datos (big data) obtenidos por los servicios hospitalarios automatizados tanto de las pruebas analíticas como de los equipos de diálisis. Esta aplicación de eHealth permite la toma de decisiones ante la predicción, al inicio de la sesión de hemodiálisis, de aparición de un episodio de hipotensión.

Abstract

This doctoral thesis aims to facilitate intelligent decision-making aimed at the main actors within the key dimensions of the digital society through the development of advanced services. For this, the concepts of Smart City, Soft-Computing and Service-Oriented Software Architectures favor the introduction of the necessary technologies for the development of flexible, scalable and quality services, in order to improve social welfare and environmental sustainability.

The results obtained throughout the research work carried out and that gives rise to this thesis, have been disseminated in the international scientific field through five significant contributions published both in journals and scientific congresses of great prestige, which make up the memory by compendium of the doctoral thesis that is exposed. This research focused on two key dimensions of the Smart City: Smart Energy and Smart Health.

In this sense, in the dimension of Smart Energy, three significant contributions have been incorporated. The first is an exhaustive review of the scientific literature, in which advanced architectures oriented to intelligent services are identified and analyzed in the context of improving energy efficiency to support decision-making regarding the optimal operation and/or behavior of energy systems in buildings. The second contribution proposes the simulation of the behavior and operations of different heating, ventilation and air conditioning (HVAC) systems of a particular building through the use of physical models suitable to these systems which allows the development of Soft-Computing algorithms, as effective control systems with multi-objective optimization, that improve the energy efficiency of buildings. Finally, the third contribution proposes a Deep Learning model, based on Long Short-Term Memory (LSTM) networks that address time series, for the forecast of the daily energy consumption of the air conditioning systems of buildings, which would improve the efficiency and sustainability of the building. This model incorporates factors such as working hours, building activities, climate, interior temperature of the building, among others.

On the other hand, in the dimension of Smart Health, two scientific contributions have been added that show the results obtained from the research carried out. The first contribution presents two eHealth platforms designed and developed to provide them with ubiquity functionality and incorporating advanced services through machine learning and interoperability techniques, in order to improve the quality of life of patients in palliative care and their families, as well as help intelligent decision-making by estimating the risk of Down syndrome. Meanwhile, the second contribution in this dimension of Smart Health proposes a Machine Learning model with the ability to discover and learn the patterns present in the data for the prediction of hypotension in patients receiving hemodialysis treatment. To achieve this, a novel dataset was used in which the variables of two well-differentiated databases (clinical and analytical data) are unified. All this comes from the management of a huge amount of data (big data) obtained by automated hospital services from both analytical tests and dialysis equipment. This eHealth application allows decision-making before the prediction, at the beginning of the hemodialysis session, of the appearance of an episode of hypotension.

Índice

1 Introducción

1.1 Dimensiones claves de la Sociedad Digital

Se estima que para el 2050 más del 68% de la población estará asentada en las ciudades [1]. Además, la superficie terrestre que ocupan las ciudades es menos del 3% y estas consumen más del 75% de los recursos naturales disponibles en el mundo [2], [3]. De hecho, este acelerado crecimiento de la población incrementará el uso de los recursos limitados y generará una mayor demanda de los servicios, por lo que las ciudades enfrentarán una mayor presión social, ambiental y de infraestructura. En consecuencia, impactará de manera significativa a distintos entornos de la industria 4.0 (I4.0), tales como la energía, la salud, el transporte, las edificaciones, entre otros [4], [5].

En vista de lo anterior, la Smart City (SC) [2], [6]–[9] es un concepto que en los últimos años se ha convertido en un enfoque muy importante dentro de la literatura científica y las políticas internacionales. Ya que, ofrece una solución alentadora mediante la adopción de las Tecnologías de la Información y Comunicación (TICs), a los problemas complejos de la sostenibilidad derivados del rápido crecimiento de la población. De manera que, la SC busca satisfacer las necesidades de la sociedad, tanto en el ámbito económico, como en el aspecto operativo, sanitario, energético, social y ambiental. Por consiguiente, impulsa el desarrollo de una ciudad más sostenible, atractiva, eficiente, interactiva, accesible y con alto valor añadido para la ciudadanía y el medio ambiente [10].

Por otra parte, un importante miembro dentro del contexto de las SCs es el ciudadano, pues, este ha sido enormemente influenciado por el desarrollo tecnológico (tales como, dispositivos móviles, sensores, internet, etc.) que han adquirido las ciudades hoy en día [11]. De modo que, el ciudadano tiene una mayor participación en lo que respecta a las decisiones que impulsan al mejoramiento de la ciudad, estableciendo así el concepto de "sociedad digital" [12], [13]. Por ende, la sociedad digital ha adquirido múltiples competencias digitales que le han permitido empoderarse, ser más educado y estar más informado debido al uso de las TICs [14].

En este sentido, las TICs son parte esencial para el desarrollo e innovación de la sociedad digital de la SC [15], [16] dado que favorece la explotación de servicios avanzados que permitan el mejoramiento de la calidad de vida, el bienestar de las personas, así como de la eficiencia y sostenibilidad de las ciudades. Algunas de las tecnologías que propician la construcción de estos novedosos servicios en las distintas dimensiones de la sociedad digital son la Cloud Computing, Internet of Things (IoT), Big data y Soft-Computing [17]–[21]. Siendo estos componentes centrales para promover el cambio de paradigma y la transformación digital de las existentes y futuras Smart Cities.

Ahora bien, la SC está constituida por diferentes dimensiones que son claves para la sociedad digital hoy en día, algunos de estos son (ver Figura 1): Smart Energy, Smart Health, Smart Building, Smart Mobility/Transportation, entre otros [8], [16], [22]–[24]. Donde, muchos de estos requieren de una gobernanza funcional y una gestión integral de los datos y servicios, a fin de superar los retos socioeconómicos más importantes de la sociedad digital. Debido a que, existe un desaprovechamiento de los recursos empleados para la correcta ejecución en los diversos procesos que comprenden a cada una de estas dimensiones. Dentro de las dimensiones inteligentes que componen a la SC, el Smart Energy y el Smart Health son los más desafiantes y estudiados de acuerdo con la comunidad científica [15], [16], [23], por lo que los resultados de esta tesis doctoral se centrará en estas dimensiones claves de la SC.

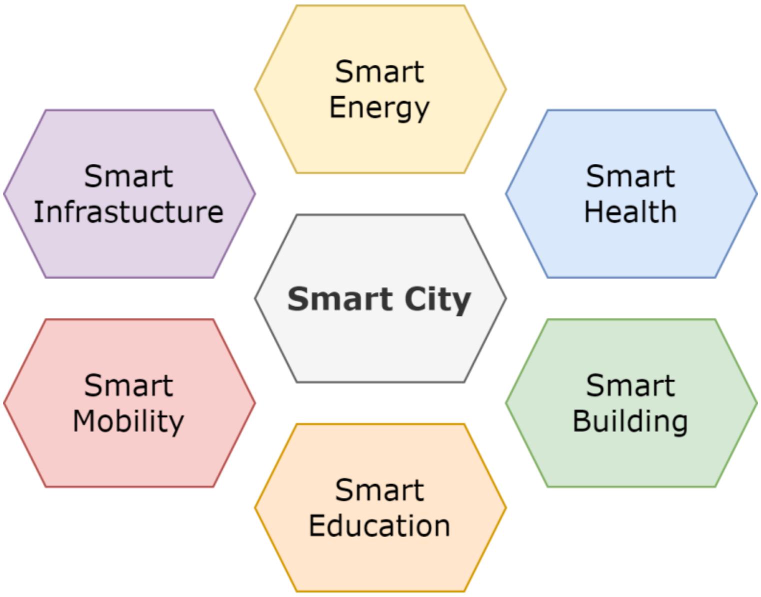


Figura 1. Algunas dimensiones que componen a la Smart City.

1.1.1 Smart Energy

La Smart Energy o "Energía Inteligente" se refiere a la utilización de tecnologías avanzadas para el monitoreo, gestión y optimización inteligente de los diferentes sistemas energéticos [15], [25]. Esto se consigue mediante la integración de diferentes componentes, como lo son los sensores, los recolectares de datos, la inteligencia artificial, el aprendizaje automático y las tecnologías de análisis de datos, entre otros. Esto permite mejorar la eficiencia y sostenibilidad energética de los diferentes sistemas.

Uno de los objetivos principales de la Smart Energy es crear un entorno confortable y sostenible para los ciudadanos mientras se mantiene la eficiencia energética. Por lo que, implementar sistemas que monitoreen, controlen y administren de manera correcta los sistemas energéticos es necesario. Por ejemplo, los edificios utilizan sistemas de gestión de edificios (BMS, por sus siglas en inglés), el cual es capaz de controlar, administrar y monitorear los componentes del edificio (por ejemplo, sistemas de climatización, sistemas de iluminación, sistemas de seguridad, elevadores, entre otros) [26].

El BMS recolecta la información proveniente de los diferentes componentes del edificio, permitiendo así analizar los registros recolectados del funcionamiento y comportamiento de estos componentes y, por lo tanto, ofreciendo información para la toma de decisiones informada sobre diferentes aspectos esenciales del edificio (como, por ejemplo, la reducción del consumo energético, mejoras en la eficiencia energética, diagnóstico preventivo de los sistemas del edificio, mejoras del confort de los ocupantes, entre otros).

No obstante, dentro de los componentes que gestiona el BMS se encuentra el sistema de climatización, el cual es uno de los más complejos y el que contribuye en gran medida en el consumo energético de los edificios, llegando este a representar más del 50% del consumo total del edificio [27], [28]. En este sentido, una gestión más eficiente de los sistemas de climatización es fundamental para mejorar la eficiencia y el consumo energético del edificio [29]. Por lo que, emplear servicios avanzados de inteligencia artificial permitirían a los administradores de edificios tomar decisiones más informadas, mejorando así la eficiencia, confort y sostenibilidad de los ciudadanos y el medio ambiente.

1.1.2 Smart Health

El Smart Health o "Salud Inteligente", presenta un gran interés por parte de la comunidad científica e industrial. El Smart Health surge de dos frentes ya conocidos, la Smart City y la salud electrónica (también conocido como e-Health); pues tienen la finalidad de mejorar los estándares de salud de la ciudadanía mediante el uso de los paradigmas tecnológicos [30]. De hecho, la Smart Health hace uso de tecnologías avanzadas y soluciones digitales para mejorar la atención médica y la salud en general.

La dimensión de Smart Health genera una cantidad ingente de información (big data) [31] proveniente de la monitorización constante de los pacientes (por ejemplo, información demográfica, enfermedades, signos vitales, pruebas de laboratorios, etc.) y, del cual, es esencial gestionar, procesar y analizar mediante la implementación de métodos confiables y rigurosos que permitirán al personal sanitario (médico, enfermera o técnico) realizar una toma de decisión más informada respecto al cuidado y/o tratamiento de la salud del paciente [32]. Por ejemplo, la historia clínica digitalizada permite al personal médico mantener registros de las enfermedades o patologías de los pacientes, a fin de conocer el estado y/o evolución del paciente y, de esta manera, ofrecer un tratamiento clínico y cuidado efectivo para el paciente.

Asimismo, la aplicación de novedosos métodos permite tanto a enfermos terminales, enfermos con trastornos genéticos y/o los enfermos renales, ser tratados y cuidados desde su domicilio, pues el cuidador tomaría decisiones más informadas de acuerdo con el análisis de los propios parámetros del paciente que recibe la atención domiciliaria. Por ejemplo, esto se aplica en el tratamiento de pacientes en cuidados paliativos domiciliario, resonancia domiciliaria, hemodiálisis domiciliaria, entre otros [33], [34]. En consecuencia, los pacientes tendrían un envejecimiento saludable o una hemodiálisis más llevadera, ya que se encuentra el paciente rodeado de sus familiares.

Teniendo en cuenta lo anteriormente expuesto, el uso, procesamiento y análisis de los datos contribuirá en gran medida a la creación de servicios avanzados de gran utilidad para las diversos actores involucrados en las dimensiones de la SC anteriormente mencionadas [24]. En este sentido, es necesario crear servicios avanzados mediante las técnicas de Soft-Computing que apoyen en la toma de decisiones inteligentes. Con la finalidad, de ofrecer una amplia gama de servicios con alto valor añadido para una mejor calidad de vida y bienestar de la sociedad, así como la mejora en la eficiencia y la sostenibilidad del medio ambiente.

1.2 Técnicas avanzadas de Soft-Computing

El término Soft-Computing fue propuesto por Lofti A. Zadeh y este señala que el Soft-Computing es un conjunto de métodos o técnicas basadas en la imprecisión, aproximación e incertidumbre, con el fin de resolver problemas complejos del mundo real [35]–[37]. Estas técnicas son capaces de aprender desde la experiencia y, por lo tanto, se pueden aplicar a cualquier dominio del problema (por ejemplo, salud, energía, biomedicina, edificaciones, agricultura, entre otros).

Debido a que, las técnicas de Soft-Computing hacen uso de cálculos aproximativos, estos ofrecen soluciones confiables y robustas a los problemas complejos que se le presenten. Asimismo, tienen un menor costo computacional en comparación a los métodos tradicionales o exactos (hard computing) [31], [37], [38]. Una característica de las técnicas de Soft-Computing, es que estas pueden utilizarse de forma combinada, ofreciendo la posibilidad de obtener resultados más fiables gracias a las características propias de cada una de las técnicas.

Las técnicas de Soft-Computing está compuesta por una gran variedad de algoritmos que han ido evolucionando a través del tiempo como, por ejemplo, computación difusa, computación evolutiva, redes neuronales, machine learning, deep learning, entre otras técnicas. En la Figura 2, se muestra una taxonomía de algunas de las técnicas de Soft-Computing.

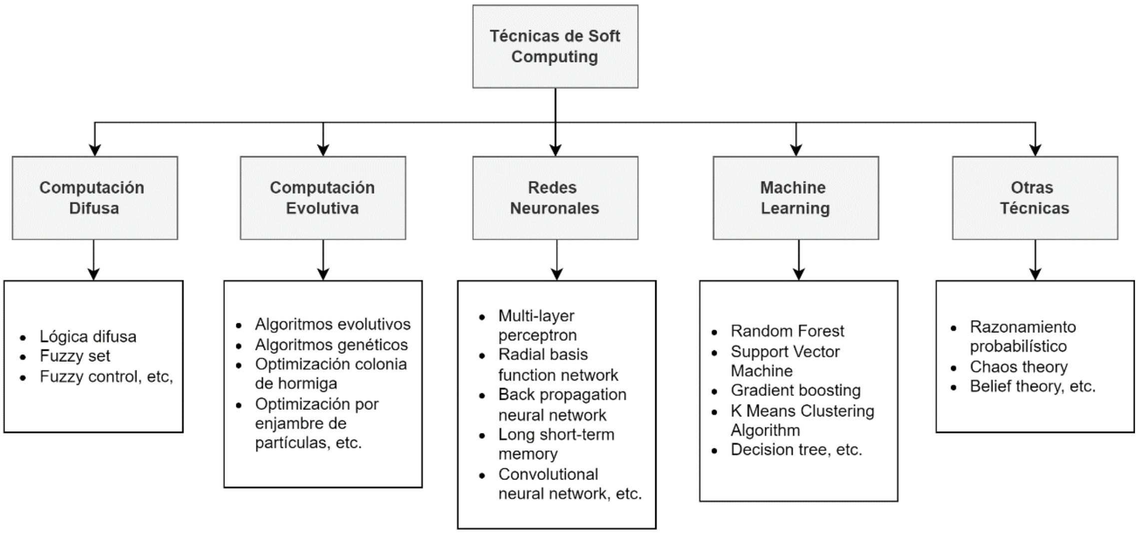


Figura 2. Taxonomía de las técnicas de Soft-Computing.

Cada una de las técnicas mostradas en la Figura 2 son utilizadas para resolver diversos problemas del mundo real. Por ejemplo, las técnicas basadas en *Computación Difusa (CD)* son empleados para resolver problemas vagos o imprecisos; es decir, permite transformar los adjetivos del lenguaje natural, con la finalidad de cuantificar el grado de pertenencia de un conjunto de datos dado [35], [39]. Además, la CD se asemeja a la capacidad humana de tomar decisiones. Esto se observa en los trabajos [40], [41], donde hacen uso de la lógica difusa para el monitoreo de pacientes. Del mismo modo, la CD se aplica en la dimensión de Smart Energy, específicamente para una mejor gestión de los sistemas de climatización, por ejemplo, en los estudios [42], [43] emplean el enfoque de lógica difusa para la gestión y control de los sistemas HVAC (Heating, Ventilation and Air-conditioning System).

De igual forma, Soft-Computing emplea la *Computación Evolutiva (CE),* esta se inspira de la propia evolución natural. La CE se aplica para la resolución de problemas de optimización, en el cual un enfoque probabilístico es utilizado para resolver estos problemas [39]. Para lograr esto, se tiene un conjunto inicial de soluciones candidatas y estas se van actualizando (mutando) de forma iterativa hasta encontrar un valor que optimice de manera global el problema dado, también conocido como fitness function. Del mismo modo, que, en las técnicas de CD, existen diversos estudios que han desarrollado investigaciones en las dimensiones de Smart Health y Smart Energy aplicando técnicas de CE [44], [45].

Por otra parte, las técnicas de *Machine Learning (ML)* son muy utilizadas en la actualidad debido a su potencial y versatilidad en los distintos sectores de la SC. Estos algoritmos aprenden directamente de la propia información que se les proporciona, por lo que son capaces de extraer información y conocimiento valioso de los propios datos [46], [47]. Los algoritmos de ML son generalmente empleados para tareas de clasificación y/o predicción. Por ejemplo, en el trabajo [48] se ha utilizado ML para predecir el consumo energético de los sistemas HVAC en edificios. De igual forma, en el estudio [49] utilizaron random forest e IoT en un sistema de monitoreo de atención médica.

Otra de las técnicas de Soft-Computing que recibe una gran atención en los últimos años son los modelos desarrollados mediante *redes neuronales artificiales (ANN)*. Las ANN se inspiran del funcionamiento de las redes neuronales biológicas. Estas son capaces de capturar patrones ocultos de la propia información histórica aprendida de los datos que se les suministran. La arquitectura de las ANN está formada por una capa de entrada, una o más capas ocultas y una capa de salida [39]. Algunos ejemplos de arquitecturas ANN son las MLP o perceptrón multicapa, RBF o radial basis function, redes SOM o mapas autoorganizados.

Dentro de este marco, la robustez y confiabilidad que brindan las ANN hoy en día, han propiciado una evolución en las arquitecturas ANN, la cual es conocido como *Deep Learning (DL)*. Las arquitecturas DL emplean procesos más complejos que les permite capturar con mayor profundidad las relaciones intrínsecas y abstractas de los datos. Algunas tareas que puede realizar estas arquitecturas o modelos DL son el reconocimiento de voz o texto, detección de objetos, procesamiento del lenguaje natural, pronóstico del consumo energético, pronóstico del clima, entre otros [50]. Como complemento, en la dimensión de Smart Health, los investigadores han desarrollado sistemas inteligentes basados en DL para predecir ataques epilépticos mediante el análisis de señales de electrocardiogramas, así como para predecir anomalías del corazón [51], [52]. Del mismo modo, en la dimensión de Smart Energy se han utilizado modelos basados en DL para predecir el consumo energético residencial, así como para controlar y optimizar los sistemas HVAC en edificios residenciales [53], [54].

1.3 Integración de servicios avanzados

Basándonos en lo descrito anteriormente, los modelos inteligentes que se desarrollen requieren ser integrados y desplegados en un ecosistema que les permita desempeñarse de forma correcta. Para lograr esto, en la actualidad existen tres paradigmas computacionales bien conocidos por parte de la comunidad científica e industrial (Cloud, Fog, Edge), los cuales proporcionan grandes beneficios y ventajas [55]–[57].

En la Figura 3 se muestra cada uno de estos paradigmas computacionales. Desde la perspectiva más general, la Cloud Computing proporciona acceso remoto a recursos computacionales tales como, almacenamiento, procesamiento, software, red, poder computacional, entre otros. Fog Computing [58] permite a los dispositivos de Internet of Things (IoT) procesar, analizar, almacenar o tomar decisiones sobre los datos antes de que sean cargados a la Cloud. Edge Computing [56] es un paradigma que se encuentra tan cerca como un salto a los dispositivos IoT; es decir, Edge Computing son aquellos dispositivos que están conectados a la red, tales como los WiFi access point, gateways, entre otros. Estos paradigmas nos ofrecen la capacidad de explorar nuevos enfoques, a fin de hacer más inteligentes los procesos del negocio o aplicación. No obstante, decidir cuál de estos se utilizará dependerá de las necesidades del proyecto o negocio [55].

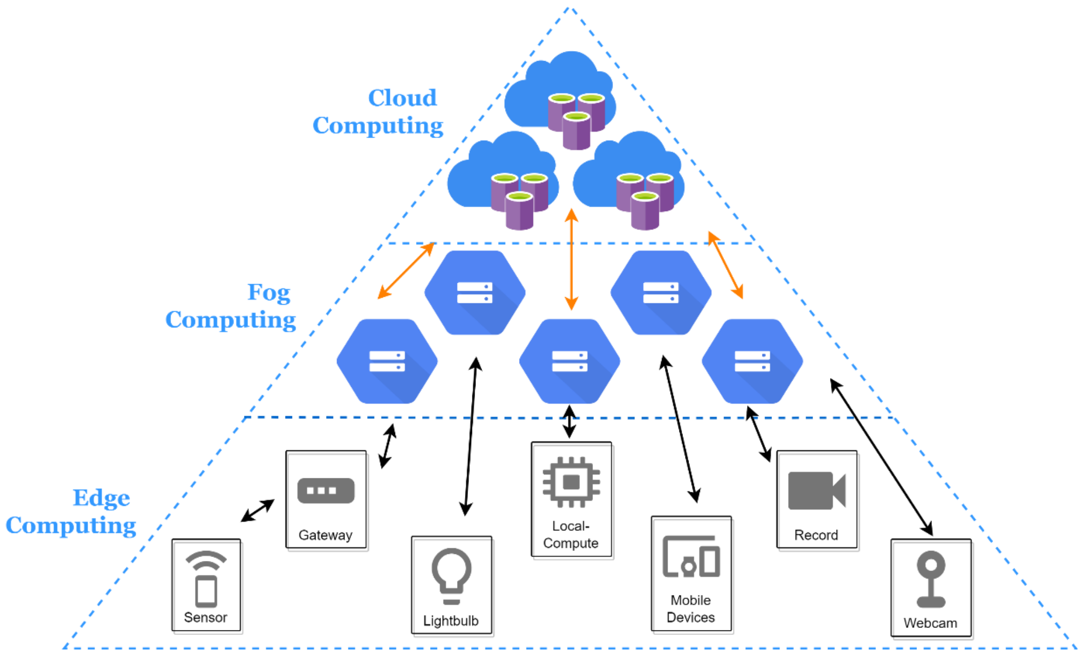


Figura 3. Paradigmas computacionales [55].

Del mismo modo, existen distintos enfoques de patrones arquitectónicos de software que se pueden implementar en los paradigmas computacionales. Estos patrones arquitectónicos presentan características que favorecen el diseño, desarrollo e implementación de sistemas inteligentes. Algunos de estos patrones arquitectónicos de software son el patrón cliente-servidor, patrón en capas, patrón maestro-esclavo, entre otros [18]. Sin embargo, existen otros patrones arquitectónicos de software que permiten desarrollar sistemas y/o aplicaciones mediante módulos (servicios), facilitando así la implementación y despliegue de sistemas inteligentes (modelos basados en técnicas de Soft-Computing).

Basándonos en lo anteriormente expuesto, el patrón de arquitectura de software orientada a servicios (SOA, por sus siglas en inglés) [59], [60] surge con la finalidad de crear uniformidad en el diseño, construcción y despliegue de los servicios para el correcto funcionamiento de los sistemas. Por lo tanto, SOA ofrece un enfoque adecuado para el desarrollo de sistemas o aplicaciones como servicios [61]. Sin embargo, es importante resaltar que los servicios que sean integrados al sistema deben cumplir con las características de reutilización, interoperabilidad y extensibilidad [62].

Por otro lado, uno de los patrones arquitectónicos que está recibiendo una gran atención por parte de la comunidad científica e industrial, es el patrón de arquitectura de microservicios (MSA) [63]. La MSA surge como una extensión del patrón SOA, donde este propone el diseño de aplicaciones o sistemas mediante el desarrollo de un conjunto de servicios pequeños y autónomos. En otras palabras, cada uno de los servicios es independiente del otro y deben satisfacer una funcionalidad del negocio o proyecto dentro del contexto dado.

Existen varios trabajos que proponen o implementan algunos de los patrones arquitectónicos orientados a servicios (SOA o MSA), dentro de las dimensiones de Smart Health y Smart Energy. Por ejemplo, en la dimensión Smart Health se ha propuesto trabajos como en [64], donde utilizan el patrón SOA para ofrecer una solución de salud móvil (mHealth) para profesionales de la salud ubicados en áreas remotas. Asimismo, en el estudio [65] proponen un sistema de monitorización de pacientes mediante el patrón SOA. De igual forma, el patrón MSA está siendo utilizado en distintos contextos de Smart Health, como en el trabajo [66], donde los investigadores proponen un sistema basado en MSA que recopila información de diversos dispositivos portátiles y, por tanto, utilizan dicha información para la construcción de un modelo inteligente que evalúa el estado de fragilidad. Igualmente, en el estudio [67] proponen un chatbot interactivo basado en MSA para el apoyo a pacientes crónicos. Por otra parte, en la dimensión de Smart Energy, los autores del trabajo [68] emplearon SOA para el monitoreo del desempeño de la ecoeficiencia en las empresas manufactureras. Mientras que, en los estudios [69]–[72], implementaron MSA y SOA para la gestión de energía en edificios.

1.4 Planteamiento del problema

Basándonos en los hechos anteriormente descritos, la presente tesis doctoral se centrará en dos dimensiones claves de la sociedad digital, específicamente, la dimensión de Smart Health y Smart Energy. Ya que, en los últimos años estos han recibido especial atención por parte de la comunidad científica e industrial, además de que exponen grandes desafíos para la SC [23], [73]. Asimismo, estas dimensiones requieren de servicios avanzados que sean capaces de proporcionar un soporte y/o facilidades en la toma de decisiones inteligente en los diferentes procesos dinámicos que los conforman.

1.5 Objetivos de la tesis doctoral

A continuación, se describe el objetivo general y los objetivos específicos de esta tesis doctoral.

1.5.1 Objetivo general

Desarrollar servicios avanzados que faciliten la toma de decisiones de los principales actores en las dimensiones claves de la sociedad digital.

1.5.2 Objetivos específicos

Para lograr el objetivo general se han definido los siguientes objetivos específicos:

I. Analizar las arquitecturas de software que faciliten la integración de servicios avanzados en las dimensiones claves de la sociedad digital.

II. Identificar las técnicas de Soft-Computing que permitan detectar patrones en los datos de las dimensiones claves de la sociedad digital.

III. Crear modelos representativos a través de las técnicas de Soft-Computing.

IV. Demostrar que los servicios avanzados basados en Soft-Computing impulsan la toma de decisión inteligente en las dimensiones claves de la sociedad digital.

1.6 Estructura de la memoria de Tesis Doctoral

El artículo 5 del Reglamento de Elaboración, Autorización y Defensa de la Tesis Doctoral [74], que engloba el programa de Doctorado de Ingeniería de la Información y del Conocimiento, señala la posibilidad de realizar una tesis doctoral como compendio de artículos de investigación. Con el fin de garantizar la calidad del trabajo doctoral, el número mínimo de artículos científicos será de tres y estas deben ser publicaciones de reconocido prestigio. Se entenderá por publicaciones de reconocido prestigio las utilizadas para la obtención de complementos de investigación (sexenios) en el ámbito en el que se desarrolle la investigación. La Tesis deberá incluir, además de los artículos, un resumen amplio que, de coherencia al conjunto de la investigación, en el que se muestre la línea argumental de la misma, así como un capítulo de conclusiones.

Para la realización de la Tesis Doctoral, se han seleccionado las siguientes cinco publicaciones en el que el autor ha desarrollado la labor investigadora de esta tesis:

1. **L. Mendoza-Pitti,** H. Calderón-Gómez, M. Vargas-Lombardo, J. M. Gómez-Pulido and J. L. Castillo-Sequera, *"Towards a Service-Oriented Architecture for the Energy Efficiency of Buildings: A Systematic Review,"* in IEEE Access, vol. 9, pp. 26119-26137, 2021, doi: 10.1109/ACCESS.2021.3057543.

2. **L. Mendoza-Pittí**, A. Garcés-Jiménez, J. Aguilar, J.M. Gómez-Pulido, M. Vargas-Lombardo, "*Proposal of physical models of multi-HVAC systems for energy efficiency in smart buildings"* . In 2019 7th International Engineering, Sciences and Technology Conference (IESTEC) (2019), IEEE, pp. 641-646, doi: 10.1109/IESTEC46403.2019.00120.

3. **L. Mendoza-Pittí,** H. Calderón-Gómez, J. M. Gómez-Pulido, M. Vargas-Lombardo, J. L. Castillo-Sequera, and C. S. de Blas*, "Developing a Long Short-Term Memory-Based Model for Forecasting the Daily Energy Consumption of Heating, Ventilation, and Air Conditioning Systems in Buildings,"* Applied Sciences, vol. 11, no. 15. 2021, doi: 10.3390/app11156722.

4. J.J., Saldaña Barrios, **L. Mendoza,** E. Pitti & M. Vargas. (2018). "*Ubiquitous and ambient-assisted living eHealth platforms for Down's syndrome and palliative care in the Republic of Panama: A systematic review*." Health Informatics Journal, 24(4), 356-367, doi:10.1177/1460458216671560.

5. **L. Mendoza-Pittí**, J. M. Gómez-Pulido, M. Vargas-Lombardo, J. A. Gómez-Pulido, M. -L. Polo-Luque and D. Rodréguez-Puyol, *"Machine-Learning Model to Predict the Intradialytic Hypotension Based on Clinical-Analytical Data,"* in IEEE Access, vol. 10, pp. 72065-72079, 2022, doi: 10.1109/ACCESS.2022.3189018.

Como seguimiento a las directrices de la normativa que comprende la modalidad de tesis por compendio, esta sigue la siguiente estructura:

I. En la primera sección, se presenta una introducción a los conceptos desarrollados en la tesis doctoral, así como al problema que dio pie al desarrollo de esta investigación. De igual forma, se presenta el objetivo general y los objetivos específicos de la tesis.

II. En la segunda sección se presentan cada uno de los artículos seleccionados para la presentación de la tesis doctoral como compendio de publicaciones. En esta se hace un breve resumen previo al artículo, se describe las aportaciones de la publicación y se detallan los índices de calidad de la publicación.

III. En la tercera sección se presentan otras publicaciones científicas realizadas, así como algunos trabajos colaborativos realizados

IV. En la cuarta sección se presenta la experiencia investigadora del doctorando.

V. En la quinta sección se presentan algunos méritos logrados durante el desarrollo de la investigación.

VI. En la sexta sección se presentan las conclusiones y los trabajos futuros resultantes de esta tesis doctoral.

2 Compendio de publicaciones

La meta de esta investigación de tesis doctoral es facilitar la toma de decisiones inteligentes de los principales actores de las dimensiones claves de la sociedad digital, mediante el desarrollo de servicios avanzados. Para evidenciar la labor investigadora realizada se redactaron cinco artículos que fueron publicados en diferentes congresos y revistas científicas de gran relevancia, por lo que describen los aspectos importantes para el desarrollo de servicios avanzados que permitan mejorar el confort, bienestar y calidad de vida de la sociedad.

Dentro de este compendio de publicaciones, el primero de ellos trata de una revisión sistemática que identifica y analiza avanzadas arquitecturas orientada a servicios inteligentes en el contexto de la mejora de la eficiencia energética para apoyar la toma de decisiones respecto al funcionamiento y/o comportamiento óptimo de los sistemas energéticos de los edificios. La segunda publicación propone el uso de modelos físicos de sistemas Multi-HVAC para la eficiencia energética en un edificio en concreto. El resultado de este estudio nos permitió identificar aquellos parámetros claves en el comportamiento de los sistemas HVAC. La tercera publicación propone un modelo de Deep Learning, basados en redes LSTM que abordan series temporales, para el pronóstico del consumo energético diario de los sistemas de climatización de los edificios, para optimizar la eficiencia y sostenibilidad del edificio incorporando factores como las horas de trabajo, las actividades del edificio, el clima y la temperatura interior. La cuarta publicación presenta dos plataformas de eHealth que muestran cómo los sistemas sanitarios están migrando de una arquitectura cliente-servidor a un paradigma basado en la web y ubicuo. Esto consigue que, se mejore la forma en que se prestan los cuidados a los pacientes terminales, debido a la aplicación de métodos de aprendizaje automático y un repositorio de resultados de pacientes centrado en datos y ubicuo. Finalmente, la quinta publicación propone un modelo basado en Machine Learning que es capaz de predecir al inicio de la sesión de hemodiálisis si un paciente sufrirá o no de hipotensión durante su tratamiento. El modelo desarrollado se fundamenta en un elevado conjunto de datos en el que se unifican la información preveniente de variables clínicas proveniente de la diálisis con la aportada por las analíticas de sangre del paciente.

Con el fin de identificar que las publicaciones realizadas cumplen los objetivos establecidos presentados en esta tesis doctoral, se resalta en la Tabla 1 a continuación, el cuadro en gris por cada objetivo cumplido respecto a la publicación realizada. Indicando de esta manera que se alcanzó el objetivo previamente establecido con la publicación.

Tabla 1. Publicaciones vs Objetivos generales

Publicaciones	**Obj. I**	**Obj. II**	**Obj. III**	**Obj. IV**
Towards a Service-Oriented Architecture for the Energy Efficiency of Buildings: A Systematic Review				
Proposal of physical models of multi-HVAC systems for energy efficiency in smart buildings				
Developing a Long Short-Term Memory-Based Model for Forecasting the Daily Energy Consumption of Heating, Ventilation, and Air Conditioning Systems in Buildings				
Ubiquitous and ambient-assisted living eHealth platforms for Down's syndrome and palliative care in the Republic of Panama: A systematic review				
Machine-Learning Model to Predict the Intradialytic Hypotension Based on Clinical-Analytical Data				

2.1 Publicación 1: Towards a Service-Oriented Architecture for the Energy Efficiency of Buildings: A Systematic Review

2.1.1 Resumen de la publicación

El primer artículo que forma parte de esta tesis doctoral trata de una revisión sistemática exhaustiva de la literatura en el que se busca identificar y analizar aquellos requerimientos de software que permitan desarrollar una arquitectura orientada a servicios (SOA) para la eficiencia energética de los edificios, así como identificar las características de big data que juegan un rol esencial en la gestión energética de los edificios. En este sentido, uno de los componentes del edificio que mayor aportan al consumo energético total final de los edificios son los sistemas HVAC, por lo que es esencial reducir su consumo energético a fin de mejorar la eficiencia energética de los edificios. Estos sistemas HVAC son gestionados, monitorizados y controlados mediante el sistema de gestión de edificio (BMS en inglés); sin embargo, el BMS tiene limitaciones para la integración de otros módulos (servicios) y, además, tiene la posibilidad de colapsar al gestionar los datos masivos (big data) generados por los diferentes componentes del edificio. De allí, pues que, en este artículo se identificó y analizó que la SOA y la MSA podrían coexistir para ofrecer grandes beneficios y suplir los requerimientos de software de interoperabilidad y versatilidad necesarios dentro de nuestro contexto de estudio. Además, se identificaron que las características de big data que mejor se adaptan para una gestión eficiente de la energía en edificios son velocidad, volumen, variedad, valor y veracidad. Por consiguiente, los resultados de este estudio proveen a los investigadores un conjunto de requerimientos de software y características de big data centrados a la implementación de arquitecturas de software orientadas a servicios a fin de mejorar la eficiencia energética de los edificios.

2.1.2 Contribuciones significativas de la publicación

En esta primera publicación, el doctorando realizó diferentes contribuciones significativas a la comunidad científica: (i) se identificaron aquellas características de Big Data (BD) que propician una correcta gestión de la energía en los edificios; (ii) se realizó un análisis exhaustivo de los diferentes requerimientos de software (funcionales y no funcionales) descritos en la literatura que permiten el desarrollo y despliegue de servicios avanzados versátiles e interoperables para la gestión eficiente de los sistemas energéticos en edificios; (iii) se identificó que los patrones arquitectónicos SOA y MSA pueden coexistir para generar un entorno robusto y confiable, a fin de desarrollar, implementar y desplegar servicios avanzados de acuerdo con las necesidades actuales de la sociedad digital; (iv) permite llenar el vacío de conocimiento y apoyar a los investigadores respecto a las formas de adaptar y desarrollar las arquitecturas orientada a servicios avanzados en el contexto estudiado.

2.1.3 Índices de calidad de la publicación

Autores: L. Mendoza-Pitti, H. Calderón-Gómez, M. Vargas-Lombardo, J. M. Gómez-Pulido and J. L. Castillo-Sequera.

Título: Towards a Service-Oriented Architecture for the Energy Efficiency of Buildings: A Systematic Review

Fecha de publicación: 5 de febrero de 2021

IEEE ACCESS, 2021

Volumen. 9, páginas. 26119-26137

ISSN: 2169-3536

DOI: 10.1109/ACCESS.2021.3057543

Ranking JCR 2020: Computer science, information systems (Q2)

Factor de impacto (JCR 2020): 3.367

Ranking SJR 2020: Computer Science (Q1)

Factor de impacto (SJR 2020): 0.59

2.1.4 Publicación

Received January 20, 2021, accepted January 30, 2021, date of publication February 5, 2021, date of current version February 17, 2021.
Digital Object Identifier 10.1109/ACCESS.2021.3057543

Towards a Service-Oriented Architecture for the Energy Efficiency of Buildings: A Systematic Review

LUIS MENDOZA-PITTI[1,2], HURIVIADES CALDERÓN-GÓMEZ[1,2], MIGUEL VARGAS-LOMBARDO[2], (Senior Member, IEEE), JOSÉ MANUEL GÓMEZ-PULIDO[1], AND JOSÉ LUIS CASTILLO-SEQUERA[1]
[1]Department of Computer Science, University of Alcala, 28801 Alcala de Henares, Spain
[2]E-health and Supercomputing Research Group (GISES), Technological University of Panama, Panama 02290149, Panama

Corresponding author: Miguel Vargas-Lombardo (miguel.vargas@utp.ac.pa)

This work was supported in part by the National Secretariat for Science, Technology and Innovation (SENACYT, Panama) and the Institute for the Training and Use of Human Resources (IFARHU), for the doctoral study scholarship, in part by the SENACYT through the National Research System (SNI-SENACYT) National Research System of which one of the authors is a member and in part by the "Smart Energy" Campus of International Excellence, promoted by the Universidad de Alcalá and the Universidad Rey Juan Carlos (Spain).

ABSTRACT Currently, smart buildings generate large amounts of data due to the many devices and equipment available. Hence, buildings implement building management systems (BMSs), which monitor, control, manage and analyze each of these components. However, current BMSs are incapable of managing a massive amount of data (big data) and therefore cannot extract knowledge or make intelligent decisions in quasi real time. In addition, there are serious limitations to integrating BMSs with other services since they generally use proprietary software. In this sense, service-oriented architecture (SOA) is an architectural style that allows one to build distributed systems and provide functionalities such as services to end users or other types of services. Therefore, an SOA has the great advantage of allowing the expansion of the functionalities of BMSs. In fact, there are several studies that address SOAs for building management. However, we have not found any description or systematic analysis in the literature that allows the development of a versatile and interoperable SOA focused on the energy efficiency of buildings and that can integrate massive data analysis features. For these reasons, this study seeks to fill this knowledge gap and, more specifically, to identify and analyze the various software requirements proposed in the literature and the characteristics of big data that allow for improving the energy efficiency of buildings. To this end, we performed an in-depth review of the literature according to the methodology proposed by Kitchenham. As a result of this review, we provide researchers with a specific vision of the requirements and characteristics to consider for software development aimed at the energy efficiency of unique or historic buildings.

INDEX TERMS Big data applications, buildings, energy efficiency, energy management, management information systems, service-oriented systems engineering, software architecture.

I. INTRODUCTION

Buildings are responsible for more than a third of the world's total energy consumption [1]–[4]. Therefore, reducing the energy consumption of buildings is a global problem. That is why countries are implementing strategies to make buildings more efficient and reduce excessive energy consumption. For example, in the design phase, Zero Energy Buildings (ZEBs) [5]–[7] have gained special attention in recent years as they make use of renewable energy sources [8], [9], thus allowing the energy efficiency indexes of a building to be improved [10]–[12]. However, in the operation phase, it is not a simple task when dealing especially with historical or unique buildings, so that immediate actions are based on reducing energy consumption [13].

In this sense, heating, ventilation and air conditioning (HVAC) systems contribute a significant part of the building's total energy consumption [14]. In addition, it is affected by various factors [15]. Consequently, to effectively improve energy efficiency and reduce energy consumption in buildings, we have conducted several investigations [16]–[18] with the aim of optimizing the performance of HVAC systems in historic or unique buildings.

The associate editor coordinating the review of this manuscript and approving it for publication was Lei Wang.

On the other hand, HVAC systems are tuned by building management system (BMS), whose purpose mainly consists of monitoring, managing, controlling and analyzing the different equipments (HVAC systems, elevators, lighting, etc.) and devices (sensors, actuators, etc.) of a building [19]. However, being a proprietary software system, the BMS presents serious limitations for the integration with other components or services, so it can crash when managing a large amount of data (big data) [20]–[23]. Hence, it is necessary to implement a software architecture capable of supporting different functionalities for the intelligent management of the building.

For this reason, components (services) must be able to be integrated and deployed in versatile way within a software architecture. Hence, we are conducting a project called ''Intelligent management system for optimizing energy consumption in building air conditioning'', driven by a consortium of Spanish universities under the Smart Energy Campus of International Excellence [24]. The goal is to develop an autonomous BMS using a versatile software architecture, which allows the energy efficiency of HVAC systems in historic or unique buildings to be improved.

In this perspective, service-oriented architecture (SOA) offers an architectural design focus suitable for the design and development of distributed systems that offer different capabilities, including end-user services and others [25]–[27]. As it can be inferred, SOA could expand the traditional features of a BMS in a way that would improve the management of the building's equipment and devices and consequently improve the energy efficiency of the building. However, before considering this in depth, it is first necessary to identify the software functionalities (features) that are best adapted to these systems.

In fact, several studies have addressed different features for an SOA aimed at building management. However, at this time, we have not found any article that provides an overview or a systematic analysis of the different features necessary for an SOA aimed at energy efficiency in buildings.

For such reasons, we conducted a systematic literature review (SLR) using the methodological procedure proposed by Kitchenham [28]. Specifically, we identify and analyze the different requirements (functional and non-functional) for a service-oriented architecture that incorporates the features (interoperability and versatility) as well as big data characteristics that will result in improved energy efficiency for buildings. We asked a series of research questions and conducted an extensive search process that allowed us to find answers to our questions. As a result, this SLR can provide researchers with a general idea of the requirements and characteristics on which investigations should focus to adapt and develop software architectures and new technologies in this field.

This article is organized as follows: Section II briefly describes the related work and the motivation of the study conducted. Section III describes the methodological procedure used for the compilation of studies, the research questions, search engine, search strategy and selection criteria. Section IV summarizes the process of applying the methodology obtained through the SLR. Section V shows the results and relevant papers that answer the research questions. Section VI presents the discussion of the results. Section VII outlines some directions for future research; and finally, Section VIII provides the conclusions derived from this study.

II. RELATED WORK AND MOTIVATION

In this section we describe some of the different software functionalities (features) that are either proposed or incorporated in the documents found. Additionally, we explain the motivation for conducting this SLR.

A. RELATED WORKS

In [29], an SOA was used in the industrial system of systems (SoS) since it provides an excellent platform to develop systems and allows the encapsulation, reuse and composition of components (services). The authors claim that services can be essential to get the desired information as in the case of the energy consumption of equipments in an industrial factory.

For its part, in [30] an SOA was adopted for the management of home energy systems which have heterogeneous components and are subject to different standards, requirements and technologies. This is mainly due to the fact that SOA allows the interoperability and flexibility issues present in the home to be addressed.

On the other hand, in [25] the SOA-based approach was used to enable the interoperability of heterogeneous smart home systems. They developed a simulator in which they incorporated different services, optimization criteria and modes of operation in order to measure the energy consumed by the smart home.

Meanwhile, in [31] they described the implementation of an SOA-based infrastructure using a service-based software layer that allows information to be obtained from various data sources that is, in turn, consumed by other applications within a factory. This allowed them to monitor the operational performance and energy consumed in one business and to meet the requirements of other applications.

Finally, in [32] an SOA was proposed for the building service integration, since SOA provides flexibility, extensibility, open design and interoperability. Therefore, business applications can be highly adaptable. The authors note that this approach would facilitate the integration of BMS and management applications along with other functions in order to share, monitor, control and manage the business environment. Table 1 shows each of the software approaches and functionalities (features) applied by the different investigations cited.

It should be noted that the aforementioned works have outlined various SOA-based approaches to building energy management and have addressed one or more features. Within this context, one of the most mentioned ones is interoperability; therefore, SOA brings in this important feature that we want to address in our investigation.

Conversely, flexibility is another feature mentioned in the studies, which is defined as the capacity of services or

TABLE 1. Summary of the approaches and software functionalities in the cited works.

Approach	Software Functionality	Refs.
Several industrial sectors	Reusability and Composition	[29]
Household management	Interoperability and Flexibility	[30]
Household management	Interoperability	[25]
Manufacturing enterprises	Interoperability	[31]
Building services integration	Flexibility, Extensibility, Open design and Interoperability	[32]
Building energy efficiency	Interoperability and Versatillity	Our work

systems to adapt to changes within a given infrastructure (a language, a database, an operating system, among others) [33]. As it can be inferred, flexibility is similar to versatility; however, for our own work, the latter is defined as the ability to build, integrate and implement components (services) that are independently and completely decoupled in different containers or servers, different languages, operating systems, databases, to mention a few [34].

B. MOTIVATION

On this basis, the abovementioned works do not identify or report versatility. Therefore, this motivated us to conduct an SLR in order to identify the studies that consider this feature through the software requirements that they propose or report in the studies, as well as the big data characteristics that best represent the energy management of buildings.

In this respect, our study provides a new perspective compared to what has been described in previous studies when addressing software functionalities. This analysis will solve several needs. In the first place, each component will be able to focus on specific building systems monitoring, automation, and control tasks, which may require CPU or GPU dependent processing (e.g. neural networks, genetic algorithms, etc.). In second place, it is necessary to control and manage the collection of large volumes of data generated from different sources and types of data (big data). Last but not least, it is necessary to integrate and implement each of the services in a scalable, interoperable and versatile way to form an agile and efficient data management system.

For these reasons, we consider that the most suitable features to carry out our study are interoperability and versatility, since they would allow us to satisfy the needs raised for the development and deployment of a service-oriented architecture in the context of energy efficiency of buildings.

III. METHODOLOGY

The methodological approach used for this study was that proposed by Kitchenham [28]. We limited our SLR to the context of the service-oriented architecture and the characteristics of big data applicable to the energy management of buildings. In the following sections, we describe the steps of the methodology employed, including the research questions we sought to answer through this study and the search strategy used to compile the different works for the SLR; establish the criteria for the selection of the studies; and finally describe the data extraction process that will be conducted to answer the research questions.

A. RESEARCH QUESTIONS

In this work, we focused on the requirements (functional and non-functional) for an SOA and the characteristics of big data applicable to energy management that has been developed for the energy efficiency of buildings. For this reason, Table 2 describes the research questions to be answered.

TABLE 2. Research questions for systematic review.

RQ1	What characteristics best represent an architecture that enables the management and application of big data to the energy management of buildings?
RQ2	What are the functional and non-functional requirements for the types of software architectures used in the energy management of buildings?
RQ3	What implementations of service-oriented software architectures have been developed for the energy efficiency of buildings?

B. SEARCH STRATEGY

1) SEARCH KEYWORDS

To construct a search string, it was necessary to perform an in-depth study of the state-of-the-art research based on the following keywords: building energy efficiency, building energy consumption, energy efficient building, software architecture, service-oriented architecture, SOA, and big data. All of these terms are related to a specific problem addressed.

Then, we built a search string combining the indicated words with different Boolean operators as follows:

(“building energy efficiency” OR “building energy consumption” OR “energy efficient building”) AND (“software architecture” OR “service-oriented architecture” OR “Big data” OR “SOA”)

2) DATA SOURCES

To collect as many relevant studies as possible, we selected the main scientific databases (Scopus, Google Scholar, Web of Science, Science Direct, IEEE Xplore, SpringerLink, and ACM Digital Lib). Some of the criteria used to select these databases were as follows: (i) to retrieve the largest number of studies regarding service-oriented software architectures, (ii) to recover the largest number of studies regarding the energy efficiency of buildings, and (iii) to cover most of the journals with an impact factor.

3) SEARCH PROCEDURE

The following procedure was used to select the main sources for our research. First, the search string was executed in each of the search engines of the selected scientific databases. Second, the titles and abstracts of the articles were prefiltered by applying the selection criteria. Third, a complete review of each of the preselected articles was performed, considering the selection criteria, to collect the sources that provide

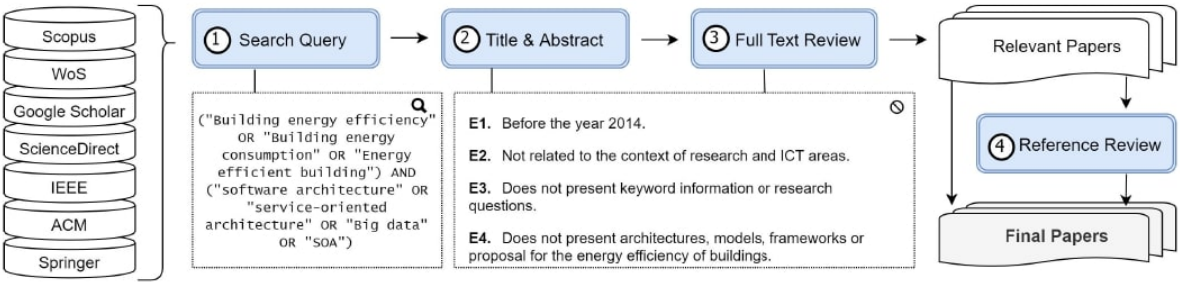


FIGURE 1. Diagram of the search procedure for the SLR execution.

relevant information for the research questions. Fourth, the full texts were thoroughly reviewed, and we retrieved the most appropriate references for our study that provided important information and that were pre-dated 2014 to add to these articles. This procedure is shown in Fig. 1.

C. SELECTION OF STUDIES

To identify the relevant articles in the second and third steps of the search procedure, we set the following exclusion criteria for use with the protocol defined above:

E1. Before the year 2014.
E2. Not related to the context of the research and ICT areas.
E3. Does not present keyword information or research questions.
E4. Does not present architectures, models, frameworks or proposals for the energy efficiency of buildings.

D. DATA EXTRACTION

We documented the entire search process with all records stored in a reference manager. Similarly, we noted the exclusion criteria for each of the excluded works. Additionally, each of the full texts was recovered only if it passed the prefiltering stage. Once this was done, we created a form where we collected the following information for each of these works:

- Does the article present concepts or definitions of the keywords?
- Does the article present characteristics of an architecture based on big data?
- Does the article present the characteristics that represent a software architecture that incorporates the use of big data for energy management?
- Does the article present the functional and non-functional requirements of software architectures aimed at the energy and building sector?
- Does the article mention or consider the limitations of service-oriented software designs, proposals, models, frameworks, platforms or architectures focused on the energy efficiency of buildings?
- In what aspect of energy efficiency is the service-oriented architecture used?

Considering this information, we answered each of the research questions posed.

IV. EXECUTION OF THE REVIEW

The search procedure was conducted in two phases. In the first phase, the relevant sources describing the articles collected in each scientific database were identified. The exclusion criteria were also applied to obtain articles that provided relevant information to answer the research questions. In the second phase, we considered the most relevant references obtained in the first phase. These phases are described below.

A. FIRST PHASE: IDENTIFYING THE RELEVANT SOURCES

Fig. 2 summarizes the numbers of articles obtained after excluding the studies found in the scientific databases. As a complement, Fig. 2 also shows the selection criteria for both the pre-filtering (only the titles and abstracts of the articles were considered) and the complete revision of the texts.

Considering this, the first exclusion criterion (E1) was applied only in the pre-filtering stage, which allowed us to exclude more than 600 publications that were outside the range of the selected years. The second exclusion criterion (E2) led to more than 3,600 exclusions, leaving those retained as the most representative since it made it easier for us to exclude articles that were outside the context of the investigation. The third exclusion criterion (E3) allowed us to discard more than 1,000 articles that did not have information on the selected keywords or RQs. Finally, the fourth exclusion criterion (E4) was especially useful since it allowed us to discard more than 300 publications without any reference to designs, proposals, models, frameworks or architectures.

B. SECOND PHASE: REVISION OF THE REFERENCES

In this second phase, all the references presented in the articles obtained in the first phase were reviewed to identify the useful information in them. We considered only those articles that satisfied each of the exclusion criteria with the exception of E1 since it applied only to the first phase. Finally, we managed to find 15 articles with information relevant to our research context, resulting in a total of 97 articles. In Fig. 3, the process of executing the SLR is summarized, and each of the phases described above is detailed.

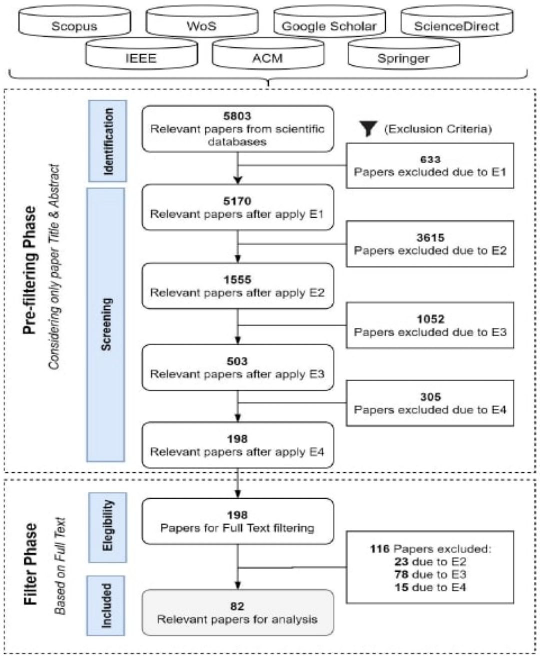


FIGURE 2. Results of the first phase after applying each selection criterion.

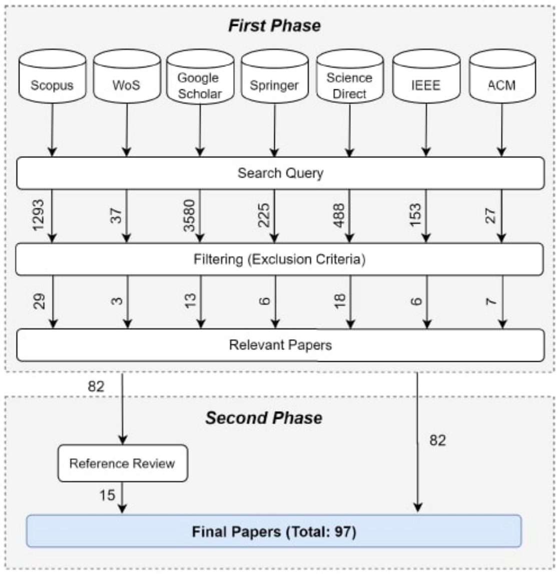


FIGURE 3. Summary of the SLR execution, showing the numbers of papers obtained in each phase.

V. RESULTS

This section presents the results of the relevant papers selected in terms of the publication year and publication type. In addition, this section also presents the answers to each research question posed, such as the various terminologies and existing concepts regarding the application of the concepts of big data and service-oriented architecture to various approaches, applications and styles of software architectures used in the context of the energy efficiency of buildings. As a complement, the various applications and approaches that were found, which were analyzed and classified in a prior study [35] and in this study, are shown in the Annex (Table 6).

A. RELEVANT WORKS BY YEAR

Fig. 4 describes the papers in terms of years of publication. We see that there are articles published in years prior to the exclusion criterion (E1) established above. The explanation for this is that the review of the references and analysis of the works published prior to criterion E1 allowed us to determine the theoretical basis necessary for our study context. Furthermore, we found that there was growing interest in the scientific community in the development of service-oriented architectures for energy efficient buildings. In addition, a decrease in the number of studies starting in 2019 was observed due to new trends and innovations that are emerging in the context of software architectures. An obvious example is the new microservice paradigm [36]–[38].

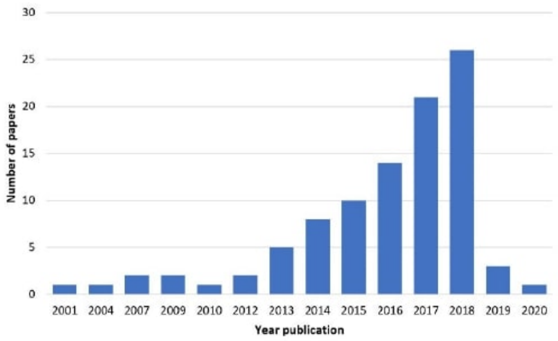


FIGURE 4. Number of papers by year of publication. Reviewed until September 2020.

B. RELEVANT WORKS BY TYPE OF PUBLICATION

Fig. 5 shows the distribution of the relevant works based on the publication type, highlighting journals with 57 relevant works, followed by conference proceedings with 24 works and books with 13 works. Finally, the publication types with the fewest works that were considered were two web pages and one general documents.

C. RQ1. WHAT CHARACTERISTICS BEST REPRESENT AN ARCHITECTURE THAT ENABLES THE MANAGEMENT AND APPLICATION OF BIG DATA TO THE ENERGY MANAGEMENT OF BUILDINGS?

The SLR allowed us to retrieve a large number of relevant studies to answer this research question.

It is true that the characteristics of big data found in the literature vary considerably. First, Laney [39] conceptualized

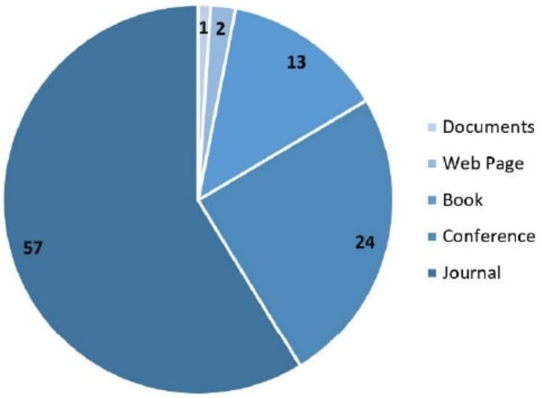


FIGURE 5. Number of papers by type of publication.

big data in terms of its three main characteristics (volume, velocity, and variety). Subsequently, other studies [40]–[49] analyzed seven characteristics. Some studies even considered more than 13 characteristics [42], [44], [50]–[68]. However, it was found that the four classic characteristics of big data, including volume, variety, velocity and value, are the ones that best represent the energy management of buildings [23], [43], [69]–[71]. Another important big data characteristic to consider in addition to the four mentioned above is veracity [42], [44]. In the current SLR being conducted, these five big data characteristics (see Fig. 6) are a set of functional requirements that a service-oriented architecture must integrate.

These characteristics are briefly defined below within the building energy efficiency field.

- ***Volume:*** Buildings constantly provide enormous amounts of data that come from various devices (sensors, smart meters, actuators, etc.) and equipment (HVAC, lighting, communications, etc.). In this sense, modern software architectures use technologies or services to handle massive data, as is the case with Industry 4.0 and the increasingly widespread use of Internet of Things (IoT) devices [72]. For all these reasons, this characteristic is of great relevance for the efficient energy management of buildings due to the large volume of information to be analyzed.
- ***Velocity:*** Developing and implementing services capable of rapidly collecting, analyzing and processing large amounts of data will offer several opportunities for improvements in building management. Therefore, these services must be able to handle data in different time intervals [70], allowing decision making in real time or near real time.
- ***Variety:*** Building data show a high degree of variety. In general, these data are structured (building energy consumption data, historical databases, etc.), semi-structured (meteorological data, schedule of different uses of the building, etc.) and unstructured (data on the behaviors and patterns of the occupants of buildings, etc.) [43], [73]. Therefore, the development of a service-oriented architecture would allow the integration and processing of the different types, sources and forms of data.
- ***Value:*** It is possible to implement intelligent strategies for the treatment of the collected data using algorithmic methods, allowing the acquisition of valuable knowledge that allows the optimization and improvement of the energy use of buildings [23]. In fact, these strategies can be implemented as services or software components that are in charge of detecting the energy consumption patterns of building equipment and generating and distributing energy according to demand (response to demand).
- ***Veracity:*** This characteristic refers to the noise present in the data. In the context of buildings, this noise is frequent due to failures or disconnections of the sensors or installed equipment. However, analyzing these data through various data mining techniques will offer the opportunity to detect hidden patterns and extract truthful and reliable information for decision making on the management of the building [42], [44].

D. RQ2. WHAT ARE THE FUNCTIONAL AND NON-FUNCTIONAL REQUIREMENTS FOR THE TYPES OF SOFTWARE ARCHITECTURES USED IN THE ENERGY MANAGEMENT OF BUILDINGS?

A system architecture is an abstract model that represents the organization, behavior and collaboration between the components that make up a system [74]. In this sense, software architectures have different styles that define how their structure is organized [75]. To deepen these architectural styles, we reviewed the following works [74], [76].

In addition, functional requirements (FRs) describe how the systems should behave functionally, whereas non-functional requirements (NFRs) describe the restrictions on the services or functions offered by a system [77]–[80]. Therefore, to develop a service-oriented architecture (SOA) that incorporates the software features (interoperability and versatility) to improve the energy efficiency of buildings, it is of great importance to identify and to analyze the FRs and NFRs that were considered within the analyzed studies. Therefore, below, we briefly describe the requirements found.

1) FUNCTIONAL REQUIREMENTS

Table 3 summarizes each of the functional requirements reported in the analyzed studies. According to the review of the works, we recovered 32 works that presented different FRs. However, the various authors described them in different ways. Therefore, they were grouped according to the context or similarity of their focuses. Next, we describe each of these FRs:

FR1. ***Data sources layer:*** Most studies addressed this requirement, which includes all data, regardless of whether they are structured, semi-structured or

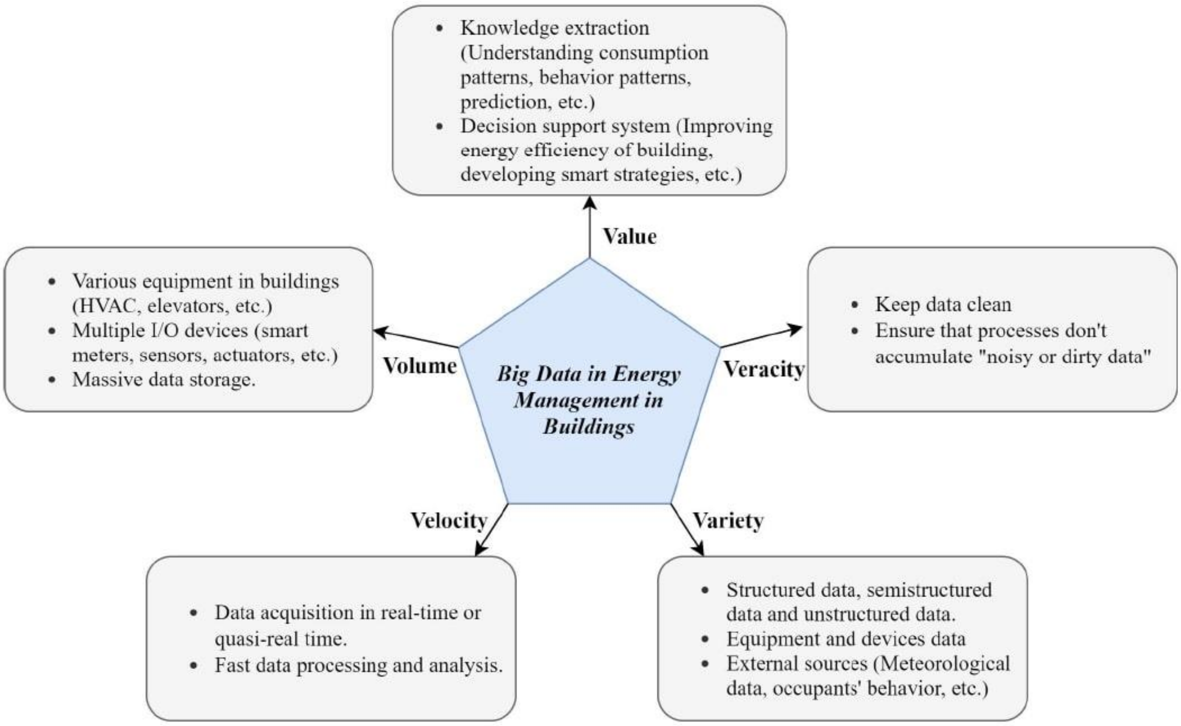


FIGURE 6. Characteristics of big data in energy management in buildings.

non-structured. Applications were based on different techniques for each type of data, as in the work of [81]; or they transformed analog to digital data, as in 82].

FR2. ***Networking layer:*** Not all the works that were analyzed study how to transmit or communicate the collected data. For example, in [82], the authors grouped the data based on explicit variables for further analysis; and in [83], the authors stated in their proposal that data must communicate correctly with the services. However, in [84], the authors used this layer as an intermediary between the data and the applications.

FR3. ***Data cleaning and preprocessing layer:*** Some of the investigations focused on data cleaning and processing. This layer is important since not all data are correct, and they generally come from various sources. In [82], the authors preprocess data that are out of range or have missing values. In addition, some works use techniques such as extraction, transformation and loading (ETL) in this layer to preprocess the data before they are loaded into the data storage layer [85]–[87].

FR3. ***Data storage and processing layer:*** Many works emphasize this layer since it is the key to maintaining the records and files to be processed later using analysis tools. Some studies [81], [82] proposed using various techniques and technologies for data storage such as HDFS, HBASE, and HIVE. Furthermore, in [101], the authors stated that there are multiple methods to store and process data (for example, centralized/distributed or locally/remotely).

FR4. ***Integration layer:*** Certain studies implemented this layer since it is used for M2M communications, as in [102]; incidentally, other studies used this layer as an intermediary layer between data and devices or services, as in [80], [88] and [103].

FR5. ***Data analysis layer:*** A large number of works considered this layer in their studies since it is the critical layer for analyzing and extracting the most knowledge from the collected data. In [64], the authors propose using Apache Spark since it is a very versatile tool for analyzing data distributed in clusters; some applications used cloud capabilities to analyze huge datasets [103], [104].

FR6. ***Application and presentation layer:*** A considerable number of articles analyzed and implemented this layer. For example, in [102], the application layer hosts business processes, protocols, and programming logic, while the presentation layer houses human task mediations such as business service registration and defining the interaction between services.

TABLE 3. Summary of the functional requirements reported.

Documents	Ref.	FR1	FR2	FR3	FR4	FR5	FR6	FR7	FR8
K. Zhou et al., 2016	[43]	■	■	■	■	■	■	■	■
H. Daki et al., 2018	[81]	■			■	■			
X. He et al., 2018	[88]				■			■	
Y. Guo et al., 2018	[71]	■		■	■		■	■	
V. Marinakis et al., 2018	[64]	■		■	■		■	■	
S. Martin. et al., 2015	[84]	■	■					■	
D. Sembroiz et al., 2018	[89]	■	■			■		■	
Y. Simmhan et al., 2018	[90]	■	■	■			■	■	■
J. Chou et al., 2016	[91]	■					■	■	
H. Gökçe; K. Gökçe, 2013	[85]			■	■			■	
H. Gökçe; K. Gökçe, 2014	[92]	■						■	
H. Gökçe; K. Gökçe, 2014	[86]			■	■		■	■	
B. Cheng et al., 2015	[93]	■			■			■	
E. Gomes et al., 2016	[94]	■		■	■			■	
L. Linder et al., 2017	[95]	■			■			■	
M. Babar et al., 2017	[82]	■	■	■	■			■	■
C. Chilipirea et al., 2017	[96]	■		■	■	■	■	■	■
M. Malik et al., 2017	[97]	■			■	■			
I. Khajenasiri et al., 2017	[98]	■	■			■		■	
S. Lazarova-Molnar et al., 2017	[99]	■						■	
E. F. Zambom Santana et al., 2017	[83]	■	■	■	■	■	■	■	
L. Kallab et al., 2017	[100]	■	■		■	■	■	■	
E. H. A. Gomes et al., 2018	[65]	■		■	■		■	■	
A. M. Shahat Osman, 2019	[63]	■			■	■	■	■	
K. Vatanparvar et al., 2018	[101]	■			■	■	■		
H. Elhoseny et al., 2016	[102]	■				■		■	
N. Mohamed et al., 2018	[103]	■			■	■	■		
J. Al-Jaroodi et al., 2018	[104]	■			■	■	■		
M. Sinderen et al., 2010	[30]	■						■	■
R. Santos; P. Carreira, 2014	[105]	■						■	
P. Angulo et al., 2017	[31]	■	■					■	
V. Degelar et al., 2013	[21]	■			■			■	
	TOTAL	**29** ***(91%)***	**9** ***(28%)***	**11** ***(34%)***	**21** ***(66%)***	**13** ***(41%)***	**14** ***(44%)***	**26** ***(81%)***	**5** ***(16%)***

Legend: ■ "Reported" □ "Do not reported"

However, some studies used this layer to interact with users or automation applications through APIs or web applications [89], [93].

FR7. Decision maker layer: Some works considered this layer due to the importance of making decisions based on previously processed and analyzed data. Some develop applications so that the decisions are automated or manual (such as expert analysis) [43], [82], [90], [96].

According to the 32 works analyzed and summarized in Table 3, we observed that from 66% – 91% of the works implemented or analyzed the FRs' "data sources layer" (FR1), "data storage and processing layer" (FR4) and "application and presentation layer" (FR7). Therefore, these functional requirements exploit the functionalities for systems focused on the energy efficiency of buildings.

However, from 34% – 44% of the researchers expressed an interest in the functional requirements "data cleaning and preprocessing layer" (FR3), "integration layer" (FR5) and "data analysis layer" (FR6) since they represent important aspects that must be provided by a software architecture aimed at improving the energy efficiency of buildings.

Finally, the "networking layer" (FR2) and "decision-maker layer" (FR8) were considered by 16% – 28% of the works; these are the functional requirements least analyzed or implemented in the studies found.

In any case, the incorporation of all these functionalities (FR1 – FR8) in the architecture will be subject to the specific application on which it is focused.

2) NON-FUNCTIONAL REQUIREMENTS

Table 4 summarizes the articles retrieved with each of the NFRs reported in the analyzed studies. A software architecture aimed at the energy efficiency of buildings is related to some NFRs of large distributed systems (such as interoperability, scalability, and heterogeneity, among others) and

TABLE 4. Summary of the non-functional requirements reported.

Documents	Ref.	NFR1	NFR2	NFR3	NFR4	NFR5	NFR6
H. Daki et al., 2018	[81]	■	■				■
X. He et al., 2018	[88]	■	■				■
V. Marinakis et al., 2018	[64]	■		■			■
Y. Guo et al., 2018	[71]	■					■
S. Martin. et al., 2015	[84]			■			
M. A. Al Faruque et al., 2016	[107]	■		■		■	
D. Sembroiz et al., 2018	[89]	■		■	■	■	
Y. Simmhan et al., 2018	[90]	■		■		■	
J. Chou et al., 2016	[91]	■					
H. Gökçe; K. Gökçe, 2014	[92]			■			
H. Gökçe; K. Gökçe, 2014	[86]			■			
B. Cheng et al., 2015	[93]	■	■				
E. Gomes et al., 2016	[94]	■					
L. Linder et al., 2017	[95]	■	■	■		■	
M. Babar et al., 2017	[82]	■		■			
C. Chilipirea et al., 2017	[96]	■					■
L. Kallab et al., 2017	[100]	■	■	■			
M. Malik et al., 2017	[97]	■	■	■			
E. F. Santana et al., 2017	[83]	■		■		■	
E. H. A. Gomes et al., 2018	[65]						
A. M. Shahat Osman, 2019	[63]	■		■	■	■	
N. Mohamed et al., 2017	[108]			■	■	■	
K. Vatanparvar et al., 2018	[101]	■		■			
H. Elhoseny et al., 2016	[102]	■		■		■	
N. Mohamed et al., 2018	[103]	■		■		■	
J. Al-Jaroodi et al., 2018	[104]	■		■		■	
M. Sinderen et al., 2010	[30]			■			
R. Santos.; P. Carreira, 2014	[105]			■	■		
P. Angulo et al., 2017	[31]	■		■	■		■
	TOTAL	**22** ***(76%)***	**6** ***(21%)***	**21** ***(72%)***	**5** ***(17%)***	**9** ***(31%)***	**5** ***(17%)***

Legend: ■ "Reported" □ "Do not reported"

sensitive and personal data (such as security and privacy). Within this context, we retrieved 29 articles with

each of the non-functional requirements reported in the different studies, which we describe below:

NFR1. ***Scalability:*** A considerable number of studies considered this requirement due to the number of users and services that can access or use the infrastructure and the enormous amount of data it can store, which increases over time. In [93], the authors stated that a data repository can be expanded and supported by a cluster of CouchDB servers. In addition, the scalability of big data processing demands is provided by an HDFS [88].

NFR2. ***Processing:*** A few investigations mentioned this requirement. For example, in [80] and 87], the authors stated that technologies such as HBASE, which greatly improve data processing, were used since they offer real-time search mechanisms, caching, and server-side programming, among others.

NFR3. ***Interoperability:*** Many works analyzed this non-functional requirement since it is a crucial part of an architecture, giving systems the ability to exchange and share information regardless of the target language or system. Some of these studies adopted different techniques to conduct it, such as through web services [92]; however, in [101], the authors used fog computing to provide the necessary interoperability.

NFR4. ***Heterogeneity:*** A few studies analyzed this requirement. It applies to different areas, as stated in [89], where the authors presented this requirement as a challenge due to the various protocols and existing formats. However, in [63], the authors referred to the heterogeneity of the data source.

NFR5. ***Security and privacy:*** Although it is true that technology offers us great opportunities to process data, the data are at risk of being collected by agents that are external to the implemented architectures. Therefore, this requirement is of great importance since it protects sensitive and private information. For example, in [90] and [106], the authors mentioned some protocols that can be used for IoT architectures. In [107], the authors indicated that implementing a platform using fog computing provides the data privacy required for energy administration; in [63], the authors used Sentry, a service provided by Apache Hadoop, as a role-based authorization system for the data and metadata stored in clusters. In addition, SmartCityWare 108]; provides security mechanisms within its functions as a service.

NFR6. ***Fault tolerance:*** Fault tolerance is extremely important for a computer system since it allows a system to respond regardless of whether any of its components or services fail. For example, in [65], [80] and [87], the authors used a NoSQL database with fault tolerance support to correct faults. Similarly, in [96], the Hadoop framework was used due to the advantages it offers in terms of fault tolerance.

According to the 29 studies analyzed and summarized in Table 4, we observed that from 72% – 76% of the works implemented or analyzed the NFRs of "scalability" (NFR1) and "interoperability" (NFR3) for the most part. From this analysis, both non-functional requirements are relevant for an SOA focused on the energy efficiency of buildings since they would allow the management of energy services in the building. Especially, "interoperability" is a software feature of special interest for an SOA focused on the energy efficiency of buildings since it allows all interfaces to be fully understood and to work without restrictions with the subsystems, whether in implementation or access.

Furthermore, 31% of the studies considered "security and privacy" (NFR5) as an important non-functional requirement of the management system of the building components vulnerable to attacks. Therefore, it is important to develop smart strategies in this field.

Finally, the non-functional requirements "processing" (NFR2), "heterogeneity" (NFR4) and "fault tolerance" (NFR6) were considered by 17% – 21% of the studies. However, this does not mean that they are not relevant since in this review the use of these non-functional requirements in these works was still incipient.

E. RQ3. WHAT IMPLEMENTATIONS OF SERVICE-ORIENTED SOFTWARE ARCHITECTURES HAVE BEEN DEVELOPED FOR THE ENERGY EFFICIENCY OF BUILDINGS?

A service-oriented architecture (SOA) is defined as one approach, independent of the protocol of a computer, that allows the determination, recording and invoking of services [26], [27]. Therefore, an SOA can contribute greatly to the energy efficiency of buildings by favoring and supporting the characteristics of big data [35]. In addition, an SOA facilitates the exchange, communication and interaction between services and users, regardless of the protocols used by the building systems. Therefore, it enables other applications to have interoperability, flexibility and scalability [26], [27], [109]–[119].

To illustrate the potential of the combined use of an SOA and big data analysis in one system for the management of the energy efficiency of buildings, it is important for a system to analyze the data from the various devices and sensors and to make control decisions, including the establishment, change or reconfiguration of the operating parameters of energy systems, within the different sections of the building. The controllers act on functional rules based on the data collected and the data history. These are processed and organized by the various services, enabling the construction of a knowledge base used to develop maintenance plans, renew or improve equipment, etc. An SOA enables the continuous flow of data and information between applications or management services to optimize the energy efficiency.

Within this perspective, only 12 articles referred to the implementation of an SOA applied to building energy efficiency. Table 5 summarizes the corresponding implementations found in these studies.

TABLE 5. Summary of SOA implementation for building energy efficiency.

Documents	Ref.	Middleware-services	Web-services
V. Degeler et al., 2013	[21]		
M. van Sinderen et al. 2010	[30]		
P. Angulo et al. 2017	[31]		
H. U. Gökçe et al. 2014	[86]		
N. Mohamed et al. 2018	[103]		
H. Ufuk Gökçe et al. 2014	[92]		
I. Khajenasiri et al. 2017	[98]		
R. Santos et al. 2014	[105]		
M. A. Al Faruque et al. 2016	[107]		
N. Mohamed et al. 2017	[108]		
H. Zhao et al. 2015	[120]		
I. Petri, et al. 2015	[121]		
	TOTAL	**8 *(67%)***	**4 *(33%)***

Legend: "Reported" "Do not reported"

According to the 12 papers analyzed and summarized in Table 5, an SOA was implemented as a middleware service in 67% of the studies, while it was implemented as a web service in only 33% of the studies. However, these results indicate that SOA implementation in the studied context is in practice scarce since it is used only for the exchange of information of certain services, such as in [92], which developed a web service for the extraction of the data registered by sensors. However, there is a new paradigm called microservices, which presents various functionalities that would expand the functionalities of a software architecture aimed at the energy efficiency of buildings [36].

VI. DISCUSSION OF THE STATE-OF-THE-ART FINDINGS

This study has been conducted with the aim of identifying and analyzing the software requirements for a service-oriented architecture (SOA) that incorporates the software functionalities (interoperability and versatility) and big data characteristics that best represent the energy management of a building, thus contributing to its energy efficiency. The main findings obtained from the research questions posed are presented below.

A. FINDINGS RELATED TO RQ1

There are several significant findings applicable to RQ1. First, the results of this study indicate that the fundamental characteristics that best represent an architecture that enables the use of big data in the energy management of buildings are the velocity, volume, variety and value. These characteristics are of great importance and utility within this area since they are capable of supporting various processes, as indicated in [71]. Similarly, it is necessary to consider the problems presented by big data in this sector [23].

Furthermore, in the second finding, the building sector is too complex to analyze due to the large amount and types of data (stochastic or deterministic) it generates. Therefore, the characteristics of big data play an important role in the energy efficiency of buildings as they allow the application of analysis techniques to extract the value of these data. In addition, it makes it easier for building managers to make key decisions related to the energy efficiency of buildings [14], [23], [69], [122]–[125].

Regarding the third finding, another important characteristic of big data to consider within the energy management of buildings is the veracity [42], [44]. This characteristic is particularly relevant since buildings have many heterogeneous systems (data sources) that generate a large number of data types (structured data, unstructured data, real-time data, event data, etc.). In addition, they contain valuable information that is often wasted. Therefore, it is necessary to establish analysis tools or techniques that allow the extraction of significant information for better decision making.

Finally, we rely on these five big data characteristics (velocity, volume, variety, value and veracity) to define data management, but not data analysis, which will be addressed in future research.

B. FINDINGS RELATED TO RQ2

In this section, we discuss the findings most relevant to the requirements of software (functional and non-functional) found in the works cited.

1) FUNCTIONAL REQUIREMENTS

There are several interesting findings on this point. First, the works cited used functional requirements (FRs) according to the needs or context that were being addressed, such as in [82], [83], [90] and [96]. However, it is important to consider the possible bias in these results since these requirements were gathered according to the context in which the studies used them.

With regard to the second finding, it is important to consider the six FRs most cited in this study (FR1 and FR3–FR7). This finding reinforces our view on the requirements of software, specifically of the FRs, necessary for a software architecture that allows the integration of software services to support the energy efficiency of buildings.

2) NON-FUNCTIONAL REQUIREMENTS

There are several significant findings on this point. Regarding non-functional requirements (NFRs), a quite interesting finding is that of the six NFRs found in the work, more than 72% of the studies cited analyzed or implemented the NFRs for scalability and interoperability. These results can be explained by the fact that buildings contain a number of devices (sensors, actuators, etc.) and equipment (HVAC system, lighting, etc.), which in turn generate information. Therefore, they require software architectures that are scalable and interoperable [23], [95]. In summary, the two NFRs found are essential attributes for an architecture within the context studied. In other words, the NFRs make it possible to guarantee that the FRs are capable of processing or managing the user or service information.

In relation to the second finding, the analyzed studies coincide with previous research in highlighting the advantage of interoperability as a software functionality (feature) in service-oriented architectures. However, we do not clearly identify the versatility within these studies applied to the energy efficiency of buildings.

C. FINDINGS RELATED TO RQ3

There are several interesting findings applicable to RQ3. First, there is little evidence of the use of an SOA in the field of the energy efficiency of buildings. In fact, these studies used it as an information exchange layer to support some legacy services.

Another interesting finding was the identification of a second generation of SOA-based software architecture called the microservice architecture (MSA) [126]. This new software architectural style structures an application as a collection of small services [36]–[38]. Additionally, it has great advantages such as low coupling, scalability, being highly maintainable, being versatile and independently deploying services, among others [38]. Furthermore, we consider that an MSA can solve the versatility feature required in our study. Therefore, both generations (SOA and MSA) can provide great advantages, as described in [126].

D. THREATS TO VALIDITY

In this section, we discuss some of the threats to the validity of our study.

1) SEARCH STRATEGY

The most likely threat in this step is that some relevant documents may have been excluded. To reduce this threat, we used seven of the most important digital scientific databases for document retrieval. We adjusted each of the keywords in the different search engines. Similarly, we created different subgroups of these keywords to retrieve as many studies as possible. These keywords were corroborated among and approved by the authors.

2) SELECTION OF STUDIES

This threat to validity could be due to the authors' objective judgment in the selection of the studies. To mitigate this threat, we rigorously followed the selected review protocol. We validated each of the studies with the aforementioned selection criteria. Additionally, any study that raised questions regarding the criteria was discussed among the authors. Furthermore, we define the study period from 2014 to 2020 in criterion E1 since we want to know the new trends that have emerged on this matter. We are aware that there were proposals in previous years; however, we wanted to build on the new trends that exist for these technologies.

3) DATA EXTRACTION

Regarding data extraction, there may be biases in this process that affected the results and analysis of the selected works. To mitigate this, the different elements extracted in this SLR for each research question were discussed and verified among the authors, who reached an agreement on each of the elements obtained. Therefore, with these measures, the bias was reduced.

VII. RECOMMENDATIONS FOR FUTURE RESEARCH DIRECTIONS

In this section, we briefly discuss some recommendations for future research directions derived from the SLR.

A. DEMAND-SIDE MANAGEMENT

Demand-side management is a very important point to consider within the energy efficiency of buildings. Anticipating the demand peaks for energy using intelligent services (artificial intelligence, machine learning, etc.) would help improve the energy efficiency of a building and the economic costs, among other aspects. In addition to anticipating demand, the suppliers or producers of energy could reduce their power generation and, as a consequence, reduce their emissions [127]–[129].

Therefore, future research should focus on studying the necessary software requirements (FRs and NFRs) to establish a smarter software architecture for demand-side management. Additionally, it is necessary to develop techniques or tools that are capable of adapting to circumstances, such as the current circumstances caused by COVID-19, so that the system can be more efficient, faster and more reliable for the management of the demand side [130], [131].

B. NORMS, STANDARDS, DIRECTIVES AND INITIATIVES

Future studies on the different software service-oriented architectures that address the energy efficiency in buildings should analyze the technical algorithms that incorporate and identify the energy use of the equipment and optimize the energy of the building. These techniques are applied to energy performance models for buildings with different sizes, locations and sectors; they must contemplate the norms, standards, guidelines and initiatives proposed by various organizations worldwide [13], [132]–[135].

As an example, the UNE-EN 15251 standard proposes criteria for the energy management of buildings; therefore, it would be relevant to study how the different software service-oriented architectures identify the main foci of the consumption and energy generation in buildings, to study how parameters are monitored by analyzing the energy consumed based on the analysis of heterogeneous sources of data and to propose actions specific to minimizing the energy consumption. In addition, since the buildings have different energy use profiles, it is also necessary to use data and models that characterize the main contributors of energy consumption. For example, according to ASHRAE [136], in residential buildings, the energy consumed is primarily due to the equipment used by their occupants; in addition, in industrial buildings, the energy consumption is associated mostly with the operations of the machinery and industry infrastructure dedicated to production processes.

VIII. CONCLUSION AND FUTURE WORK

In this state-of-the-art study, we developed and executed an SLR following Kitchenham's methodological procedure to identify relevant works that analyze the software requirements for a service-oriented software architecture and the characteristics of big data that have been implemented in recent years to study energy-efficient buildings.

Once the SLR was conducted, we obtained a total of 97 relevant works that allowed us to generate favorable responses to our research questions.

We found that the characteristics of big data play essential roles in the development of a software architecture focused on the energy efficiency of buildings. Furthermore, the software requirements (FRs and NFRs) found, which to the best of our knowledge are classified for the first time herein, allowed us to provide a general idea of the necessary requirements for the development of software architectures in this area. In addition, it is important to highlight that these requirements must be considered in accordance with the needs to be solved and the area to which they are addressed. For example, the software requirements for a data center, factory, and office are not the same.

In addition, we detected that an SOA does comply with the functionality of interoperability software since several studies use it as an intermediary for the heterogeneous components or systems that buildings possess. Additionally, we did not find studies that consider the software functionality of versatility.

Considering this, we have identified a trend in the second generation of the service-oriented architecture (SOA), called the microservices architecture (MSA), which treats services in a distributed manner and translates into multiple benefits, as indicated in the work of [137]. Therefore, this MSA could support the software functionality of versatility within our study context.

TABLE 6. Summary of the findings in the literature.

Year	Authors	Approach	Application	Styles	Functional requirements	Non-Functional requirements	Ref.
2010	Sinderen, M. et al.	Architecture	Houshold Energy Management	SOA	✓ Context sensor services ✓ Control process for decisions ✓ Applications	✓ Interoperability	[30]
2013	H. Gökçe; K. Gökçe	Architecture	Building monitoring and control	Layered	✓ ETL layer ✓ Data warehouse core layer ✓ Information representation layer	-	[85]
2013	Degelar, V. et al.	Architecture	Smart Environments	SOA	✓ Physical layer ✓ Data repository ✓ Composition layer for interface	-	[21]
2014	H. Gökçe; K. Gökçe	Architecture	Energy Efficient Building	Layered	✓ Data Layer ✓ Information Layer ✓ Tool layer	✓ Interoperability ✓ Scalability	[92]
2014	H. Gökçe; K. Gökçe	Architecture	Building monitoring, analysis and optimization	Layered	✓ Data warehouse core ✓ ETL ✓ Intelligent control module ✓ Information representation	✓ Interoperability ✓ Monitoring	[86]
2014	Santos, R.; Carreira, P.,	Architecture	Building Energy Management Systems	SOA	✓ Devices layer ✓ Application layer	✓ Interoperability ✓ Heterogeneity handling	[105]
2015	B. Cheng et al.	Architecture	Smart City	Layered	✓ Data collection ✓ Data storage ✓ Data processing ✓ Applications	✓ Interoperability ✓ Scalability ✓ Processing	[93]
2015	H. Zhao et al.	Framework	Building design	SOA	-	-	[120]
2015	S. Martin. et al.	Architecture	Smart Grid	Layered	✓ Data layer ✓ Communication Logic Layer ✓ Application Layer	✓ Interoperability ✓ Security ✓ Privacy	[84]
2015	I. Petri et al.	Platform	Energy Efficient Building	SOA	-	-	[121]
2016	E. Gomes et al.	Model	Smart City	Layered	✓ Data sources ✓ ETL ✓ Data warehouse ✓ Data processing ✓ Users Visualization	✓ Scalability	[94]
2016	K. Zhou et al.	Model	Smart Energy	Architecture for Big Data	✓ Data collection, transmission and storage ✓ Data cleaning and processing ✓ Data integration and feature selection ✓ Data mining and knowledge Discovery ✓ Representation, visualization and application ✓ Intelligent decision-making and real-time interaction ✓ Smart energy management	-	[43]
2016	J. Chou et al.	Framework	Energy Efficient Building		-	✓ Scalability ✓ Security	[91]
2016	H. Elhoseny et al.	Architecture	Smart City	SOA and Layered	✓ Data layer ✓ Integration layer ✓ Application layer ✓ Presentation layer	✓ Interoperability ✓ Scalability	[102]
2016	M. A. Al Faruque et al.	Platform	Energy management	Cloud computing	-	✓ Interoperability ✓ Scalability ✓ Privacy	[107]

TABLE 6. *(Continued.)* Summary of the findings in the literature.

2017	Angulo, P. et al.	Architecture	Building monitoring	SOA	✓ Sensor layer ✓ Communication layer ✓ Application	✓ Interoperability	[31]
2017	L. Linder et al.	Architecture	Smart Building	Layered	✓ Data sources ✓ Data storage and processing ✓ Application layer	✓ Interoperability ✓ Privacy ✓ Scalability ✓ Processing	[95]
2017	M. Babar et al.	Architecture	Smart City	Layered	✓ Data acquistion and aggregation module ✓ Data computation and processing module ✓ Decision and application module	✓ Interoperability ✓ Scalability ✓ Fault tolerance	[82]
2017	C. Chilipirea et al.	Architecture	Smart City	Layered	✓ Data sources ✓ Data normalization ✓ Data brokering ✓ Data storage ✓ Data analysis ✓ Data visualization ✓ Decisions	✓ Scalability ✓ Fault tolerance	[96]
2017	M. Malik et al.	Proposal	Energy data analysis	Layered	✓ Data collection ✓ Data distribution ✓ Data storage	✓ Interoperability ✓ Scalability ✓ Processing ✓ Fault tolerance	[97]
2017	I. Khajenasiri et al.	Model	Energy control	Layered	✓ Sensing layer ✓ Network layer ✓ Service layer ✓ Interface layer	-	[98]
2017	S. Lazarova-Molnar et al.	Framework	Building data analysis	Layered	✓ Data tier ✓ Application tier ✓ Presentation tier	-	[99]
2017	E. F. Zambom Santana et al.	Proposal	Smart City	Layered	✓ Cloud & Networking ✓ Middleware ✓ User Management ✓ Social Gateway ✓ Big Data Management ✓ Application	✓ Interoperability ✓ Scalability ✓ Security ✓ Privacy	[83]
2017	L. Kallab et al.	Framework	Building energy management	Layered	✓ Field Level ✓ Core Platform Level ✓ Management Level ✓ End-User Level	✓ Interoperability ✓ Scalability ✓ Processing ✓ Heterogeneous	[100]
2017	N. Mohamed et al.	Platform	Smart City	SOA	-	✓ Interoperability ✓ Security	[108]
2018	H. Daki et al.	Architecture	Storage system electrical	Architecture for Big Data	✓ Data collection ✓ Data integration ✓ Data storage	✓ Scalability ✓ Processing ✓ Fault tolerance	[81]
2018	X. He et al.	Architecture	Smart City	Architecture for Big Data	✓ Data storage plane ✓ Data processing plane ✓ Data application plane	✓ Scalability ✓ Processing ✓ Fault tolerance	[88]
2018	N. Mohamed et al.	Architecture	Smart Building	Cloud computing	-	-	[138]
2018	Y. Zhang et al.	Framework	Energy data analysis	-	-	-	[87]
2018	Y. Guo et al.	Platform	Energy management	Architecture for Big Data	✓ Data sources ✓ Data interface ✓ Big data management ✓ Analytic engine ✓ Application and display	✓ Scalability ✓ Heterogeneous	[71]
2018	E. H. A. Gomes et al.	Model	Smart City	Layered	✓ Data sources ✓ ETL ✓ Data storage ✓ Data processing ✓ Search and Analysis	✓ Scalability	[65]
2018	V. Marinakis et al.	Architecture	Smart Energy	Architecture for Big Data	✓ Data interoperability and semantification layer ✓ Data storage cluster	✓ Interoperability ✓ Scalability ✓ Fault tolerance	[64]

TABLE 6. *(Continued.)* Summary of the findings in the literature.

					✓ Analytics services		
2018	D. Sembroiz et al.	Architecture	Smart Building	Cloud computing	✓ Perception layer ✓ Network layer (Data aggregator) ✓ Middleware ✓ Application layer	✓ Interoperability ✓ Scalability ✓ Privacy ✓ Security ✓ Heterogeneous	[89]
2018	N. Mohamed et al.	Architecture	Smart Building	SOA and Layered	✓ Control layer ✓ Configuration layer ✓ Planning layer	✓ Interoperability ✓ Security ✓ Scalability	[103]
2018	J. Al-Jaroodi et al.	Architecture	Smart City	SOA and Layered	✓ Control layer ✓ Configuration layer ✓ Planning layer	✓ Interoperability ✓ Security ✓ Scalability	[104]
2018	Y. Simmhan et al.	Architecture	Smart City	Cloud computing	✓ Sensing and actuation ✓ Networking ✓ Data acquisition and curation ✓ Data analytics and visualization ✓ Decision making	✓ Scalability ✓ Security ✓ Privacy ✓ Interoperability	[90]
2018	K. Vatanparvar et al.	Architecture	Energy management	SOA	✓ IoT ✓ Fog computing ✓ Cloud computing	✓ Interoperability ✓ Scalability	[101]
2019	A. M. Shahat Osman	Framework	Smart City	Layered	✓ Platform layer ✓ Data processing layer	✓ Scalabilty ✓ Heterogeneous ✓ Privacy ✓ Security ✓ Interoperability	[63]
2019	Fiware Foundation	Framework	Internet of Future	-	-	-	[106]

Finally, these findings suggest that, in general, both generations can coexist in a so-called intergenerational architecture (SOA and MSA), which is why this issue will be addressed in future studies.

APPENDIX

See Table 6.

ACKNOWLEDGMENT

The author would like to thank the National Secretariat for Science, Technology and Innovation (SENACYT, Panama) and the Institute for the Training and Use of Human Resources (IFARHU), for the doctoral study scholarship awarded within the IFARHU-SENACYT doctoral scholarship program, 2018. Also, the SNI-SENACYT National Research System of which one of the authors is a member and the Smart Tool for the Design of Residential Projects and Energetically Sustainable Buildings project, FONDO I + D + I SENACYT, 2018. In addition, they are grateful to the ''Smart Energy'' Campus of International Excellence, promoted by the Universidad de Alcalá and the Universidad Rey Juan Carlos, Spain, for the financial support received by granting the authors the ACCÉSIT of the 2019 Awards Call for the research project: ''Intelligent management system for optimizing energy consumption in building air conditioning.''

REFERENCES

[1] S. S. K. Kwok and E. W. M. Lee, ''A study of the importance of occupancy to building cooling load in prediction by intelligent approach,'' *Energy Convers. Manage.*, vol. 52, no. 7, pp. 2555–2564, Jul. 2011, doi: 10.1016/j.enconman.2011.02.002.

[2] H. Yoshino, T. Hong, and N. Nord, ''IEA EBC annex 53: Total energy use in buildings—Analysis and evaluation methods,'' *Energy Buildings*, vol. 152, pp. 124–136, Oct. 2017, doi: 10.1016/j.enbuild.2017.07.038.

[3] L. Belussi, B. Barozzi, A. Bellazzi, L. Danza, A. Devitofrancesco, C. Fanciulli, M. Ghellere, G. Guazzi, I. Meroni, F. Salamone, F. Scamoni, and C. Scrosati, ''A review of performance of zero energy buildings and energy efficiency solutions,'' *J. Building Eng.*, vol. 25, Sep. 2019, Art. no. 100772, doi: 10.1016/j.jobe.2019.100772.

[4] J. R. Vázquez-Canteli, S. Ulyanin, J. Kämpf, and Z. Nagy, ''Fusing tensorflow with building energy simulation for intelligent energy management in smart cities,'' *Sustain. Cities Soc.*, vol. 45, pp. 243–257, Feb. 2019, doi: 10.1016/j.scs.2018.11.021.

[5] European Commission. (May 16, 2019). *Energy Performance of Buildings Directive*. Accessed: Sep. 09, 2020. [Online]. Available: https://ec.europa.eu/energy/topics/energy-efficiency/energy-efficient-buildings/energy-performance-buildings-directive_en

[6] A. Magrini, G. Lentini, S. Cuman, A. Bodrato, and L. Marenco, ''From nearly zero energy buildings (NZEB) to positive energy buildings (PEB): The next challenge—The most recent European trends with some notes on the energy analysis of a forerunner PEB example,'' *Develop. Built Environ.*, vol. 3, Aug. 2020, Art. no. 100019, doi: 10.1016/j.dibe.2020.100019.

[7] M. Economidou, V. Todeschi, P. Bertoldi, D. D'Agostino, P. Zangheri, and L. Castellazzi, "Review of 50 years of EU energy efficiency policies for buildings," *Energy Buildings*, vol. 225, Oct. 2020, Art. no. 110322, doi: 10.1016/j.enbuild.2020.110322.

[8] S. Asadi, M. Nazari-Heris, S. R. Nasab, H. Torabi, and M. Sharifironizi, "An updated review on net-zero energy and water buildings: Design and operation," in *Food-Energy-Water Nexus Resilience and Sustainable Development: Decision-Making Methods, Planning, and Trade-Off Analysis*, S. Asadi and B. Mohammadi-Ivatloo, Eds. Cham, Switzerland: Springer, 2020, pp. 267–290.

[9] D. D'Agostino, P. Zangheri, and L. Castellazzi, "Towards nearly zero energy buildings in Europe: A focus on retrofit in non-residential buildings," *Energies*, vol. 10, no. 1, Jan. 2017, Art. no. 1, doi: 10.3390/en10010117.

[10] N. N. A. Bakar, M. Y. Hassan, H. Abdullah, H. A. Rahman, M. P. Abdullah, F. Hussin, and M. Bandi, "Energy efficiency index as an indicator for measuring building energy performance: A review," *Renew. Sustain. Energy Rev.*, vol. 44, pp. 1–11, Apr. 2015, doi: 10.1016/j.rser.2014.12.018.

[11] N. N. A. Bakar, M. Y. Hassan, H. Abdullah, H. A. Rahman, M. P. Abdullah, F. Hussin, and M. Bandi, "Identification building energy saving using energy efficiency index approach," in *Proc. IEEE Int. Conf. Power Energy (PECon)*, Dec. 2014, pp. 366–370, doi: 10.1109/PECON.2014.7062472.

[12] M. V. Moreno, L. Dufour, A. F. Skarmeta, A. J. Jara, D. Genoud, B. Ladevie, and J.-J. Bezian, "Big data: The key to energy efficiency in smart buildings," *Soft Comput.*, vol. 20, no. 5, pp. 1749–1762, May 2016, doi: 10.1007/s00500-015-1679-4.

[13] *Directive 2012/27/EU of the European Parliament and of the Council of 25 October 2012 on Energy Efficiency*, Eur. Parliament, Brussels, Belgium, 2012.

[14] A. Ridi, N. Zarkadis, G. Bovet, N. Morel, and J. Hennebert, "Towards reliable stochastic data-driven models applied to the energy saving in buildings," in *Proc. Int. Conf. Cleantech Smart Cities Buildings Nano Urban Scale*, Lausanne, Switzerland, Sep. 2013, pp. 1–7. Accessed: Sep. 9, 2020. [Online]. Available: https://hal.archives-ouvertes.fr/hal-00872270.

[15] C. Zhou, Z. Fang, X. Xu, X. Zhang, Y. Ding, X. Jiang, and Y. Ji, "Using long short-term memory networks to predict energy consumption of air-conditioning systems," *Sustain. Cities Soc.*, vol. 55, Apr. 2020, Art. no. 102000, doi: 10.1016/j.scs.2019.102000.

[16] J. Aguilar, A. Garcès-Jimènez, N. Gallego-Salvador, J. A. G. De Mesa, J. M. Gomez-Pulido, and À. J. Garcìa-Tejedor, "Autonomic management architecture for multi-HVAC systems in smart buildings," *IEEE Access*, vol. 7, pp. 123402–123415, 2019, doi: 10.1109/ACCESS.2019.2937639.

[17] L. Mendoza-Pittí, A. Garcès-Jimenez, J. Aguilar, J. M. Gómez-Pulido, and M. Vargas-Lombardo, "Proposal of physical models of multi-HVAC systems for energy efficiency in smart buildings," in *Proc. 7th Int. Eng., Sci. Technol. Conf. (IESTEC)*, Oct. 2019, pp. 641–646, doi: 10.1109/IESTEC46403.2019.00120.

[18] L. M. Escobar, J. Aguilar, A. Garcés-Jiménez, J. A. G. De Mesa, and J. M. Gomez-Pulido, "Advanced fuzzy-logic-based context-driven control for HVAC management systems in buildings," *IEEE Access*, vol. 8, pp. 16111–16126, 2020, doi: 10.1109/ACCESS.2020.2966545.

[19] H. V. V. Priyadarshana, W. K. I. Madushanaka, L. L. L. Anuruddha, G. T. Chathura, H. W. D. Hettiarachchi, and K. T. M. U. Hemapala, "Multi-agent controlled building management system," in *Proc. Innov. Power Adv. Comput. Technol. (i-PACT)*, Apr. 2017, pp. 1–5, doi: 10.1109/IPACT.2017.8244900.

[20] M. W. Ahmad, M. Mourshed, D. Mundow, M. Sisinni, and Y. Rezgui, "Building energy metering and environmental monitoring—A state-of-the-art review and directions for future research," *Energy Buildings*, vol. 120, pp. 85–102, May 2016, doi: 10.1016/j.enbuild.2016.03.059.

[21] V. Degeler, L. I. L. Gonzalez, M. Leva, P. Shrubsole, S. Bonomi, O. Amft, and A. Lazovik, "Service-oriented architecture for smart environments (short paper)," in *Proc. IEEE 6th Int. Conf. Service-Oriented Comput. Appl.*, Dec. 2013, pp. 99–104, doi: 10.1109/SOCA.2013.26.

[22] T. Teich, S. Wolf, T. Neumann, S. Junghans, and S. Franke, "Concept for a service-oriented architecture in building automation systems," *Procedia Eng.*, vol. 69, pp. 597–602, Jan. 2014, doi: 10.1016/j.proeng.2014.03.031.

[23] N. Koseleva and G. Ropaite, "Big data in building energy efficiency: Understanding of big data and main challenges," *Procedia Eng.*, vol. 172, pp. 544–549, Jan. 2017, doi: 10.1016/j.proeng.2017.02.064.

[24] *Proyecto CEI 'Energía Inteligente'*. Accessed: Sep. 09, 2020. [Online]. Available: https://www.campusenergiainteligente.es/en/presentacion/

[25] A. Chehri and H. T. Mouftah, "Service-oriented architecture for smart building energy management," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2013, pp. 4099–4103, doi: 10.1109/ICC.2013.6655203.

[26] M. P. Papazoglou and W.-J. van den Heuvel, "Service oriented architectures: Approaches, technologies and research issues," *VLDB J.*, vol. 16, no. 3, pp. 389–415, Jul. 2007, doi: 10.1007/s00778-007-0044-3.

[27] A. Bouguettaya *et al.*, "A service computing manifesto: The next 10 years," *Commun. ACM*, vol. 60, no. 4, Mar. 2017, doi: 10.1145/2983528.

[28] B. Kitchenham, "Procedures for performing systematic reviews," Dept. Comput. Sci., Keele Univ., Keele, U.K., Tech. Rep. TR/SE-0401, NICTA Tech. Rep. 0400011T.1, Aug. 2004, vol. 33.

[29] D. Mora, M. Taisch, A. W. Colombo, and J. M. Mendes, "Service-oriented architecture approach for industrial 'system of systems': State-of-the-art for energy management," in *Proc. IEEE 10th Int. Conf. Ind. Inform.*, Jul. 2012, pp. 1246–1251, doi: 10.1109/INDIN.2012.6301130.

[30] M. van Sinderen, M. Warnier, and B. Shishkov, "Service-oriented architecture for household energy management," in *Proc. 4th Int. Workshop Enterprise Syst. Technol.*, Athens, Greece, 2010, pp. 51–68, doi: 10.5220/0004465400510068.

[31] P. Angulo, C. C. Guzmán, G. Jiménez, and D. Romero, "A service-oriented architecture and its ICT-infrastructure to support eco-efficiency performance monitoring in manufacturing enterprises," *Int. J. Comput. Integr. Manuf.*, vol. 30, no. 1, pp. 202–214, Jan. 2017, doi: 10.1080/0951192X.2016.1145810.

[32] A. Malatras, A. H. Asgari, T. Baugé, and M. Irons, "A service-oriented architecture for building services integration," *J. Facilities Manage.*, vol. 6, no. 2, pp. 132–151, May 2008, doi: 10.1108/14725960810872659.

[33] A. Karlsen. (Jul. 18, 2019). *Flexibility—A Software Architecture Principle*. Accessed: Jan. 02, 2021. [Online]. Available: https://medium.com/faun/flexibility-a-software-architecture-principle-6eafe045a1d4

[34] H. Calderón-Gómez, L. Mendoza-Pitti, M. Vargas-Lombardo, J. M. Gomez-Pulido, J. L. Castillo-Sequera, J. Sanz-Moreno, and G. Sencion, "Telemonitoring system for infectious disease prediction in elderly people based on a novel microservice architecture," *IEEE Access*, vol. 8, pp. 118340–118354, 2020, doi: 10.1109/ACCESS.2020.3005638.

[35] L. Mendoza-Pittí, J. M. Gómez-Pulido, and M. Vargas-Lombardo, "Software architectures for energy efficiency of buildings: A systematic review," *Revista Iberica de Sistemas e Tecnologias de Informacao*, vol. 2019, no. E23, pp. 40–52, 2019.

[36] T. Cerny, M. J. Donahoo, and J. Pechanec, "Disambiguation and comparison of SOA, microservices and self-contained systems," in *Proc. Int. Conf. Res. Adapt. Convergent Syst. (RACS)*, Kraków, Poland, Sep. 2017, pp. 228–235, doi: 10.1145/3129676.3129682.

[37] F. Rademacher, S. Sachweh, and A. Zundorf, "Differences between model-driven development of service-oriented and microservice architecture," in *Proc. IEEE Int. Conf. Softw. Archit. Workshops (ICSAW)*, Apr. 2017, pp. 38–45, doi: 10.1109/ICSAW.2017.32.

[38] *What are Microservices?* Accessed: Sep. 28, 2019. [Online]. Available: http://microservices.io/index.html

[39] K. Krishnan, *3-D Data Management: Controlling Data Volume, Velocity, and Variety*. Amsterdam, The Netherlands: Morgan Kaufmann, 2013.

[40] M. Bilal, L. O. Oyedele, J. Qadir, K. Munir, S. O. Ajayi, O. O. Akinade, H. A. Owolabi, H. A. Alaka, and M. Pasha, "Big data in the construction industry: A review of present status, opportunities, and future trends," *Adv. Eng. Informat.*, vol. 30, no. 3, pp. 500–521, Aug. 2016, doi: 10.1016/j.aei.2016.07.001.

[41] A. Oussous, F.-Z. Benjelloun, A. A. Lahcen, and S. Belfkih, "Big data technologies: A survey," *J. King Saud Univ., Comput. Inf. Sci.*, vol. 30, no. 4, pp. 431–448, Oct. 2018, doi: 10.1016/j.jksuci.2017.06.001.

[42] A. Gandomi and M. Haider, "Beyond the hype: Big data concepts, methods, and analytics," *Int. J. Inf. Manage.*, vol. 35, no. 2, pp. 137–144, Apr. 2015, doi: 10.1016/j.ijinfomgt.2014.10.007.

[43] K. Zhou, C. Fu, and S. Yang, "Big data driven smart energy management: From big data to big insights," *Renew. Sustain. Energy Rev.*, vol. 56, pp. 215–225, Apr. 2016, doi: 10.1016/j.rser.2015.11.050.

[44] I. Lee, "Big data: Dimensions, evolution, impacts, and challenges," *Bus. Horizons*, vol. 60, no. 3, pp. 293–303, May 2017, doi: 10.1016/j.bushor.2017.01.004.

[45] B. Furht and F. Villanustre, "Introduction to big data," in *Big Data Technologies and Applications*. Cham, Switzerland: Springer, 2016, pp. 3–11.

[46] S. M. Borodo, S. M. Shamsuddin, and S. Hasan, ''Big data platforms and techniques,'' *Indonesian J. Electr. Eng. Comput. Sci.*, vol. 1, no. 1, p. 191, Jan. 2016, doi: 10.11591/ijeecs.v1.i1.pp191-200.

[47] A. Cravero, O. Saldana, R. Espinosa, and C. Antileo, "Big data architecture for water resources management: A systematic mapping study," *IEEE Latin Amer. Trans.*, vol. 16, no. 3, pp. 902–908, Mar. 2018, doi: 10.1109/TLA.2018.8358672.

[48] I. Taleb, M. A. Serhani, and R. Dssouli, "Big data quality: A survey," in *Proc. IEEE Int. Congr. Big Data (BigData Congress)*, San Francisco, CA, USA, Jul. 2018, pp. 166–173, doi: 10.1109/BigDataCongress.2018.00029.

[49] E. Y. T. Adesta and D. Agusman, ''Internet of Things and big data revolution in the context of green technology,'' Seminar Nasional Inovasi dan Aplikasi Teknologi di Industri, Indonesia, Tech. Rep. Prosiding SENIATI 2018, 2018, p. 9.

[50] M. A.-U.-D. Khan, M. F. Uddin, and N. Gupta, "Seven V's of big data understanding big data to extract value," in *Proc. Zone 1 Conf. Amer. Soc. Eng. Educ.*, Bridgeport, CT, USA, Apr. 2014, pp. 1–5, doi: 10.1109/ASEEZone1.2014.6820689.

[51] E. A. Nuaimi, H. A. Neyadi, N. Mohamed, and J. Al-Jaroodi, "Applications of big data to smart cities," *J. Internet Services Appl.*, vol. 6, no. 1, Aug. 2015, Art. no. 25, doi: 10.1186/s13174-015-0041-5.

[52] X. Jin, B. W. Wah, X. Cheng, and Y. Wang, ''Significance and challenges of big data research,'' *Big Data Res.*, vol. 2, no. 2, pp. 59–64, Jun. 2015, doi: 10.1016/j.bdr.2015.01.006.

[53] C. K. Emani, N. Cullot, and C. Nicolle, "Understandable big data: A survey," *Comput. Sci. Rev.*, vol. 17, pp. 70–81, Aug. 2015, doi: 10.1016/j.cosrev.2015.05.002.

[54] H. Daki, A. E. Hannani, A. Aqqal, A. Haidine, and A. Dahbi, ''Big data management in smart grid: Concepts, requirements and implementation,'' *J. Big Data*, vol. 4, no. 1, Dec. 2017, Art. no. 13, doi: 10.1186/s40537-017-0070-y.

[55] V. C. Storey and I.-Y. Song, "Big data technologies and management: What conceptual modeling can do," *Data Knowl. Eng.*, vol. 108, pp. 50–67, Mar. 2017, doi: 10.1016/j.datak.2017.01.001.

[56] J. I. Rodríguez-Molano, L. E. Contreras-Bravo, and E. R. López-Santana, "Big data tools for smart cities," in *Data Mining and Big Data*, vol. 10943, Y. Tan, Y. Shi, and Q. Tang, Eds. Cham, Switzerland: Springer, 2018, pp. 649–658.

[57] K. N. Singh, R. K. Behera, and J. K. Mantri, "Big data ecosystem: Review on architectural evolution," in *Emerging Technologies in Data Mining and Information Security*, vol. 813, A. Abraham, P. Dutta, J. K. Mandal, A. Bhattacharya, and S. Dutta, Eds. Singapore: Springer, 2019, pp. 335–345.

[58] C. Costa and M. Y. Santos, "Big data: State-of-the-art concepts, techniques, technologies, modeling approaches and research challenges," *Int. J. Comput. Sci.*, vol. 44, no. 3, pp. 285–301, 2017.

[59] N. E. Aboudi and L. Benhlima, ''Big data management for healthcare systems: Architecture, requirements, and implementation,'' *Adv. Bioinf.*, vol. 2018, pp. 1–10, Jun. 2018, doi: 10.1155/2018/4059018.

[60] M. Usama, M. Liu, and M. Chen, "Job schedulers for big data processing in Hadoop environment: Testing real-life schedulers using benchmark programs,'' *Digit. Commun. Netw.*, vol. 3, no. 4, pp. 260–273, Nov. 2017, doi: 10.1016/j.dcan.2017.07.008.

[61] F. Almeida, "Big data: Concept, potentialities and vulnerabilities," *Emerg. Sci. J.*, vol. 2, no. 1, pp. 1–10, Mar. 2018, doi: 10.28991/esj-2018-01123.

[62] A. F. C. Santos, I. P. Teles, O. M. P. Siqueira, and A. A. de Oliveira, "Big data: A systematic review," in *Information Technology—New Generations*, vol. 558, S. Latifi, Ed. Cham, Switzerland: Springer, 2018, pp. 501–506.

[63] A. M. S. Osman, "A novel big data analytics framework for smart cities," *Future Gener. Comput. Syst.*, vol. 91, pp. 620–633, Feb. 2019, doi: 10.1016/j.future.2018.06.046.

[64] V. Marinakis, H. Doukas, J. Tsapelas, S. Mouzakitis, Á. Sicilia, L. Madrazo, and S. Sgouridis, "From big data to smart energy services: An application for intelligent energy management," *Future Gener. Comput. Syst.*, vol. 110, pp. 572–586, Sep. 2020, doi: 10.1016/j.future.2018.04.062.

[65] E. H. A. Gomes and M. A. R. Dantas, "An infrastructure model for smart cities based on big data," *Int. J. Grid Utility Comput.*, vol. 9, no. 4, p. 11, 2018.

[66] P. V. Desai, "A survey on big data applications and challenges," in *Proc. 2nd Int. Conf. Inventive Commun. Comput. Technol. (ICICCT)*, Coimbatore, India, Apr. 2018, pp. 737–740, doi: 10.1109/ICICCT.2018.8472999.

[67] D. P. Acharjya and K. Ahmed, "A survey on big data analytics: Challenges, open research issues and tools," *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 2, pp. 511–518, 2016, doi: 10.14569/IJACSA.2016.070267.

[68] A. Fernández, S. del Río, V. López, A. Bawakid, M. J. del Jesus, J. M. Benítez, and F. Herrera, "Big data with cloud computing: An insight on the computing environment, mapreduce, and programming frameworks: Big data with cloud computing," *Wiley Interdiscipl. Rev., Data Mining Knowl. Discovery*, vol. 4, no. 5, pp. 380–409, Sep. 2014, doi: 10.1002/widm.1134.

[69] L. Wen, K. Zhou, S. Yang, and L. Li, "Compression of smart meter big data: A survey," *Renew. Sustain. Energy Rev.*, vol. 91, pp. 59–69, Aug. 2018, doi: 10.1016/j.rser.2018.03.088.

[70] K. Zhou and S. Yang, "Understanding household energy consumption behavior: The contribution of energy big data analytics," *Renew. Sustain. Energy Rev.*, vol. 56, pp. 810–819, Apr. 2016, doi: 10.1016/j.rser.2015.12.001.

[71] Y. Guo, Z. Yang, S. Feng, and J. Hu, ''Complex power system status monitoring and evaluation using big data platform and machine learning algorithms: A review and a case study," *Complexity*, vol. 2018, pp. 1–21, Sep. 2018, doi: 10.1155/2018/8496187.

[72] S. Pancholi, ''Solving big data challenges us electric utility industry,'' in *Proc. IEEE PES Gen. Meeting Presentation*, National Harbor, MD, USA, Jul. 2014, pp. 27–31.

[73] B. P. Bhattarai, S. Paudyal, Y. Luo, M. Mohanpurkar, K. Cheung, R. Tonkoski, R. Hovsapian, K. S. Myers, R. Zhang, P. Zhao, M. Manic, S. Zhang, and X. Zhang, ''Big data analytics in smart grids: State-of-the-art, challenges, opportunities, and future directions,'' *IET Smart Grid*, vol. 2, no. 2, pp. 141–154, Jun. 2019, doi: 10.1049/iet-stg.2018.0261.

[74] H. Jaakkola and B. Thalheim, "Architecture-driven modelling methodologies," in *Proc. Inf. Modelling Knowl. Bases*, 2011, p. 21.

[75] L. Dobrica and E. Niemela, "A survey on software architecture analysis methods," *IEEE Trans. Softw. Eng.*, vol. 28, no. 7, pp. 638–653, Jul. 2002, doi: 10.1109/TSE.2002.1019479.

[76] A. Sharma, M. Kumar, and S. Agarwal, "A complete survey on software architectural styles and patterns," *Procedia Comput. Sci.*, vol. 70, pp. 16–28, Jan. 2015, doi: 10.1016/j.procs.2015.10.019.

[77] M. Dabbagh and S. P. Lee, "An approach for integrating the prioritization of functional and nonfunctional requirements," *Sci. World J.*, vol. 2014, Apr. 2014, Art. no. e737626, doi: 10.1155/2014/737626.

[78] J. Eckhardt, A. Vogelsang, and D. M. Fernández, "Are 'non-functional' requirements really non-functional?: An investigation of non-functional requirements in practice," in *Proc. 38th Int. Conf. Softw. Eng. (ICSE)*, May 2016, pp. 832–842, doi: 10.1145/2884781.2884788.

[79] D. Ameller, X. Burgués, D. Costal, C. Farré, and X. Franch, "Non-functional requirements in model-driven development of service-oriented architectures," *Sci. Comput. Program.*, vol. 168, pp. 18–37, Dec. 2018, doi: 10.1016/j.scico.2018.08.001.

[80] I. Sommerville and P. Sawyer, *Requirements Engineering: A Good Practice Guide*, 1st ed. Hoboken, NJ, USA: Wiley, 1997.

[81] H. Daki, A. E. Hannani, and H. Ouahmane, "HBase-based storage system for electrical consumption forecasting in a Moroccan engineering school," in *Proc. 4th Int. Conf. Optim. Appl. (ICOA)*, Mohammedia, Morocco, Apr. 2018, pp. 1–6, doi: 10.1109/ICOA.2018.8370520.

[82] M. Babar and F. Arif, "Smart urban planning using big data analytics to contend with the interoperability in Internet of Things," *Future Gener. Comput. Syst.*, vol. 77, pp. 65–76, Dec. 2017, doi: 10.1016/j.future.2017.07.029.

[83] E. F. Z. Santana, A. P. Chaves, M. A. Gerosa, F. Kon, and D. S. Milojicic, "Software platforms for smart cities: Concepts, requirements, challenges, and a unified reference architecture," *ACM Comput. Surv.*, vol. 50, no. 6, pp. 1–37, Jan. 2018, doi: 10.1145/3124391.

[84] S. Martin, J. Hernandez, and C. Valmaseda, "A novel middleware for smart grid data exchange towards the energy efficiency in buildings," in *Proc. Int. Conf. Workshops Netw. Syst. (NetSys)*, Cottbus, Germany, Mar. 2015, pp. 1–8, doi: 10.1109/NetSys.2015.7089063.

[85] H. U. Gökçe and K. U. Gökçe, "Holistic system architecture for energy efficient building operation," *Sustain. Cities Soc.*, vol. 6, pp. 77–84, Feb. 2013, doi: 10.1016/j.scs.2012.07.003.

[86] H. U. Gökçe and K. U. Gökçe, "Multi dimensional energy monitoring, analysis and optimization system for energy efficient building operations," *Sustain. Cities Soc.*, vol. 10, pp. 161–173, Feb. 2014, doi: 10.1016/j.scs.2013.08.004.

[87] Y. Zhang, S. Ma, H. Yang, J. Lv, and Y. Liu, "A big data driven analytical framework for energy-intensive manufacturing industries," *J. Cleaner Prod.*, vol. 197, pp. 57–72, Oct. 2018, doi: 10.1016/j.jclepro.2018.06.170.

[88] X. He, K. Wang, H. Huang, and B. Liu, "QoE-driven big data architecture for smart city," *IEEE Commun. Mag.*, vol. 56, no. 2, pp. 88–93, Feb. 2018, doi: 10.1109/MCOM.2018.1700231.

[89] D. Sembroiz, S. Ricciardi, and D. Careglio, "A novel cloud-based IoT architecture for smart building automation," in *Security and Resilience in Intelligent Data-Centric Systems and Communication Networks*. Amsterdam, The Netherlands: Elsevier, 2018, pp. 215–233.

[90] Y. Simmhan, P. Ravindra, S. Chaturvedi, M. Hegde, and R. Ballamajalu, "Towards a data-driven IoT software architecture for smart city utilities," *Softw., Pract. Exp.*, vol. 48, no. 7, pp. 1390–1416, Jul. 2018, doi: 10.1002/spe.2580.

[91] J.-S. Chou, N.-T. Ngo, W. K. Chong, and G. E. Gibson, "Big data analytics and cloud computing for sustainable building energy efficiency," in *Start-Up Creation*. Amsterdam, The Netherlands: Elsevier, 2016, pp. 397–412.

[92] H. U. Gökçe and K. U. Gökçe, "Integrated system platform for energy efficient building operations," *J. Comput. Civil Eng.*, vol. 28, no. 6, Nov. 2014, Art. no. 05014005, doi: 10.1061/(ASCE)CP.1943-5487.0000288.

[93] B. Cheng, S. Longo, F. Cirillo, M. Bauer, and E. Kovacs, "Building a big data platform for smart cities: Experience and lessons from santander," in *Proc. IEEE Int. Congr. Big Data*, New York, NY, USA, Jun. 2015, pp. 592–599, doi: 10.1109/BigDataCongress.2015.91.

[94] E. Gomes, M. A. R. Dantas, D. D. J. de Macedo, C. De Rolt, M. L. Brocardo, and L. Foschini, "Towards an infrastructure to support big data for a smart city project," in *Proc. IEEE 25th Int. Conf. Enabling Technol., Infrastruct. Collaborative Enterprises (WETICE)*, Paris, France, Jun. 2016, pp. 107–112, doi: 10.1109/WETICE.2016.31.

[95] L. Linder, D. Vionnet, J.-P. Bacher, and J. Hennebert, "Big building data—A big data platform for smart buildings," *Energy Procedia*, vol. 122, pp. 589–594, Sep. 2017, doi: 10.1016/j.egypro.2017.07.354.

[96] C. Chilipirea, A.-C. Petre, L.-M. Groza, C. Dobre, and F. Pop, "An integrated architecture for future studies in data processing for smart cities," *Microprocessors Microsyst.*, vol. 52, pp. 335–342, Jul. 2017, doi: 10.1016/j.micpro.2017.03.004.

[97] M. Malik, M. Tabone, D. P. Chassin, E. C. Kara, R. V. Guha, and S. Kiliccote, "A common data architecture for energy data analytics," in *Proc. IEEE Int. Conf. Smart Grid Commun. (SmartGridComm)*, Dresden, Germany, Oct. 2017, pp. 417–422, doi: 10.1109/SmartGridComm.2017.8340736.

[98] I. Khajenasiri, A. Estebsari, M. Verhelst, and G. Gielen, "A review on Internet of Things solutions for intelligent energy control in buildings for smart city applications," *Energy Procedia*, vol. 111, pp. 770–779, Mar. 2017, doi: 10.1016/j.egypro.2017.03.239.

[99] S. Lazarova-Molnar and N. Mohamed, "Collaborative data analytics for smart buildings: Opportunities and models," *Cluster Comput.*, vol. 22, no. S1, pp. 1065–1077, Jan. 2019, doi: 10.1007/s10586-017-1362-x.

[100] L. Kallab, R. Chbeir, P. Bourreau, P. Brassier, and M. Mrissa, "HIT2GAP: Towards a better building energy management," *Energy Procedia*, vol. 122, pp. 895–900, Sep. 2017, doi: 10.1016/j.egypro.2017.07.399.

[101] K. Vatanparvar and M. A. A. Faruque, "Control-as-a-service in cyber-physical energy systems over fog computing," in *Fog Computing in the Internet of Things*, A. M. Rahmani, P. Liljeberg, J.-S. Preden, and A. Jantsch, Eds. Cham, Switzerland: Springer, 2018, pp. 123–144.

[102] H. Elhoseny, M. Elhoseny, S. Abdelrazek, H. Bakry, and A. Riad, "Utilizing service oriented architecture (SOA) in smart cities," *Int. J. Adv. Comput. Technol.*, vol. 8, no. 3, pp. 77–84, 2017.

[103] N. Mohamed, J. Al-Jaroodi, and I. Jawhar, "Service-oriented big data analytics for improving buildings energy management in smart cities," in *Proc. 14th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Limassol, Cyprus, Jun. 2018, pp. 1243–1248, doi: 10.1109/IWCMC.2018.8450469.

[104] J. Al-Jaroodi and N. Mohamed, "Service-oriented architecture for big data analytics in smart cities," in *Proc. 18th IEEE/ACM Int. Symp. Cluster, Cloud Grid Comput. (CCGRID)*, Washington, DC, USA, May 2018, pp. 633–640, doi: 10.1109/CCGRID.2018.00052.

[105] R. Santos and P. Carreira, "Service oriented development of building energy management systems—An architecture blueprint," in *Proc. 3rd Int. Conf. Smart Grids Green IT Syst.*, Barcelona, Spain, 2014, pp. 315–323, doi: 10.5220/0004974903150323.

[106] FIWARE. *What is Fiware?* Accessed: May 02, 2019. [Online]. Available: https://www.fiware.org/developers/

[107] M. A. A. Faruque and K. Vatanparvar, "Energy management-as-a-service over fog computing platform," *IEEE Internet Things J.*, vol. 3, no. 2, pp. 161–169, Apr. 2016, doi: 10.1109/JIOT.2015.2471260.

[108] N. Mohamed, J. Al-Jaroodi, I. Jawhar, S. Lazarova-Molnar, and S. Mahmoud, "SmartCityWare: A service-oriented middleware for cloud and fog enabled smart city services," *IEEE Access*, vol. 5, pp. 17576–17588, 2017, doi: 10.1109/ACCESS.2017.2731382.

[109] J. Minguez, M. Jakob, U. Heinkel, and B. Mitschang, "A SOA-based approach for the integration of a data propagation system," in *Proc. IEEE Int. Conf. Inf. Reuse Integr.*, Las Vegas, NV, USA, Aug. 2009, pp. 47–52, doi: 10.1109/IRI.2009.5211609.

[110] N. P. Kotsiopoulos and D. Vrakas, "Policies production system for ambient intelligence environments," in *Artificial Intelligence: Methods and Applications*, vol. 8445, A. Likas, K. Blekas, and D. Kalles, Eds. Cham, Switzerland: Springer, 2014, pp. 251–263.

[111] F. Fayaza, "Service oriented architecture in enterprise application," Dept. Inf. Technol., Univ. Moratuwa, Katubedda, Sri Lanka, Tech. Rep., 2014, doi: 10.13140/RG.2.1.3597.8087.

[112] C. Kyriazopoulou, "Architectures and requirements for the development of smart cities: A literature study," in *Smart Cities, Green Technologies, and Intelligent Transport Systems*, vol. 579, M. Helfert, K.-H. Krempels, C. Klein, B. Donellan, and O. Guiskhin, Eds. Cham, Switzerland: Springer, 2015, pp. 75–103.

[113] A. Dutta, M. S. Devi, and M. Arora, "Online census-based information sharing for delivery of e-Governance services," in *Proc. Special Collection eGovernment Innov. India (ICEGOV)*, New Delhi, India, 2017, pp. 45–51, doi: 10.1145/3055219.3055226.

[114] A. Dutta, M. S. Devi, and M. Arora, "Census Web service architecture for e-Governance applications," in *Proc. 10th Int. Conf. Theory Pract. Electron. Governance (ICEGOV)*, New Delhi, India, Mar. 2017, pp. 1–4, doi: 10.1145/3047273.3047390.

[115] T. H. Ashrafi, S. E. Arefin, K. D. J. Das, M. A. Hossain, and A. Chakrabarty, "FOG based distributed IoT infrastructure," in *Proc. 2nd Int. Conf. Internet Things, Data Cloud Comput. (ICC)*, Cambridge, U.K., Mar. 2017, pp. 1–13, doi: 10.1145/3018896.3036365.

[116] J. A. Oliveira, M. Vargas, and R. Rodrigues, "SOA reuse: Systematic literature review updating and research directions," in *Proc. XIV Brazilian Symp. Inf. Syst. (SBSI)*, Caxias do Sul, Brazil, 2018, pp. 1–8, doi: 10.1145/3229345.3229419.

[117] E. G. Nadhan. (Apr. 2004). *Service-Oriented Architecture: Implementation Challenges*. Accessed: Apr. 12, 2019. [Online]. Available: https://docs.microsoft.com/en-us/previous-versions/aa480029(v%3dmsdn.10)

[118] G. Agugiaro, J.-L. Robineau, and P. Rodrigues, "Project Ci-nergy: Towards AN integrated energy urban planning system from a data modelling and system architecture perspective," *ISPRS Ann. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. IV–4/W3, pp. 5–12, Sep. 2017, doi: 10.5194/isprs-annals-IV-4-W3-5-2017.

[119] N. M. Josuttis, *SOA in Practice: The Art of Distributed System Design*. Newton, MA, USA: O'Reilly Media, 2007.

[120] H. Zhao, Z. Nagy, D. Thomas, and A. Schlueter, "Service-oriented architecture for data exchange between a building information model and a building energy model," in *Proc. Int. Conf. CISBAT Future Buildings Districts Sustain. Nano Urban Scale*, 2015, pp. 761–766.

[121] I. Petri, Y. Rezgui, T. Beach, H. Li, M. Arnesano, and G. M. Revel, "A semantic service-oriented platform for energy efficient buildings," *Clean Technol. Environ. Policy*, vol. 17, no. 3, pp. 721–734, Mar. 2015, doi: 10.1007/s10098-014-0828-2.

[122] I. A. T. Hashem, V. Chang, N. B. Anuar, K. Adewole, I. Yaqoob, A. Gani, E. Ahmed, and H. Chiroma, "The role of big data in smart city," *Int. J. Inf. Manage.*, vol. 36, no. 5, pp. 748–758, Oct. 2016, doi: 10.1016/j.ijinfomgt.2016.05.002.

[123] H. Ahmed. (Aug. 2014). *Applying Big Data Analytics for Energy Efficiency*. Accessed: Sep. 09, 2020. [Online]. Available: https://aaltodoc.aalto.fi:443/handle/123456789/13899

[124] V. Marinakis, "Big data for energy management and energy-efficient buildings," *Energies*, vol. 13, no. 7, Mar. 2020, Art. no. 7, doi: 10.3390/en13071555.

[125] A. Ashabi, S. B. Sahibuddin, and M. S. Haghighi, "Big data: Current challenges and future scope," in *Proc. IEEE 10th Symp. Comput. Appl. Ind. Electron. (ISCAIE)*, Apr. 2020, pp. 131–134, doi: 10.1109/ISCAIE47305.2020.9108826.

[126] A. Andriyanto, R. Doss, and P. Yustianto, "Adopting SOA and microservices for inter-enterprise architecture in SME communities," in *Proc. Int. Conf. Electr., Electron. Inf. Eng. (ICEEIE)*, vol. 6, Oct. 2019, pp. 282–287, doi: 10.1109/ICEEIE47180.2019.8981437.

[127] P. Kumar, G. S. Brar, and L. Singh, "Energy efficiency evaluation in commercial and residential buildings with demand side management: A review," in *Proc. 8th Int. Conf. Power Syst. (ICPS)*, Dec. 2019, pp. 1–6, doi: 10.1109/ICPS48983.2019.9067675.

[128] A. Ozadowicz and J. Grela, "An event-driven building energy management system enabling active demand side management," in *Proc. 2nd Int. Conf. Event-Based Control, Commun., Signal Process. (EBCCSP)*, Jun. 2016, pp. 1–8, doi: 10.1109/EBCCSP.2016.7605082.

[129] Y. Li, J. De La Ree, and Y. Gong, "The smart thermostat of HVAC systems based on PMV-PPD model for energy efficiency and demand response," in *Proc. 2nd IEEE Conf. Energy Internet Energy Syst. Integr. (EI2)*, Oct. 2018, pp. 1–6, doi: 10.1109/EI2.2018.8582341.

[130] C. Chase, "Rapid demand response forecasting techniques are helping companies adapt during covid-19," *J. Bus. Forecasting*, vol. 39, no. 2, pp. 21–27, Jun. 2020.

[131] IEA. (Jun. 2020). *Energy Integration—The Covid-19 Crisis and Clean Energy Progress—Analysis*. Accessed: Aug. 27, 2020. [Online]. Available: https://www.iea.org/reports/the-covid-19-crisis-and-clean-energy-progress/energy-integration

[132] T.-Y. Chiu, S.-L. Lo, and Y.-Y. Tsai, "Establishing an integration-energy-practice model for improving energy performance indicators in ISO 50001 energy management systems," *Energies*, vol. 5, no. 12, Dec. 2012, Art. no. 12, doi: 10.3390/en5125324.

[133] *The New EN ISO 52000 Family of Standards to Assess the Energy Performance of Buildings Put in Practice—NASA/ADS*. Accessed: Aug. 27, 2020. [Online]. Available: https://ui.adsabs.harvard.edu/abs/2019E3SWC.11104047V/abstract

[134] R. J. Cole and M. J. Valdebenito, "The importation of building environmental certification systems: International usages of BREEAM and LEED," *Building Res. Inf.*, vol. 41, no. 6, pp. 662–676, Nov. 2013, doi: 10.1080/09613218.2013.802115.

[135] C. Eu. *Directive (EU) 2018/2002 of the European Parliament and of the Council—Of 11 December 2018—Amending Directive 2012/27/EU on Energy Efficiency*. European Parliament. Accessed: Aug. 27, 2020. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32018L2002&from=EN

[136] L. Berglund, "Mathematical models for predicting thermal comfort response of building occupants," *Ashrae J.-Amer. Soc. Heating Refrigerating Air-Conditioning Eng.*, vol. 19, no. 12, p. 38, 1977.

[137] D. Guaman, L. Yaguachi, C. C. Samanta, J. H. Danilo, and F. Soto, "Performance evaluation in the migration process from a monolithic application to microservices," in *Proc. 13th Iberian Conf. Inf. Syst. Technol. (CISTI)*, Jun. 2018, pp. 1–8, doi: 10.23919/CISTI.2018.8399148.

[138] N. Mohamed, J. Al-Jaroodi, and S. Lazarova-Molnar, "Energy cloud: Services for smart buildings," in *Sustainable Cloud and Energy Services*, W. Rivera, Ed. Cham, Switzerland: Springer, 2018, pp. 117–134.

LUIS MENDOZA-PITTÍ was born in Panama, in 1992. He received the B.E. degree in engineering systems and computing from the Technological University of Panama, Panama, in 2015. He is currently pursuing the Ph.D. degree in information and knowledge engineering with the University of Alcalá, Alcalá de Henares, Spain. From 2013 to 2016, he was a Student Research Assistant with the Technological University of Panama. His current research interests include smart building, energy optimization, cloud computing, and software engineering.

HURIVIADES CALDERÓN-GÓMEZ was born in Panama, in 1993. He received the B.E. degree in software development from the Technological University of Panama, Panama, in 2016, and the M.E. degree in higher education from University Santander, Panama, Panama, in 2019. He is currently pursuing the Ph.D. degree in information and knowledge engineering with the University of Alcalá, Alcalá de Henares, Spain.

From 2014 to 2016, he was a Research Assistant with the Technological University of Panama. In 2017, he joined the Technological University of Panama, as a Research. Since March 2019, he has been an Associate Professor with the Department of Computer Systems Engineering, Technological University of Panama. His current research interests include eHealth, cloud computing, and software engineering.

MIGUEL VARGAS-LOMBARDO (Senior Member, IEEE) received the Ph.D. degree from the Polytechnic University of Madrid. He is currently a member of the National Research System (SNI-SENACYT), Panama. He is also a Professor and a Researcher with the Technological University of Panama.

JOSÉ MANUEL GÓMEZ-PULIDO was born in Cáceres, Spain. He received the B.S. and M.S. degrees in telecommunications engineering from the University Polytechnic of Madrid, Spain, in 1988 and 1994, respectively, and the Ph.D. degree in telecommunication from the University of Alcala, Madrid, in 2004. In 1992, he was an Assistant Researcher and a Professor with the University Polytechnic of Madrid. From 1993 to 2004, he was with the Signal Theory and Communications Department, University of Alcala, where he is currently an Assistant Professor with the Computational Science Department. He has participated in more than 24 research projects several research projects with Spanish and European companies, related with radio propagation in mobile communications, signal processing in radar, ultrasound and laser applications, EMI–EMC design in on-board satellite instrumentation, radio propagation materials proprieties, numerical techniques applied in telecommunication, and so on. His research interests include wireless sensors networks, smart grids, and new technologies in the data mining and artificial intelligence world applied to the ICT.

JOSÉ LUIS CASTILLO-SEQUERA received the degree in computing from the National University of San Marcos, Peru, the master's degree in computer project management and the master's degree in university teaching, Spain, and the Ph.D. degree in information systems, documentation, and knowledge from the University of Alcala, Spain. He is currently a Full Professor with the Department of Computer Science, University of Alcala. He also leads projects of technological innovation oriented to the use of ICTs in university research groups. Their areas of interests focus on business technologies and intelligent technologies for knowledge management with interdisciplinary technologies. He has multiple publications at the level of high-level indexed JCR journals. He is the author of books and book chapters related to the field of artificial intelligence. His research interests include information retrieval, knowledge extraction based on evolutionary learning and neural networks, data mining and big data, documentation, learning, and teaching innovation with Web 2.0.

2.2 Publicación 2: Proposal of Physical Models of Multi-HVAC Systems for Energy Efficiency in Smart Buildings

2.2.1 Resumen de la publicación

Este artículo propone la simulación del comportamiento y las operaciones de diferentes sistemas de calefacción, ventilación y aire acondicionado (HVAC) en un edificio concreto mediante el uso de modelos físicos. Esto se logró al desarrollar de sistemas de control efectivos con optimización multiobjetivo para mejorar la eficiencia energética y la reducción de las emisiones de gases de efecto invernadero (GEI) en las ciudades. Por lo que, se logró aproximaciones muy cercanas a las que incorporan las especificaciones funcionales proporcionadas por los fabricantes de los sistemas HVAC. En este sentido, el modelado adecuado de estos sistemas permitiría desarrollar algoritmos de Soft-Computing que mejoren la eficiencia energética de los edificios.

2.2.2 Contribuciones significativas de la publicación

En esta segunda publicación, el doctorando realizó diferentes contribuciones significativas a la comunidad científica: (i) se propone el uso de modelos físicos para simular el comportamiento y funcionamiento de los diferentes subsistemas HVAC de un edificio en concreto; (ii) se logró describir cuatro modelos físicos simplificados de los sistemas multi-HVAC del edificio, en la cual pueden ser utilizados dentro de algoritmos de optimización multiobjetivo, a fin de mejorar y optimizar el desempeño físico de estos sistemas del edificio; (iii) los modelos físicos simplificados empleados en los sistemas multi-HVAC permitirían desarrollar algoritmos de Soft-Computing, como sistemas de control efectivos con optimización multiobjetivo, que mejoren la eficiencia energética de los edificios.

2.2.3 Índices de calidad de la publicación

Autores: L. Mendoza-Pittí, A. Garcés-Jiménez, J. Aguilar, J.M. Gómez-Pulido, M. Vargas-Lombardo.

Título: Proposal of physical models of multi-HVAC systems for energy efficiency in smart buildings

Conference Proceedings: 2019 7th International Engineering, Sciences and Technology Conference (IESTEC)

Fecha de publicación: 09 de octubre del 2019

Ubicación: Ciudad de Panamá, Panamá

Páginas: 641-646

DOI: 10.1109/IESTEC46403.2019.00120

Editor: IEEE

Date Added to IEEE Xplore: 30 December 2019

Electronic ISBN: 978-1-7281-1691-4

Print on Demand (PoD) ISBN: 978-1-7281-1692-1

INSPEC Accession Number: 19244670

2.2.4 Publicación

Proposal of physical models of multi-HVAC systems for energy efficiency in smart buildings

Luis Mendoza-Pittí
Depto. Ciencias de la Computación
Universidad de Alcalá
Alcalá de Henares, España
l.mendoza@edu.uah.es

Alberto Garcés-Jiménez
Centro de Innovación Experimental del Conocimiento (CEIEC)
Universidad Francisco de Victoria
Pozuelo, España
albertogarces0@gmail.com

José Aguilar
CEMISID, Escuela de Ingeniería de Sistemas
Universidad de Los Andes
Mérida, Venezuela
aguilar@ula.ve

José Manuel Gómez-Pulido
Dpto. Ciencias de la Computación
Universidad de Alcalá
Alcalá de Henares, España
jose.gomez@uah.es

Miguel Vargas-Lombardo
E-health and Supercomputing Research Group (GISES)
Universidad Tecnológica de Panamá
Panamá City, Panamá
miguel.vargas@utp.ac.pa

***Abstract*— One of main contributions to the consumption of energy in buildings is the heating, ventilation and air condition (HVAC) system. In fact, it consumes more than half of the energy. Multi-HVAC systems can be seen as an intermediate level of modeling, composed by different HVAC subsystems (e.g. boilers, chillers, heat pumps, among others), that define their operations and enabling the simulation of their behavior at a given time. The main purpose is to develop effective control systems with multi-objective optimization to improve the energy efficiency and the reduction of greenhouse gas (GHG) emissions in the cities. This article proposes the simulation of the behavior and operations of different HVAC subsystems in a particular building making use of physical models. The objective is achieved with different approaches, taking the functional specifications provided by the vendors of the HVAC subsystems.**

Keywords— multi-HVAC systems, energy efficiency, physical model, smart building, optimization multi-objectives.

***Resumen*— Uno de los principales contribuidores al consumo de energía en los edificios es el sistema de calefacción, ventilación y aire acondicionado (HVAC, por sus siglas en inglés). De hecho, estos consumen más de la mitad de la energía. Los sistemas multi-HVAC pueden ser vistos como un nivel intermedio de modelado, compuesto por diferentes subsistemas HVAC (como, por ejemplo, calderas, enfriadoras, bombas de calor, entre otros), que definen sus operaciones y permiten la simulación de su comportamiento en un momento dado. El objetivo principal es desarrollar sistemas de control efectivos con optimización multi-objetivo, para mejorar la eficiencia energética y la reducción de las emisiones de gases de efecto invernadero (GEI) en las ciudades. Este artículo propone la simulación del comportamiento y operaciones de diferentes subsistemas HVAC en un edificio particular, empleando modelos físicos. El objetivo se logra con diferentes enfoques, de acuerdo con las especificaciones funcionales proporcionadas por los proveedores de los subsistemas de HVAC.**

Palabras Claves— Sistemas multi-HVAC, eficiencia energética, modelo físico, edificio inteligente, optimización multi-objetivos.

I. INTRODUCCIÓN

El crecimiento demográfico va en aumento cada año, obligando a que más personas migren hacia las ciudades. Según cifras de las Naciones Unidas, para el año 2014 la población urbana era de 3900 millones, y se prevé que para el 2050 ascienda a 6300 millones [1]. Esta situación plantea grandes desafíos para las ciudades actuales, debido a que requieren brindar servicios que satisfagan las necesidades de los ciudadanos sin comprometer el medio ambiente.

En este sentido, surgen las Smart Cities, que se apoyan de las Tecnologías de la Información y Comunicación (TICs) para responder a la necesidad de su sostenibilidad y mejorar la calidad de vida de los habitantes y a los requerimientos de sus organismos y empresas. Es por ello, que las Smart Cities engloban distintos ámbitos, como, por ejemplo: Smart Health, Smart Transportation, Smart Grid, Smart Building, entre otros [2], [3].

Dentro de este ámbito, los Smart Building requieren controlar y administrar los diferentes componentes de las edificaciones, a fin de mejorar la eficiencia, garantizar la accesibilidad, aumentar la seguridad e incrementar el ahorro energético. Más de la mitad del consumo energético de los edificios deriva de los sistemas de calefacción, ventilación y aire acondicionado (HVAC, por sus siglas en inglés) [4], donde el 30 % - 40 % de este consumo es causado por las enfriadoras [5]. Por este motivo, la climatización es un factor importante para la eficiencia energética y el ahorro en las políticas energéticas de los países [6]. Los sistemas HVAC están compuestos por diferentes subsistemas que refrigeran o calefactan los edificios dependiendo de la región y la época del año, con la finalidad de

978-1-7281-1691-4/19/$31.00 ©2019 IEEE
DOI 10.1109/IESTEC46403.2019.00120

ofrecer condiciones adecuadas de temperatura y confort a los ocupantes del edificio.

La complejidad del manejo de los elementos de climatización exige sistemas de gestión energética de edificios (BEMS, por sus siglas en inglés), que recopilan datos de diferentes sensores y actúan sobre dispositivos de control, sistemas lógicos, actuadores, entre otros [7]. Estos sistemas permiten configurar y ajustar las temperaturas consignas de los HVAC para las épocas de invierno y verano, sin embargo, no son capaces de utilizar datos históricos o externos (como, por ejemplo, datos meteorológicos) para regular de manera óptima las temperaturas que deben ofrecer en un momento dado estos sistemas.

Un considerable número de investigadores ha empleado la optimización multiobjetivo (OM) en los sistemas HVAC con enfriadoras, logrando reducir el consumo energético de estos subsistemas HVAC, sin comprometer la productividad y el confort de los ocupantes de las edificios [5], [8], [9]. Sin embargo, hay una menor investigación de los sistemas descomponiéndolos a nivel de máquina y sus elementos asociados, que denominaremos subsistema multi-HVAC.

Por las razones enunciadas anteriormente, este estudio tiene el objetivo de proponer mediante modelos físicos simplificados el comportamiento y funcionamiento de los diferentes subsistemas HVAC presentes en el Edificio del Teatro Real de Madrid, en España, para que al momento de buscar puntos de optimización que mejoren de manera significativa la eficiencia energética de los sistemas reales, estos sean representativos de un caso real. En este sentido, este trabajo es una primera fase en el cual está mejorándose para emplear la optimización en mención.

Este documento ha sido organizado de la siguiente manera. Sección II, se presenta el marco teórico de los sistemas HVAC y BEMS. Sección III, se presentan algunos estudios relacionados al contexto de la investigación. Sección IV, se detalla y se describen los modelos propuestos para los sistemas HVAC. Sección V, presentamos algunos resultados obtenidos tras ejecutar los modelos propuestos y su discusión. Las conclusiones se detallan en la Sección VI.

II. Marco teórico

A. Sistemas HVAC

Los sistemas HVAC son parte esencial en las edificaciones, ya que se encargan de ofrecer una temperatura y calidad de aire interior aceptable en edificios como oficinas, hospitales, centros comerciales, entre otros. Asimismo, son los principales consumidores de energía de los edificios [4], [10]. Por lo tanto, aplicar una adecuada gestión energética en estos sistemas, sin comprometer el confort interior, contribuye en gran medida a la eficiencia energética en las edificaciones [7].

Los sistemas HVAC están compuestos por una gran variedad de subsistemas (por ejemplo, calderas, bombas de calor, enfriadoras, tuberías, fan-coils, unidades de tratamiento de aire, entre otros). Además, cuenta con diversos sensores, controladores y actuadores que permiten regular la temperatura de la zona, la temperatura de agua enfriada, la presión estática de las tuberías, la temperatura de aire suministrada, la situación de las válvulas de paso, etc. En consecuencia, los sistemas HVAC son sistemas no lineales muy complejos, ya que están compuesto por diferentes subsistemas, por lo que la representación matemática desde la perspectiva de optimización de estos sistemas está considerada como de "problema complejo" [11], siendo el desarrollo de un modelo preciso y efectivo para los sistemas HVAC un gran desafío. Según los autores en [12], señalan que existen tres tipos de enfoque de modelado para los sistemas HVAC:

1) Enfoque basado en el comportamiento físico (caja blanca)

Este enfoque se apoya de las leyes de la física, tales como transferencia térmica, balance de flujo y energía, entre otros, con el objeto de exponer un conjunto de ecuaciones que describan los sistemas HVAC.

2) Enfoque basado en datos (caja negra)

Este enfoque emplea los datos recopilados del sistema HVAC (datos normales o anormales) para definir una relación entre los parámetros de entrada y salida, mediante técnicas como regresión lineal o algoritmos heurísticos.

3) Enfoque híbrido (caja gris)

Este enfoque se basa en una combinación de los enfoques anteriores, donde se define un modelo básico empleando métodos basados en la física y los parámetros del modelo se determinan mediante algoritmos de estimación de parámetros de acuerdo a los datos recopilados del sistema HVAC.

Sobre las bases de las ideas expuestas, interesa estudiar los sistemas multi-HVAC para definir un mejor modelo operacional de los subsistemas HVAC a ser utilizados en un momento dado.

B. Sistemas BEMS

Los sistemas BEMS se encargan de monitorear y controlar diferentes elementos y equipos que presentan los edificios, como, por ejemplo, iluminación, sistemas energéticos, sistemas de seguridad, sistemas HVAC, ascensores y demás [13], [14]. Estos sistemas pueden ser empleados tanto para el sector residencial, comercial o público [15]. Por ello, una apropiada gestión de las necesidades energéticas de las edificaciones representa grandes beneficios.

Estos sistemas se caracterizan por el intercambio de información de los procesos y funciones del edificio, permitiendo así a los administradores tomar decisiones de acuerdo a la información recibida [14], de modo que mejore sustancialmente la eficiencia energética en los edificios. Además, algunos estudios señalan que el ahorro de energía en edificios que emplean la funcionalidad de optimización de los BEMS es de hasta un 27% [16], asimismo, otros estudios estiman que existe un ahorro energético de hasta un 20 % para la calefacción de espacios y hasta un 10 % en iluminación y ventilación [17]. Algunos autores consideran los sistemas de gestión de edificios (BMS, por sus siglas en inglés) que se encargan de administrar y controlar además el resto de equipos eléctricos, mecánicos y electromecánicos del edificio, como la iluminación, los elevadores, así como otras funciones del edificio [18]. En este sentido, los BMS son un concepto más general.

III. TRABAJOS RELACIONADOS

La optimización de multiobjetivo (OM) es una técnica utilizada para tratar problemas de optimización con múltiples funciones de coste, con la finalidad de minimizarlos o maximizarlos identificando un equilibrio entre los objetivos [19], [20]. De esta manera no existe una única solución que optimice cada objetivo, sino un conjunto de soluciones no dominantes: Pareto óptimo. Se denomina Pareto óptimo cuando se alcanza un resultado en el que ningún objetivo se puede mejorar sin empeorar las otras funciones objetivo [8], [21]. Dentro de esta perspectiva, en los últimos años, ha habido un interés creciente en la implementación de OM en los sistemas HVAC de los edificios. Por ejemplo, en [22], presentan un enfoque basado en datos para optimizar el consumo total de energía del sistema HVAC de un edificio. Para esto, mediante perceptrón multicapa, dentro del área de las redes neuronales, se construye el modelo de energía total que integra tres modelos de calidad de aire interior de la instalación (modelo de temperatura, modelo de humedad relativa y modelo de concentración de CO_2). En este sentido, desarrollan un problema de optimización de cuatro objetivos, el cual fue resuelto aplicando el algoritmo de optimización de enjambre de partículas modificado, donde asignaron diferentes pesos a los objetivos, logrando optimizar el sistema HVAC con el equilibrio entre el consumo energético y confort térmico.

Otros autores [8] proponen un marco OM para optimizar la configuración HVAC con la temperatura consigna en edificios comerciales, el consumo energético, el confort térmico y la productividad de los ocupantes, dadas ciertas restricciones para los ajustes de temperatura de refrigeración y calefacción, tanto para horas ocupadas como desocupadas. Sus hallazgos, identifican un Pareto óptimo entre las métricas planteadas, señalando la posibilidad de crear estrategias óptimas de operación para los sistemas HVAC.

En [5] proponen una estrategia de operaciones de enfriadoras basada en OM. La estrategia consta de tres pasos: (I), se adquieren dos indicadores de función de efectividad, confort interior y consumo de energía; (II) estos indicadores se integran en un indicador exhaustivo; (III) emplean un algoritmo de optimización para maximizar este indicador. Llegaron a la conclusión de que la estrategia es capaz de reducir el consumo energético con una pérdida escasa de confort interior, además, de poder equilibrar el tiempo de trabajo de las enfriadoras, ajustando el peso de los objetivos, lo que posibilita satisfacer la necesidad de optimización de un edificio en concreto. También señalan que esta estrategia es adecuada para enfriadoras con caudal de agua enfriada constante.

Por otro lado, el estudio Schito [23] propone un método para tratar simultáneamente con tres objetivos: (1) el objetivo de eficiencia energética, donde emplean el indicador de carga termo-higrométrica total que debe proporcionar el sistema HVAC; (2) el objetivo de confort térmico, utilizan como indicador el Porcentaje de Predicción de Insatisfechos (PPD) y (3) el objetivo de conservación de arte, hacen uso del indicador multiplicador de vida útil (LM) para la evaluación del riesgo de arte de la naturaleza química. Este método está basado en OM con la finalidad de elegir adecuadamente las temperaturas consignas y humedad relativa, mediante la simulación dinámica y evaluación de estos objetivos. Los autores señalan que esta estrategia puede ser efectiva en museos alojados en edificios históricos, dado que la renovación está restringida por la arquitectura.

Por otra parte, en [9], proponen una metodología de optimización de energía para disminuir el consumo energético de una planta enfriadora, derivando las operaciones de la planta de acuerdo a la carga de enfriamiento y las condiciones climáticas. El modelo energético desarrollado simula el comportamiento de cada subsistema de la planta enfriadora incluso el consumo de energía. Sus hallazgos tras aplicar esta metodología en diferentes estudios de caso, fue que las operaciones de las plantas enfriadoras de dimensiones moderadas fueron optimizadas hasta un 40% y en torno a un 20% las de menor dimensión, disminuyendo así el costo energético y las emisiones de GEI.

IV. MODELADO DEL SISTEMA MULTI-HVAC

En esta sección describiremos la base física de los modelos del sistema multi-HVAC, para esto, se asumió de forma individual los diferentes subsistemas del HVAC del edificio. En este sentido, consideramos los siguientes componentes para desarrollar los modelos físicos: torre de refrigeración, bombas de agua de refrigeración, bombas del circuito primario con caudal constante y los compresores de las bombas de calor y grupos de frío.

A. *Modelo Potencia Consumida del sistema multi-HVAC*

Las enfriadoras y bombas de calor son los principales componentes de los sistemas HVAC [12], los cuales se encargan de transferir o extraer energía térmica de una zona de alta o baja temperatura a otra. Estos subsistemas están compuestos por cuatro partes principales: evaporador, condensador, compresor y válvula de expansión.

Para el modelado físico del consumo energético del momento actual de los subsistemas HVAC, nos basamos en la ecuación definida en [9]:

$$P_{consumida}(hvac,t) = \sum_{h \in hvac} \Big(P_c(h,t) + P_{tr}(h) + P_{bar}(h) + P_{bpae}(h)\Big) \quad (1)$$

Donde, h son los diferentes subsistemas HVAC, t es el consumo en un momento dado por hora, $P_c(h,t)$ es la potencia consumida del compresor de cada h en un momento t, $P_{tr}(h)$ es la potencia consumida de la torre de refrigeración, $P_{bar}(h)$ es la potencia consumida de la bomba de agua de refrigeración y $P_{bpae}(h)$ es la potencia consumida de la bomba primaria de agua enfriada. Cada uno se expresa en kW. Cabe señalar que los subsistemas HVAC pueden ser condensados tanto por aire, agua u otro elemento [24], [25]. En este sentido, en (1) se definen las variables para el subsistema HVAC condensado por agua, sin embargo, para cuando es condensado por aire, las variables $P_{tr}(h)$ y $P_{bar}(h)$ no existen (en nuestro modelo serán iguales a cero). Además, $P_{tr}(h)$, $P_{bar}(h)$ y $P_{bpae}(h)$ son dadas por el propietario, ya que fueron definidas durante el diseño del sistema HVAC. Para nuestro modelo, este puede ser minimizado.

La ecuación (2) describe la carga frigorífica o calorífica que el subsistema HVAC es capaz ceder o extraer del agua enfriada en el circuito.

$$Carga_{f,c}(h,t) = Q_f(h,t) * Ce_f(h) * \rho_f(h) * \Delta T(h,t) \quad (2)$$

Donde, $Q_f(h,t)$ es el caudal del fluido del subsistema HVAC (m^3/s), $Ce_f(h)$ es el calor específico del fluido del subsistema HVAC (kJ/kgK), $\rho_f(h)$ es la densidad del fluido utilizado en el subsistema HVAC (kg/m^3), y $\Delta T(h,t)$ es la diferencia de temperatura de entrada y salida del subsistema HVAC (K), el fluido puede ser agua o gas, en nuestro modelo es agua

La ecuación (3) describe la capacidad de generación térmica del subsistema HVAC. Esta depende de la carga frigorífica o calorífica y la capacidad nominal del subsistema HVAC.

$$\mathrm{Cap}(h,t) = \begin{cases} Carga_{f,c}(h,t), & Carga_{f,c}(h,t) < CAP_{nom}(h) \\ CAP_{nom}(h), & Carga_{f,c}(h,t) \geq CAP_{nom}(h) \end{cases} \quad (3)$$

Donde, $Carga_{f,c}(h,t)$ se obtiene en (2) y la $CAP_{nom}(h)$ es dada por parte del fabricante durante el diseño del sistema HVAC en (kW).

$$P_c(h,t) = Cap(h,t)/COP_{pr}(h) \quad (4)$$

La ecuación (4) describe la potencia consumida por parte del compresor del subsistema HVAC, es una relación entre la capacidad de generación térmica $Cap(h,t)$ y el coeficiente de rendimiento $COP_{pr}(h)$ o $EER_{pr}(h)$ de cada subsistema HVAC (este rendimiento está dado por el modelo de datos del fabricante tanto para modo refrigeración y modo calefacción).

B. *Modelo Costo Energético*

Para el modelo del costo energético de los subsistemas HVAC utilizamos la siguiente ecuación:

$$Coste(h,t) = P_{consumida}(h,t) * TE_i \ para\ t \in i \quad (5)$$

Donde, $P_{consumida}(h,t)$ es la potencia consumida total (kW) de los sistemas HVAC obtenida en (1) y TE_i es el costo de la potencia para el periodo tarifario *i* que aplica para el momento *t* (h) para la potencia consumida (€/kWh). En nuestro modelo el periodo tarifario *i* puede ser optimizado.

C. *Modelo de temperatura consigna*

El modelo de la temperatura de consigna es la capacidad del sistema HVAC en alcanzar una temperatura consigna en un rango confortable dentro del edificio. Para ello empleamos la siguiente ecuación:

$$Tsp_{rango} = (T_{sp} - T_{actual})/T_{sp} \quad (6)$$

Donde, T_{sp} es la temperatura consigna que se requiere alcanzar para el modo refrigeración y calefacción (K) y T_{actual} es la temperatura actual del edificio (K).

D. *Modelo Rendimiento del sistema multi-HVAC*

El modelo del rendimiento del sistema multi-HVAC es la relación entre la capacidad generada por cada sistema HVAC ($CapGen_{HVAC}$) y la potencia consumida ($P_{consumida}$) de los sistemas multi-HVAC. Esta relación, la representamos mediante la siguiente ecuación:

$$COP_{global} = CapGen_{HVAC}/P_{consumida} \quad (7)$$

V. RESULTADOS Y DISCUSIÓN

En esta sección discutimos algunos de los resultados obtenidos tras ejecutar los modelos descritos en la sección anterior. Para evaluar estos modelos, se utilizó el software MATLAB® (v. R2018b) como plataforma de desarrollo, de igual forma nos basamos en los datos físicos ofrecidos por los fabricantes de bombas de calor Galletti®, modelo LEW H 214 [26] y grupos de frío Carrier®, modelo 30HXC-120 [27].

En la Tabla I se muestran una comparativa de los datos del fabricante y los datos aproximados del caudal del fluido (Q) y del coeficiente de rendimiento (COP o EER), tanto para modo refrigeración y modo calefacción de los subsistemas HVAC. Para la aproximación de los parámetros desarrollamos modelos de regresión lineal de acuerdo con los datos físicos de cada fabricante. Estos resultados están dados por las siguientes condiciones de temperatura: para el modo de refrigeración, 30ºC de temperatura de ambiente exterior y 7ºC de temperatura de agua suministrada; mientras que para el modo calefacción, 10ºC de temperatura de ambiente exterior y 45ºC de temperatura de agua suministrada.

TABLA I. COMPARATIVA DE LOS DATOS DEL FABRICANTE Y DATOS APROXIMADOS

	Modo Refrigeración				**Modo Calefacción**			
	Datos fabricante		***Datos aproximados***		***Datos fabricante***		***Datos aproximados***	
	Q	***EER***	***Qa***	***EERa***	***Q***	***COP***	***Qa***	***COPa***
LEW-H-214	11.2	5.51	**11.3**	**5.67**	13.4	4.94	**13.3**	**4.99**
30HXC-120	20.7	4.86	**19.8**	**4.87**	22.3	4.34	**21.3**	**4.38**

La Tabla II proporciona una visión general de los diferentes coeficientes de determinación (R^2) de los subsistemas HVAC mostrados en la Tabla I. Lo interesante de estos datos es que cada modelo presenta una alta precisión para la aproximación de los datos; sin embargo, para el modelo de aproximación del caudal de los grupos de frío en modo calefacción presenta solo un 64% de precisión, lo que nos permite concluir que estos equipos no son óptimos para brindar calefacción a las distintas zonas del edificio.

TABLA II. COEFICIENTE DE DETERMINACIÓN (R^2) DE LOS SUBSISTEMAS HVAC ANALIZADOS

Bomba de calor LWE-H-214		R^2
EERa	Modo Refrigeración	96 %
Qa		99 %
COPa	Modo Calefacción	97 %
Qa		99 %
Grupo de frío 30HXC-120		R^2
EERa	Modo Refrigeración	99 %
Qa		99 %
COPa	Modo Calefacción	99 %
Qa		64 %

Por otro lado, para realizar la simulación del comportamiento y funcionamiento de los subsistemas HVAC, nos basamos en los datos climatológicos de los últimos años de la zona de Madrid, en España, utilizando dos curvas sinusoides (curva para la temporada y una curva diaria) para simular las temperaturas de un año en horas (8760 horas). Asimismo, acotamos el comportamiento de los subsistemas HVAC para emplear el cambio de modo calefacción a refrigeración de los equipos.

La Fig. 1. muestra el comportamiento de una bomba de calor y un grupo de frío en el transcurso de un año, donde, se aprecia que las bombas de calor presentan un mayor consumo para ofrecer calefacción, sin embargo, este consumo va decreciendo mientras se acerca la temporada para refrigerar, donde el mismo mantiene un comportamiento prácticamente constante.

En contraste, el grupo de frío presenta una relación inversa a la bomba de calor, donde mantiene un consumo más o menos constante y mucho mayor que las bombas de calor en modo calefacción, mientras que para el modo refrigeración el equipo presenta un comportamiento variable comparado a las bombas de calor.

La Fig. 2 muestra el comportamiento del costo energético de los subsistemas HVAC durante los diferentes periodos tarifarios en el año. De igual forma, se puede apreciar un comportamiento similar de ambos subsistemas HVAC con respecto a la Fig.1. ya que existe una relación directa entre ambos modelos.

VI. CONCLUSIONES

Este trabajo contribuye al conocimiento existente sobre sistemas HVAC proporcionando un análisis del comportamiento de diferentes subsistemas HVAC presentes en las edificaciones, dado que existen numerosos estudios que analizan las estrategias operacionales de solo un tipo de subsistema HVAC (enfriadores).

En este sentido, este estudio ha propuesto cuatro modelos basados en la física que analizan el comportamiento y funcionamiento de los subsistemas HVAC (bombas de calor y grupos de frío) de un edificio en particular (Teatro Real, aunque puede ser empleado para otro tipo de edificios). Para este análisis incorporamos en nuestro estudio, ajustes mediante modelos de aproximación de los datos físicos de los sistemas HVAC de los fabricantes (Tabla I y II), ya que estos no proporcionan una ecuación que represente el funcionamiento y comportamiento de sus equipos, de modo que estos modelos se aproximan en gran medida con los reales.

Como trabajo futuro, implementaremos estos modelos físicos en algoritmos de optimización multiobjetivo, con la finalidad de optimizar y obtener un conjunto de soluciones no dominadas (Pareto óptimo), que nos permita seleccionar en un momento dado los diferentes modos de operación de los sistemas multi-HVAC, con la finalidad de mejorar significativamente la eficiencia energética de estos sistemas en casos reales.

AGRADECIMIENTOS

Me gustaría hacer una mención especial a la Secretaría Nacional de Ciencia, Tecnología e Innovación (SENACYT) y al Instituto para la Formación y Aprovechamiento de Recursos Humanos (IFARHU), por la beca doctoral otorgada dentro del programa de becas doctorales IFARHU-SENACYT. De igual forma, agradezco a aquellos profesores e investigadores que dieron sus aportes y recomendaciones para la redacción de este artículo.

REFERENCIAS

[1] United Nations Department of Economic and Social Affairs, *Concise report on the world population situation in 2014*. 2014.

[2] C. Lim, K.-J. Kim, and P. P. Maglio, "Smart cities with big data: Reference models, challenges, and considerations," *Cities*, vol. 82, pp. 86–99, Dec. 2018.

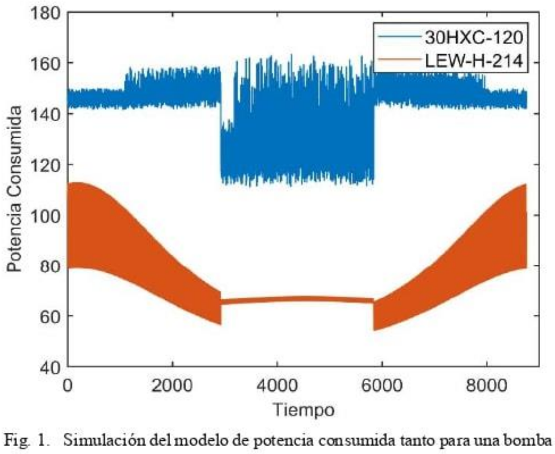


Fig. 1. Simulación del modelo de potencia consumida tanto para una bomba de calor como para un grupo de frío en el transcurso de un año.

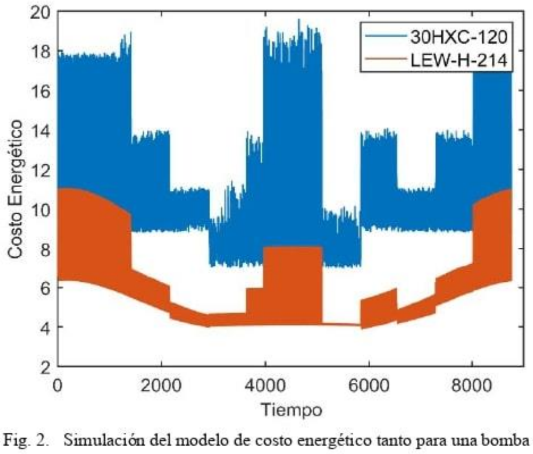


Fig. 2. Simulación del modelo de costo energético tanto para una bomba de calor como para un grupo de frío en el transcurso de un año

[3] B. N. Silva, M. Khan, and K. Han, "Towards sustainable smart cities: A review of trends, architectures, components, and open challenges in smart cities," *Sustainable Cities and Society*, vol. 38, pp. 697–713, Apr. 2018.

[4] G. M. Mauro, M. Hamdy, G. P. Vanoli, N. Bianco, and J. L. M. Hensen, "A new methodology for investigating the cost-optimality of energy retrofitting a building category," *Energy and Buildings*, vol. 107, pp. 456–478, Nov. 2015.

[5] S. Qiu, W. Zhang, J. Li, J. Chen, Z. Li, and Z. Li, "A chiller operation strategy based on multiple-objective optimization," *Energy Procedia*, vol. 152, pp. 318–323, 2018.

[6] L. Pérez-Lombard, J. Ortiz, and C. Pout, "A review on buildings energy consumption information," *Energy and Buildings*, vol. 40, no. 3, pp. 394–398, Jan. 2008.

[7] P. Rocha, A. Siddiqui, and M. Stadler, "Improving energy efficiency via smart building energy management systems: A comparison with policy measures," *Energy and Buildings*, vol. 88, pp. 203–213, Feb. 2015.

[8] S. Papadopoulos and E. Azar, "Optimizing HVAC operation in commercial buildings: A genetic algorithm multi-objective optimization framework," in *2016 Winter Simulation Conference (WSC)*, 2016, pp. 1725–1735.

[9] S. R. Thangavelu, A. Myat, and A. Khambadkone, "Energy optimization methodology of multi-chiller plant in commercial buildings," *Energy*, vol. 123, pp. 64–76, Mar. 2017.

[10] M. W. Ahmad, M. Mourshed, B. Yuce, and Y. Rezgui, "Computational intelligence techniques for HVAC systems: A review," *Build. Simul.*, vol. 9, no. 4, pp. 359–398, Aug. 2016.

[11] M. Porowski, "The optimization method of HVAC system from a holistic perspective according to energy criterion," *Energy Conversion and Management*, vol. 181, pp. 621–644, Feb. 2019.

[12] Z. Afroz, G. Shafiullah, T. Urmee, and G. Higgins, "Modeling techniques used in building HVAC control systems: A review," *Renewable and Sustainable Energy Reviews*, vol. 83, pp. 64–84, Mar. 2018.

[13] J. Cser, R. Beheshti, and P. van der Veer, "Towards the development of an Integrated Building Management System," in *Innovation in Technology Management. The Key to Global Leadership. PICMET '97*, 1997, pp. 740-.

[14] "Building Energy Management Systems (BEMS) | ClimateTechWiki." [Online]. Available: https://www.climatetechwiki.org/technology/jiqweb-bems. [Accessed: 24-May-2019].

[15] S. Beucker, J. D. Bergesen, and T. Gibon, "Building Energy Management Systems: Global Potentials and Environmental Implications of Deployment: Building Energy Management Systems," *Journal of Industrial Ecology*, vol. 20, no. 2, pp. 223–233, Apr. 2016.

[16] J. Reynolds, Y. Rezgui, A. Kwan, and S. Piriou, "A zone-level, building energy optimisation combining an artificial neural network, a genetic algorithm, and model predictive control," *Energy*, vol. 151, pp. 729–739, May 2018.

[17] F. Wang, H. Yoshida, and E. Ono, "Methodology for optimizing the operation of heating/cooling plants with multi-heat-source equipments," *Energy and Buildings*, vol. 41, no. 4, pp. 416–425, Apr. 2009.

[18] H. V. V. Priyadarshana, W. K. I. Madushanaka, L. L. L. Anuruddha, G. T. Chathura, H. W. D. Hettiarachchi, and K. T. M. U. Hemapala, "Multi-agent controlled building management system," in *2017 Innovations in Power and Advanced Computing Technologies (i-PACT)*, 2017, pp. 1–5.

[19] G. Chiandussi, M. Codegone, S. Ferrero, and F. E. Varesio, "Comparison of multi-objective optimization methodologies for engineering applications," *Computers & Mathematics with Applications*, vol. 63, no. 5, pp. 912–942, Mar. 2012.

[20] K. Deb and K. Deb, "Multi-objective Optimization," in *Search Methodologies: Introductory Tutorials in Optimization and Decision Support Techniques*, E. K. Burke and G. Kendall, Eds. Boston, MA: Springer US, 2014, pp. 403–449.

[21] E. Zitzler and L. Thiele, "Multiobjective optimization using evolutionary algorithms — A comparative case study," in *Parallel Problem Solving from Nature — PPSN V*, 1998, pp. 292–301.

[22] X. Wei, A. Kusiak, M. Li, F. Tang, and Y. Zeng, "Multi-objective optimization of the HVAC (heating, ventilation, and air conditioning) system performance," *Energy*, vol. 83, pp. 294–306, 2015.

[23] E. Schito, P. Conti, and D. Testi, "Multi-objective optimization of microclimate in museums for concurrent reduction of energy needs, visitors' discomfort and artwork preservation risks," *Applied Energy*, vol. 224, pp. 147–159, 2018.

[24] Climatek, "Tres tipos de condensadores HVAC," *Aire acondicionado Madrid*, 19-Mar-2018. [Online]. Available: https://climatekmadrid.com/tres-tipos-de-condensadores-hvac/. [Accessed: 25-May-2019].

[25] "Cooling Technology - Evaporative Condensed Chillers." [Online]. Available: http://www.coolingtechnology.com/about_process_cooling/evaporative_condensed_chillers/default.html. [Accessed: 26-May-2019].

[26] Galletti, "Reversible Chillers and Heat Pumps - LEW SERIES 40-480 KW." .

[27] Carrier, "Water-Cooled Screw-Compressor Liquid Chillers - 30HXC." .

2.3 Publicación 3: Developing a Long Short-Term Memory-Based Model for Forecasting the Daily Energy Consumption of Heating, Ventilation, and Air Conditioning Systems in Buildings

2.3.1 Resumen de la publicación

Pronosticar el consumo energético de los sistemas HVAC en edificios es muy complejo e imprevisible, debido a la influencia de factores como, las horas de trabajo, las actividades del edificio, el clima, la ubicación geográfica, entre otros. De hecho, los modelos tradicionales tienen limitaciones al capturar los comportamientos dinámicos y no-lineales presentes en los sistemas HVAC. Es por ello, que, este artículo propone desarrollar y evaluar un modelo basado en Deep Learning, más específicamente el modelo Long Short-Term Memory (LSTM), para el pronóstico diario del consumo energético de los sistemas HVAC en los edificios. Las redes LSTM son muy útiles para tratar con series temporales, debido a que tienen la capacidad de "memorizar" la información de grandes secuencias de datos durante el aprendizaje. Asimismo, son muy utilizadas para tareas de pronóstico del clima, procesamiento del lenguaje, traducción, entre otros. En esta perspectiva, se llevó a cabo un procedimiento riguroso a través del ajuste de diversos hiperparámetros, a fin de obtener un modelo robusto, confiable y generalizable. Para esto, realizamos múltiples ejecuciones de los hiperparámetros con diferentes valores, detectando así que el modelo propuesto logró un desempeño significativo sobre la métrica del coeficiente de variación de la raíz del error cuadrático medio (CVRMSE) del 9.5%; es decir, obtuvo un valor más bajo respecto a lo recomendado por las agencias internacionales.

2.3.2 Contribuciones significativas de la publicación

En esta tercera publicación, el doctorando realizó diferentes contribuciones significativas a la comunidad científica: (i) se presenta un modelo de Deep Learning (DL), basado en redes LSTM, dirigido a pronosticar el consumo energético diario de los sistemas HVAC en edificios; (ii) el modelo propuesto es capaz de reconocer los patrones intrínsecos dentro de los parámetros de temperatura exterior, ocupación del edificio, temperatura interior del edificio, horas de trabajo, las actividades del edificio, entre otros; (iii) el modelo resultante logra un desempeño inferior al 30% propuesto por las diferentes agencias internacionales; (iv) el modelo es capaz de capturar las variaciones temporales propias de los sistemas HVAC en los edificios, así como de generar pronósticos de consumo energético confiables; (v) este modelo basado en redes LSTM puede ser empleado para la toma de decisiones inteligentes, tales como, mantenimiento preventivo, detección y diagnóstico de fallas, análisis del mercado energético, optimización del funcionamiento del sistema HVAC, entre otros.

2.3.3 Índices de calidad de la publicación

Autores: L. Mendoza-Pittí, H. Calderón-Gómez, J. M. Gómez-Pulido, M. Vargas-Lombardo, J. L. Castillo-Sequera, and C. S. de Blas

Título: Developing a Long Short-Term Memory-Based Model for Forecasting the Daily Energy Consumption of Heating, Ventilation, and Air Conditioning Systems in Buildings

Fecha de publicación: 22 de julio de 2021

MDPI - Applied Science, 2021

Volumen 11, Issue 15

ISBN: 2076-3417

DOI: 10.3390/app11156722

Ranking JCR 2020: Engineering, multidisciplinary (Q2).

Factor de impacto (JCR 2020): 2.679

Ranking SJR 2020: Computer Science Applications (Q2).

Factor de impacto (SJR 2020): 0.435

2.3.4 Publicación

applied sciences

MDPI

Article

Developing a Long Short-Term Memory-Based Model for Forecasting the Daily Energy Consumption of Heating, Ventilation, and Air Conditioning Systems in Buildings

Luis Mendoza-Pittí [1,2], Huriviades Calderón-Gómez [1,2], José Manuel Gómez-Pulido [1,*], Miguel Vargas-Lombardo [2,*], José Luis Castillo-Sequera [1] and Clara Simon de Blas [3]

1 Department of Computer Science, University of Alcalá, 28805 Alcalá de Henares, Spain; l.mendoza@edu.uah.es (L.M.-P.); huriviades.calderon@edu.uah.es (H.C.-G.); jluis.castillo@uah.es (J.L.C.-S.)
2 SNI, Senacyt Panama, E-Health and Supercomputing Research Group (GISES), Technological University of Panama, Panama City 0819-07289, Panama
3 Department of Computer's Sciences, Technical School of Computer Engineering, Rey Juan Carlos University, 28933 Madrid, Spain; clara.simon@urjc.es
* Correspondence: jose.gomez@uah.es (J.M.G.-P.); miguel.vargas@utp.ac.pa (M.V.-L.)

Abstract: Forecasting the energy consumption of heating, ventilating, and air conditioning systems is important for the energy efficiency and sustainability of buildings. In fact, conventional models present limitations in these systems due to their complexity and unpredictability. To overcome this, the long short-term memory-based model is employed in this work. Our objective is to develop and evaluate a model to forecast the daily energy consumption of heating, ventilating, and air conditioning systems in buildings. For this purpose, we apply a comprehensive methodology that allows us to obtain a robust, generalizable, and reliable model by tuning different parameters. The results show that the proposed model achieves a significant improvement in the coefficient of variation of root mean square error of 9.5% compared to that proposed by international agencies. We conclude that these results provide an encouraging outlook for its implementation as an intelligent service for decision making, capable of overcoming the problems of other noise-sensitive models affected by data variations and disturbances without the need for expert knowledge in the domain.

Keywords: daily energy consumption; deep learning; forecasting model; HVAC systems; long short-term memory; short-term forecast

check for updates

Citation: Mendoza-Pittí, L.; Calderón-Gómez, H.; Gómez-Pulido, J.M.; Vargas-Lombardo, M.; Castillo-Sequera, J.L.; de Blas, C.S. Developing a Long Short-Term Memory-Based Model for Forecasting the Daily Energy Consumption of Heating, Ventilation, and Air Conditioning Systems in Buildings. *Appl. Sci.* **2021**, *11*, 6722. https://doi.org/10.3390/app11156722

Academic Editor: Jürgen Reichardt

Received: 29 June 2021
Accepted: 20 July 2021
Published: 22 July 2021

1. Introduction

According to the US Energy Information Administration (EIA) [1] and the European Commission [2], buildings represent between 39% and 40% of world energy consumption, where the heating, ventilating, and air conditioning (HVAC) systems account for more than half of the final energy consumption of buildings [3]. In fact, the energy consumption of HVAC systems is very complex and unpredictable at times because the use given to a building is very different throughout the day due to factors such as working hours, different types of personnel and their activities in the building, time dedicated to equipment maintenance, and events that take place in the building. In addition, these systems are affected by other factors, such as climate, geographical location, and building envelope, among others [4,5].

The dynamic and nonlinear behavior of the HVAC system is difficult to manage, due to different subsystems working together. Some of these subsystems are: the compressor, fan-coils, heat pumps, chillers, and the air distribution system, among others [6,7]. Therefore, the recorded time series of HVAC system operations are chaotic.

In this sense, it is crucial to understand and analyze the energy consumption of the HVAC system. This means that anticipating the energy consumption of HVAC systems has great importance in the context of the energy efficiency and sustainability of buildings.

This would allow the implementation of planning strategies, optimization of the operations of these systems, and reduction in energy peaks and associated costs [8,9]. However, before building a model for forecasting the energy consumption of HVAC systems, it is important to consider two fundamental factors: the past observations or time lags and the time horizon to be forecasted [4,8].

On the one hand, the time lags have been of interest in recent years since they allow for an increase in the accuracy capacity of the predictions of the dynamic environment models [8] and, on the other hand, the forecast time horizon delimits from which point in time the model will carry out the forecast. The steps of this action plan can be regrouped into three phases: short-, medium- and long-term [10,11]. Short-term forecasts are focused on the day-to-day operations of energy systems, peak energy demand, daily energy consumption, short-term electric market, fault detection, and diagnosis, etc. [12,13]. Medium-term and long-term forecasts are directed to determine equipment modernization, create energy saving strategies, and modify electric market plans, among others [8].

Within this framework, numerous data-driven models (i.e., statisticians and machine learning (ML)) have been used for the short-term forecasting of the energy consumption of HVAC systems in buildings. Some of these are: autoregressive integrated moving average (ARIMA), multiple-linear regression (MLR), support vector machine (SVM), decision tree (DT), random forest (RF), and artificial neural networks (ANNs), among others [14–19]. While these models offer a good approximation when forecasting/predicting the energy consumption of the HVAC system in buildings, they are very generalist, noise-sensitive models and have a lower level of abstraction [16,20–25] since the data of these systems have many disturbances and variations while they are in ordinary operation. In addition, the various processes and complex features inherent in buildings significantly affect the behavior of HVAC systems [5,18,26,27].

To overcome the shortcomings of the aforementioned models, deep learning (DL)-based models have exposed promising alternatives since they allow automatic discovery of the intrinsic structure of the data without the need for expert knowledge in the domain [15,20,28,29]. Therefore, DL models are applicable to the analysis and forecasting of complex and dynamic building energy systems such as HVAC systems. In fact, DL models based on the long short-term memory (LSTM) neural network are very useful in this context (as discussed in the following sections).

However, the implementation of these DL models brings with it the correct and precise tuning and implementation of different hyperparameters (HPs), such as the number of hidden neurons, activation function, number of epochs, learning rate, optimizer, regularization technique, etc. Therefore, these HPs will determine the abstraction capability of a DL model.

In this context, the present study has focused on the development of an LSTM-based model that is capable of forecasting the daily energy consumption of HVAC systems in buildings. Another important objective is to evaluate the proposed model with respect to different time lags, HP configurations, and to implement regularization techniques (such as dropout and early stopping). To identify the best configuration of the model in this scope, multiple experiments have been conducted to compare the performance of each configuration. Therefore, the developed research provides an insight into the strengths and weaknesses of applying some of the configurations analyzed and described in the present article. It should be noted that this research is part of the research project called "Intelligent management system for optimizing energy consumption in building air conditioning" [7,30–33].

We can summarize the contributions of this study as follows: (i) we analyze in depth real-world data on the energy consumption of the HVAC system in a complex building such as the Teatro Real; (ii) we deeply examine the past behavior of a chaotic time series (noisy data and sensor errors) to make a daily forecast of the energy consumption of the HVAC system; (iii) we apply a comprehensive methodological process that allows us to obtain a robust, generalizable and reliable model by tuning different parameters; (iv) we evaluate the advantages and disadvantages of applying regularization techniques in the context of HVAC systems; (v) we identify that the proposed model achieves reliable performance on the CVRMSE metric of 9.5% which is significantly lower than recommended by international agency guidelines; (vi) we determine the configuration of our model to achieve reliable and robust forecasts of HVAC energy consumption.

The article is organized as follows. In Section 2, we present the related work regarding DL-based models aimed at forecasting energy consumption in HVAC systems, including motivation for the use of LSTM networks. In Section 3, we briefly describe how to address the problem in a particular real-world use case and how it has been addressed in this study. In Section 4, we detail the methodological procedure we carried out to build the LSTM-based DL model. Section 5 shows different configurations of the evaluated model. In Section 6, we analyze and discuss how the model configurations were performed. In addition, we identify the best configuration of the model regarding its time lags, HPs and regularization techniques. Finally, Section 7 provides some conclusions from the study carried out and presents future research work to be developed.

2. Related Works

In recent years, a large number of investigations have focused on the area of artificial intelligence (AI) work with models based on deep learning (DL) techniques, in which they have obtained surprising results in both academic and industrial fields [3,34–36]. Furthermore, these DL techniques have been used to model complex and sophisticated patterns for the prediction/forecasting of time series in the energy sector [37]. DL techniques use different neural network structures that allow learning from significant data representations, and in this sense, neural networks, such as the multilayer perceptron, recurrent neural network, have provided useful information for energy decision making for building managers/administrators, such as the prediction/forecast of the energy consumption variables of residential and nonresidential buildings [38–43], building temperature [37,44], energy load of the building [45–47], and heating or cooling load of the building [48–50].

Within the aforementioned perspective, there are different proposals for DL-based model architectures aimed at forecasting the energy consumption of HVAC systems in buildings. Next, we describe some relevant works in this context.

Hwang, J. K. et al. [3] proposed DL models for predicting the performance and energy consumption of a heating and cooling (HC) system. They used a feed-forward ANN algorithm based on back-propagation in which they selected several hyperparameters (HPs), such as the activation function, number of hidden layers and nodes, and regularization techniques, to adjust their model. Moreover, they also tested the proposed models using different time intervals and selected the most appropriate input variables for the models. Finally, they emphasized that HC system operating variables are more important than the internal and external temperatures of the building and noted that the DL approach could establish strategies to predict the energy performance of the HC system.

Likewise, J. Cho et al. [51] benchmarked DL-based algorithms (MLP and LSTM) for air conditioning load forecasting to reduce the energy consumption of the building. Their tests found that the LSTM had fewer errors than MLP. They verified that the DL algorithms are capable of solving load forecasting problems.

Moreover, Machida, Y. et al. [52] proposed an estimation method for energy consumption for building air conditioning systems. They used an RNN with five neurons in the middle layer and time lags of three states. In addition, they used data from a simulated building in the summer and winter seasons, incorporating variables such as preset temperature, on-off state, and the number of people in the zone/room, among others. In this study, the authors demonstrated that, with the proposed method, it is possible to estimate energy consumption.

On the other hand, Sendra-Arranz, R. and Gutiérrez, A. [4], developed different LSTM-based models for short-term prediction of the HVAC system of a real self-sufficient solar house based on the previous day's behavior. They only experimented with the number of neurons and learning rate. In their work, they highlighted that these results drive the prediction of energy consumption in real-time in buildings.

Similarly, Zhou, C. et al. [25] compared predictive models, such as the ARIMA model, Multilayer neural network model, and LSTM model for predicting the daily and time energy consumption of air conditioning systems. The authors noted that some HPs tuning had to be made with respect to the LSTM model. They validated their models by comparing the mean average percentage error (MAPE) of each. According to the results obtained, the LSTM model had high adaptability and accuracy in predicting only the energy consumption data of the air conditioning systems.

In the work of Ellis M. J. and Chinde, V. [53], an encoder-decoder LSTM-based economic model predictive control (EMPC) framework was developed. They only tuned one, which was the number of nodes in the LSTM layers. This framework was applied to a building HVAC system and used the EnergyPlus building energy simulation program to simulate a multizone building. The authors emphasize that their model can accurately predict the indoor temperature and sensitive HVAC cooling rate of an area in the building for a two-day horizon.

Additionally, Hwang, I. et al. [54] proposed a *seq2seq* (encoder-decoder) model based on LSTM for estimating the energy consumed by air conditioning. They described some HPs such as the number of LSTM layers for the encoder and decoder, the number of nodes, the activation function, and the regularization technique. The authors contrasted the proposed model with other algorithms and concluded that the LSTM-based *seq2seq* model is superior because it effectively combines past and future knowledge.

From the researches cited above, we can deduce that different architectures based on DL models have been used to forecast the energy consumption of HVAC systems in buildings, such as MLP, RNN, and LSTM models. However, HVAC system data are a sequence of values recorded in a period of time (time series) [29], so future predictions of HVAC energy consumption depend not only on the current values in the inputs but also on the overall trend itself that past data maintain. This concept is known as long-term dependency, and both the MLP network and standard RNN do not have the capacity to handle this dependency. MLP models do not consider data as time series and RNNs do not have the ability to retain and consider long time lags during learning [55]. From this perspective, LSTM networks have the ability to transmit the cell state information through the different LSTM units, which is why the model is able to maintain the information of large historical sequences, as well as to consider the data as a sequence of ordered data [48,56]. For this reason, an LSTM-based model is ideal for forecasting the energy consumption of HVAC systems in buildings.

Taking into account the above, most of the previous studies only selected some hyperparameters for the development of DL-based models. However, few studies deeply examined the past behavior of the time series to perform a short-term forecast, as well as the use of regularization techniques such as dropout and early stopping within the training environment of the model. Therefore, it is necessary to build and evaluate a DL model based on LSTM in which different time lags are considered, various HPs settings are used, and the regularization techniques mentioned for the deep neural network (DNN) are incorporated, to identify the best configuration for performing daily forecasting of HVAC system energy consumption in buildings with complex usage.

3. Problem Formulation

This research focuses on short-term energy consumption forecasting of HVAC systems in buildings. Since the predicted consumed energy for the HVAC system would act as a baseline, this would allow managers/administrators to carry out intelligent strategies on the HVAC system, such as optimal control, predicting its behavior, performing preventive maintenance, enabling fault detection, and diagnosis, etc., through the use of artificial intelligence models such as that proposed in this study [57–59]. In this section, we will briefly describe the problem and how we addressed it in our study.

3.1. HVAC System in Buildings

The main function of HVAC systems is to provide building occupants with thermal comfort and good indoor air quality. In fact, these systems are characterized by being composed of multiple subsystems, such as heat pumps, chillers, cooling towers, boilers, and air handling units (AHU) [6,7]. In addition, the HVAC system is responsible for more than half of the total energy consumption of the building and has great potential for energy savings [3,60,61].

On the other hand, buildings are adopting a centralized management system based on software, called the Building Management System (BMS). This is characterized by having various analog and digital devices (actuators, sensors, etc.), whose objective is to provide the building with an architecture for the management, monitoring, and control of HVAC and lighting systems, among others [7,31,33]. The BMS is responsible for collecting and storing information on building energy consumption and some other variables, such as environmental variables (outdoor temperature, wind speed, humidity, etc.), building variables (internal temperature, building usage, programming, occupant behavior, etc.), and equipment operating variables (HVAC sensors, energy consumption, operating status, etc.) [27], as shown in Figure 1.

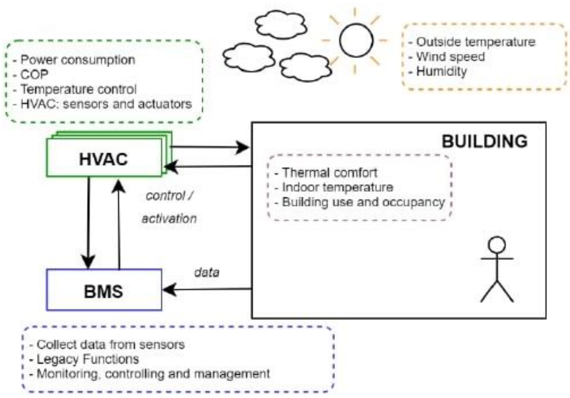


Figure 1. Management of the building environment and the HVAC system using a BMS.

Now, through the BMS, the building data are registered and indexed in time intervals. Therefore, data from different building systems (i.e., HVAC, lighting, etc.) are treated as time series. Within this framework, HVAC data can be considered a chaotic time series since they present high volatility, nonlinearity, and nonstationarity due to the different subsystems that compose it, as well as many factors that affect the HVAC system, such as the building usage, the specifications and features of the building, the behavior of the occupant, and the weather conditions [4,6,25,62]. Even the complexity of these time series increases in buildings that have areas/zones of mixed use or activity such as theaters, offices, and shopping centers [5].

In this sense, it is necessary to build and evaluate a DL-based model that considers these temporal variations inherent to HVAC data and some external and internal factors that influence it.

3.2. Long Short-Term Memory (LSTM)

Taking into account the above, and as indicated in Section 2 (Related works), a model based on LSTM is suitable for the development of models for time series predictions, as they have a better capacity to deal with temporal dependencies as well as the detection of complex patterns, which are inherent in HVAC data in buildings [51,63]. In other words, LSTM networks have greater robustness for the handling of continuous values and present good performance with time series data of large time lags [25]. Within this context, this neural network offers a solution to the two main problems that RNNs present at the moment of learning: (i) it is not able to consider time lags far from the time series and (ii) it tends to present unstable gradients (vanishing or exploding) [4,29,55].

The LSTM network, which is an improved variant of RNN, was first introduced in 1997 by Hochreiter and Schmidhuber [27,56] and has been widely used for the analysis and prediction of time series or sequential data, such as: voice recognition, electrocardiograms, stock price, etc. [64]. This network is composed of a memory cell state denoted by (C_t) and three main gates called forget gate (F_t), input gate (I_t), and output gate (O_t), as shown in Figure 2.

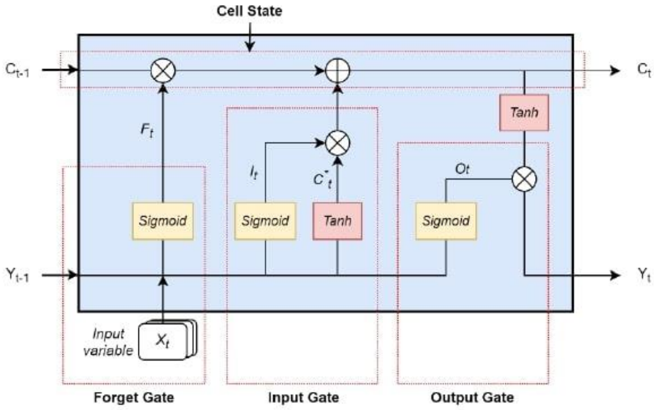


Figure 2. Diagram of the internal structure of an LSTM network.

These gates function as valves to manage and control the information learned during training. More specifically, the forget gate determines what information to remove or keep from the memory cell; the input gate selects the information from the candidate memory cell state (C_t^*) to update the cell state; the output gate decides the information from the memory cell so that the model only considers meaningful information to predict. The values for each of these gates are calculated according to the following equations

$$I_t = \text{sigmoid}(W_i \cdot [y_{t-1}, X_t] + b_i) \tag{1}$$

$$F_t = \text{sigmoid}(W_f \cdot [y_{t-1}, X_t] + b_f) \tag{2}$$

$$O_t = \text{sigmoid}(W_o \cdot [y_{t-1}, X_t] + b_o) \tag{3}$$

$$C_t^* = \tanh(W_c \cdot [y_{t-1}, X_t] + b_c) \tag{4}$$

The memory cell (C_t) and output (y_t) values of the LSTM network are calculated using the following equations

$$C_t = (C_{t-1} \cdot F_t) + (C_t^* \cdot I_t) \tag{5}$$

$$y_t = O_t * \tanh(C_t) \tag{6}$$

where $W_{(i,\,f,\,o,\,C)}$, and $b_{(i,\,f,\,o,\,C)}$ represent the weights and bias variables, respectively, of the three gates and the memory cell state, y_{t-1} is the previous hidden unit information, and X_t is the value of the current input. Thus, the use of these gates allows the LSTM network to have the ability to learn long-term dependencies in the input sequence.

Now, for the development of the LSTM-based model, we consider different data from the building BMS system. Some of these data come from the environment, HVAC system, and building. By supplying the information as a time series to the model, the model will predict the behavior and energy consumption of the building's HVAC system. Therefore, building administrators/managers will obtain information to carry out the aforementioned intelligent strategies.

4. Methodology

In this study, we performed the procedure shown in Figure 3 to train and test the robustness of the LSTM-based DL model for forecasting the short-term (daily) energy consumption of HVAC systems in buildings. The following is a description of each of the steps involved.

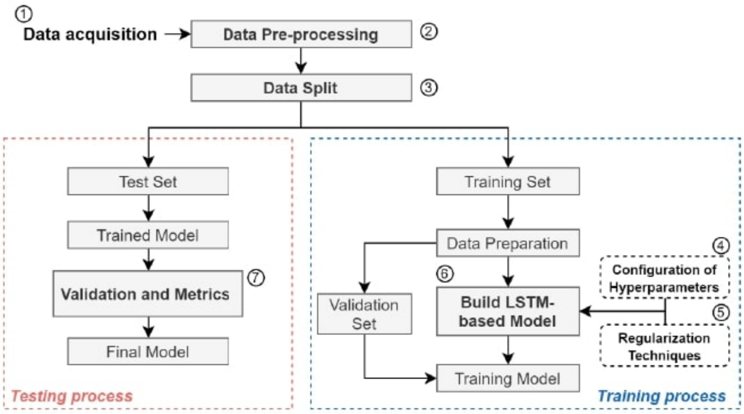


Figure 3. Flow diagram of the model development.

4.1. Data Acquisition

We used a raw dataset from a BMS system to train and test the robustness of the model. This dataset was kindly provided by the Teatro Real of Madrid, which is the most important Spanish opera house.

The Teatro Real is a historic building built in 1850 that has an area of 65,000 m^2 and a capacity for more than 1800 people. It also has several rooms for various activities and uses such as events, rehearsal rooms, studios, technical areas, and a surrounding office area.

On the other hand, this building is in operation every day of the year, with working days of 16 h (8:00 to 00:00). The HVAC system at the Teatro Real is made up of multi-HVAC subsystems [7] incorporating two chiller systems with a nominal cooling power of 350 kW and two heat pump systems with a nominal power of 195 kW for cooling and heating. These multi-HVAC subsystems are monitored and controlled by a BMS that collects data from the different sensors in the building. These data are sampled at different time intervals: one sample every 15 min (96 samples per day) for operating data from multi-HVAC subsystems and one sample every hour (24 samples per day) for internal building data. Some of the variables that the BMS of the Teatro Real keeps stored are the outside and interior temperature of the building and, relative to the multi-HVAC system, the capacity, the energy consumed, and the thermal power generated.

4.2. Data Pre-Processing

For the aforementioned case study, we performed the detection and cleaning of outliers or noisy values, since these greatly affect the precision of the predictive models [65]. For this reason, we used correlation and clustering techniques (k-means) to identify outliers.

On the other hand, the variables used for the model were indicated by the professional experts of the building, which were the following: the interior temperature of the building, the external temperature, and the energy consumed by a HVAC system. We retrieved 6912 data points distributed in each of these variables. Table 1 presents the statistics of them. For this work, the data comes from a heat pump supplied by the expert staff of the

Teatro Real building. These data include the winter season since this season is the one with the highest energy consumption and cost. In addition, these data cover a longer and more uniform period of time in the thermal demand. It should be noted that this case is easily applicable to other seasons if data of the same type are provided.

Table 1. Statistics of the variables.

Stats	Outside Temperature	Indoor Temperature	Heat Pump Power
Mean	8.717577	20.942005	80.989117
Std	3.214003	1.715789	50.924245
Minimum	−0.768667	16.068332	0.48
25%	6.5815	19.617981	26.405333
50%	8.534667	21.256344	94.066675
75%	10.816498	22.388645	117.143975
Maximum	18.72963	23.722167	143.04

In addition, in order to obtain a uniform internal temperature of the building, we considered calculating the average of this temperature recorded in the different areas, rooms, and offices of the Teatro Real in which it was influenced by the heat pump.

As a follow-up to this activity, this new variable (average internal temperature) was sampled per hour, while the other variables were sampled every 15 min. In this sense, we applied a data transformation process (subsampling) for this new variable, since it was crucial for the training of the model [66,67]. Therefore, we transformed the time interval from 1 h to 15 min., using the Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) method [68,69] to equalize the resolution intervals in the Teatro Real data. It should be noted that we performed tests with the data at a 1 h resolution; however, the model lost precision when training the model due to the complexity of HVAC systems [25].

The variables used to develop a DL model are found on different scales (for example °C, kW, etc.); consequently, it was important to standardize them. From this perspective, the method we used to normalize the variables was standard normalization, also known as standard score or z-score [65].

4.3. Data Split

Once the data from the Teatro Real building had been pre-processed, we defined the forecast horizon of the energy consumption of the HVAC system in one day (short-term) since this allows for anticipating the energy that these systems will use; therefore, it will help building managers/administrators to design and implement optimal control strategies over these systems and access optimal energy rates in the continuous electricity market [8,70,71]. Therefore, the one-day forecast horizon for the building is a 96 samples/time step.

Subsequently, we divided the data into three subsets (training-set, validation-set, and testing-set) because the division of these subsets was essential for the accuracy of the model [72]. From this perspective, the pre-processed data were divided into 80% for training and 20% as unseen data to be used in the final evaluation (testing) of the model. It should be noted that 20% of the data from the training-set we used for validation during model training (see Figure 3).

Because the different variables were recorded as a time series, it is of great importance that these data sets were kept in order and preserved the temporal behavior. In addition, we had multivariate data, so it was more difficult to analyze them using classical techniques [73]. Therefore, we applied the sliding window method [72,74–77], which allows transforming time series data into supervised learning. To achieve this, the method maps the past observations (time lags) as input variables. The time lag can be 15 min., one hour, one day, or more than one day. Subsequently, it uses the time steps after this time lagged (one day of samples for the case study) as forecast output, as shown in Figure 4.

It should be noted that the input variables included the three variables presented in Section 4.2. and the output was the daily energy consumption of the HVAC system to be forecasted.

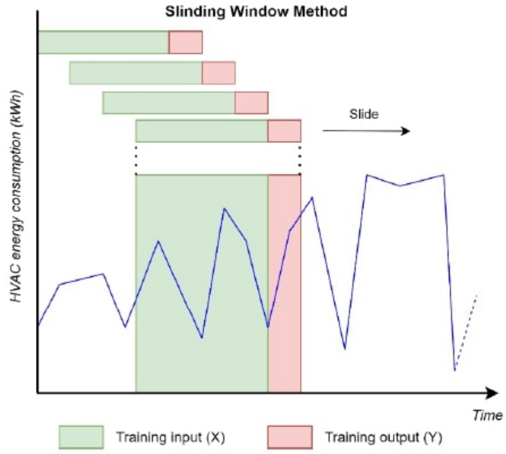


Figure 4. Example of use of sliding window method to transform time series data to supervised learning.

4.4. *Configuration of Hyperparameters*

The tuning of the hyperparameters (HPs) in a DL-based model as an LSTM network has a great impact on the speed, learning capacity, and forecast accuracy of the model [27,78]. For this reason, we evaluated the model by adjusting the most relevant HPs that characterize the structure of the LSTM network, such as the activation function and number of neurons. However, before considering the HPs mentioned above, we will explain some considerations that we made with other HPs required for the training process of the model.

There are different optimizers for learning a neural network, some of which are the stochastic gradient descent (SGD), adaptive gradient algorithm (AdaGrad), Adadelta, root mean square propagation (RMSprop), and adaptive moment estimation (Adam) [29,79,80]. Although all of these optimizers perform for learning DL models, we opted for the Adam optimizer, which, according to the literature, is the most widely used for DL architectures [43] as it offers rapid convergence, a learning rate adaptive, and little tuning of HPs [81]. Similarly, batch size is a hyperparameter that determines the learning performance of the model [39,82]. It should be noted that, from the observation of performance through multiple initial training executions, it was established that the learning rate parameter was set at 0.001, batch size at 2 and the number of iterations or epochs at 50.

4.4.1. Number of Neurons

The number of neurons must be sufficient to provide the LSTM network with learning power. However, this number of neurons is selected by researchers through trial and error. Additionally, the determination of the number of neurons in the LSTM layer will depend on the application and the field in which it is to be used [83]. In this sense, we performed several preliminary tests with a number of neurons between 40 and 80 in the LSTM layer, in which we identified that between 50, 60, and 70 neurons the model performed better. Therefore, we evaluated our model with this number of neurons in the LSTM layer, to find which one best fitted our case study. Furthermore, for the output layer, the number of neurons corresponds to the one-day forecast horizon of our case study (96 neurons).

4.4.2. Activation Functions

The main purpose of the activation function is to introduce nonlinearities to the neural network [84], allowing us to model with greater certainty the nonlinear behavior or pattern typical of real-world data. Therefore, the rectified linear unit (ReLu), leaky ReLu, and exponential linear unit (ELU) activation functions, provide a good level of performance in DL-based models [29,84]. However, we must take into account the strengths and weaknesses of each of these activation functions. Below, Table 2 provides a summary of the strengths and weaknesses of each of these activation functions.

Table 2. Strengths and weaknesses of the ReLu, ELU, Leaky ReLu activation functions.

Activation Function	Strengths	Weaknesses
ReLu	• Fix vanishing gradient problem. • Low computational cost.	• The components are not updated, due to the "Dead ReLu" problem. • It does not avoid the exploding gradient problem. • Transforms negative values linearly.
ELU	• Avoid the "Dead ReLu" problem. • Allow negative outputs. • Fix vanishing gradient problem.	• High computational cost. • It converges slower due to the exponential.
Leaky ReLu	• Avoid the "Dead ReLu" problem. • Lower computational cost than ELU. • Fix vanishing gradient problem.	• It does not avoid the exploding gradient problem. • When differentiating it, it becomes a linear function, causing problems when training.

Considering the above, each of these activation functions avoids the vanishing gradient problem; however, this is not the same with the exploding gradient problem [29,84]. For these reasons, we evaluated the following activation functions on the LSTM network:

- **Hyperbolic Tangent *(TanH)*:** This activation function is shaped like an "S" similar to the sigmoid function. However, unlike the latter, which has an output value of 0 to 1, the *Tanh* has an output value that ranges from -1 to 1. Therefore, it allows the layer output to be normalized around zero when starting the training, helping to accelerate the convergence of the model [29].
- **Scaled exponential linear unit *(SeLu)*:** This was introduced by Günter Klambauer as a variant of the exponential linear unit (ELU) [85]. An advantage of this activation function is that it performs an internal normalization (self-normalized) of the data; that is, the outputs of this function are normally distributed. Therefore, it has fast convergence and solves the problem of gradients vanishing and exploding [29,42,86].

It should be noted that we evaluated these activation functions on the input and output gates of the LSTM network. In this sense, and based on the considerations described above, Table 3 summarizes the different hyperparameters that we evaluated and used for model training.

Table 3. Hyperparameters were evaluated and used to train the model.

Hyperparameters	Values
Activation function for LSTM layer	[SeLu, Tanh]
Number of neurons for LSTM layer	[50,60,70]
Number of neurons for output layer	96
Iterations or Epochs	50
Optimizer	Adam
Learning rate	0.001

4.5. Regularization Techniques

One of the challenges presented by DL-based models is overfitting. This occurs when the model is overparameterized or very complex; consequently, it learns the statistical noise from the training data, resulting in a model that overfits these data excessively. Moreover, when evaluating the model, it does not generalize correctly on the test data [29,78].

To avoid these drawbacks, we considered implementing dropout and early stopping regularization techniques.

4.5.1. Dropout

Dropout is a regularization technique applied during neural network training in a way that ignores or eliminates some neurons randomly [87]. In this way, this technique allows the neural network to be trained as if it presented a different architecture or maintained different internal connections in the layer that implements it. This means that, in each update, the network will be trained with different "forms" of the layer that maintains this regularization. The result of this process is a model that presents a higher robustness and a lower probability of overfitting to the training data [87,88]; consequently, the model is more generalizable and usable.

In addition, this technique is effective when there are limited data for training since, if there are few data, the model will fit these data and will not generalize correctly; on the other hand, if there is a large amount of data, this technique will generate a high computational cost [89].

As a follow-up to this process, we established a dropout rate of 20% since it generated good results when evaluating several executions of the model with the test.

4.5.2. Early Stopping

This technique allows network training to be regularized as soon as the model reaches a minimum threshold of the validation error of the loss function (LF) (training error metric, such as MSE or MAE). In other words, it stops training when the LF of the network increases after waiting for a number of epochs, thus providing good model generalization performance and avoiding network overfitting [29,90].

Taking into account the above, and as indicated in Section 4.3., we took 20% of the training data set as the validation-set, which allowed us to monitor and stop training when the LF did not show an improvement in the validation error. In this context, for our case study, the early stopping technique took the results obtained from LF-validation error and compared them every three epochs if there was a reduction in this error. In case this error increased after passing these three epochs, the training would stop, and the model would be selected three previous epochs before this validation error increased. An example of this can be seen in Figure 5.

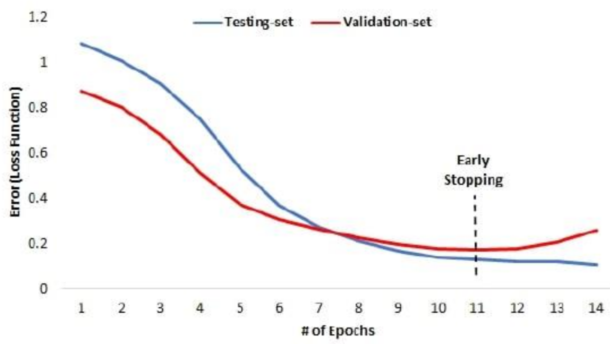


Figure 5. Example of the early stopping regularization technique.

4.6. Build LSTM-Based Model

The architecture of the model proposed in this study for the problem of forecasting the daily energy consumption of the HVAC system in buildings is composed as follows: first, an LSTM input layer that receives all the temporal variables presented in Section 4.2. and which will be evaluated with different HPs settings (see Table 3). Second, an intermediate dropout layer will be implemented to prevent overfitting during the training. Finally, the data from this dropout layer will pass to an output layer made up of a fully-connected layer or dense layer, which will make the forecast for the next day (as shown in Figure 6). It should be noted that the weight and bias matrices were updated (learned) at each iteration during the training.

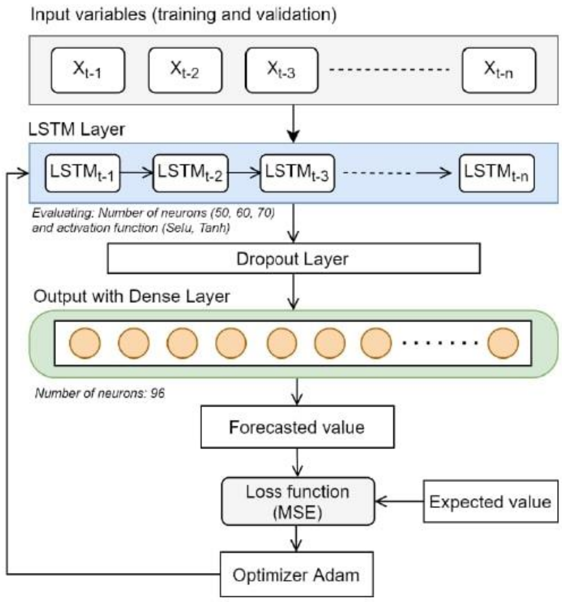


Figure 6. Architecture proposal and training flow of the model for forecasting the daily energy consumption of the HVAC system in buildings.

Once we obtained the forecast result for the day, the actual values registered by the building BMS are compared by calculating the LF. Subsequently, the error obtained by the LF will go through Adam's optimizer during training, which will be in charge of feedback and updating the weights of the network to minimize the LF in each epoch. Similarly, the early stopping technique will be monitored every three epochs with the validation data so that the network does not have overfitting. Therefore, the validation data follow the same architecture flow of the proposed model (see Figure 6).

4.7. Validation and Metrics

Standard DL model validation techniques such as train-test division or k-fold cross validation are not useful for evaluating and validating time series. They ignore the temporal behavior of this type of problem. For this reason, to validate our proposed model we used the Walk-Forward Validation (WFV) technique [91,92].

Taking into account the above, WFV takes the last training-set as time lagged to make a forecast; thus, it compares the output of this model with the unseen data (testing-set). Moreover, this technique allows taking different time lags to evaluate in the model; furthermore, it works well in conjunction with the sliding window method.

One of the most commonly used metrics to evaluate the performance of a forecast is MAPE since it is easy to understand and explain. However, the use of this indicator is restricted in the case of having null or close to zero measured values, which would provide undefined or very extreme values [8]. For this reason, to compare the forecasting performance of our model, we used the following metrics: coefficient of determination (R^2), root mean square error (RMSE), and coefficient of variation of root mean square error (CVRMSE). It should be noted that the CVRMSE metric is especially used in this research, since the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE), the International Performance Measurement and Verification Protocol (IPMVP), and the Federal Energy Management Program (FEMP) established this indicator as a metric of the goodness of fit of a mathematical model with respect to the data or reference estimates measured in different operating conditions of an HVAC system [93–96]. Specifically, if the CVRMSE forecast error of the model is less than 0.30 it is considered valid and adequate for engineering purposes. Next, we define the equations for the error metrics

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{m}(\hat{y}_i - y_i)^2}{m}} \tag{7}$$

$$\mathrm{CVRMSE} = \left(\frac{\mathrm{RMSE}}{\overline{y}}\right) * 100\% \tag{8}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{m}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{m}(y_t - \overline{y})^2} \tag{9}$$

where m is the number of samples, y_t is the expected value of the i-th point, $\overline{y}$ is the mean of the expected value, and $\hat{y}_i$ is the predicted value.

5. Evaluations

The proposed architecture of the LSTM-based model was developed and implemented under the Python language (v3.6) using the Keras Application Programming Interface (API) (v2.4) from the open source TensorFlow library [29,97]. Additionally, the hardware in which the different tests were executed was carried out in an NVIDIA Jetson Nano Developer Kit [98].

On the other hand, the forecasts obtained were compared using different configurations that we will describe below. The architecture configuration C1 represents our proposed model applying the two regularization techniques dropout and early stopping. Configuration C2 represents the model that only applies dropout. Configuration C3 represents the model that only applies early stopping. Finally, configuration C4 represents the model without these regularization techniques.

Similarly, for all these configurations, we evaluated the different hyperparameters (HPs) shown in Table 3, where each of these HPs is labeled as follows: (S50, S60, S70) represent the *SeLu* activation function with 50, 60 and, 70 neurons, respectively, and (T50, T60, T70) represent the *Tanh* activation function with 50, 60 and, 70 neurons, respectively. In addition, we used various time lags (one to seven days), to identify the best time lag setting for the forecast of the daily (short-term) energy consumption of the building's heat pump.

We then executed each of these configurations five times for each time lag, due to the randomness of the DNNs. Tables 4–7 present the performance of the metrics evaluated in our study for the time lags analyzed in the different configurations of the model and HPs mentioned above. Each metric obtained is the average calculated from the executions performed.

Each of the tables presented above is divided as follows: time lags, performance metrics, and the nomenclature given for the HP sets studied. Subsequently, we analyzed the performance obtained in each of the tables and identified the best time lags, activation function, and number of neurons for each configuration, according to the evaluation metrics analyzed. The best performance obtained in Tables 4–7 are highlighted in bold.

From this perspective, all configurations achieved good performance with time lags of seven days. Furthermore, the C1, C3, and C4 configurations performed better with a *Tanh* activation function of 50 neurons (T50) in the LSTM layer. C2 achieved good results with a *Tanh* activation function of 60 neurons (T60) in the LSTM layer.

In the same way, we carried out an in-depth analysis of the previous evidence, in which we analyzed and evaluated each configuration obtained with respect to its epochs, LF of the training-set, and LF of the validation-set, to identify the overfitting in these model configurations. In this sense, we named each configuration C1-T50, C2-T60, C3-T50, and C4-T50. Figure 7 shows each of the configurations with their values obtained for the LF, where the bar chart represents the average error of the LF (training-set and validation-set) and the red line represents the average number of epochs in which the training was completed. It should be noted that these results are the average of the five runs performed.

Table 4. Performance in the metrics evaluated when forecasting the energy consumption of the HVAC, using different time lags and HPs, applying the C1 configuration.

Time Lags	Evaluation Metrics	S50	S60	S70	T50	T60	T70
1	R^2 (%)	0.817	0.818	0.818	0.826	0.825	0.824
	RMSE (kWh)	22.61	22.53	22.55	22.05	22.09	22.14
	CVRMSE (%)	0.249	0.248	0.248	0.243	0.243	0.244
2	R^2 (%)	0.829	0.826	0.821	0.832	0.830	0.827
	RMSE (kWh)	21.86	22.06	22.34	21.64	21.76	21.96
	CVRMSE (%)	0.240	0.243	0.246	0.238	0.239	0.242
3	R^2 (%)	0.841	0.837	0.843	0.847	0.846	0.849
	RMSE (kWh)	21.07	21.33	20.96	20.69	20.71	20.55
	CVRMSE (%)	0.232	0.235	0.231	0.228	0.228	0.226
4	R^2 (%)	0.846	0.854	0.836	0.862	0.860	0.864
	RMSE (kWh)	20.75	20.20	21.37	19.65	19.77	19.52
	CVRMSE (%)	0.228	0.222	0.235	0.216	0.217	0.215
5	R^2 (%)	0.849	0.858	0.852	0.871	0.874	0.869
	RMSE (kWh)	20.56	19.91	20.25	18.99	18.74	19.10
	CVRMSE (%)	0.226	0.219	0.223	0.209	0.206	0.210
6	R^2 (%)	0.860	0.861	0.856	0.874	0.874	0.875
	RMSE (kWh)	19.75	19.72	20.07	18.75	18.73	18.68
	CVRMSE (%)	0.217	0.217	0.221	0.206	0.206	0.205
7	R^2 (%)	0.858	0.852	0.840	**0.876**	0.874	0.876
	RMSE (kWh)	19.94	20.30	21.06	**18.61**	18.74	18.64
	CVRMSE (%)	0.219	0.223	0.232	**0.205**	0.206	0.205

The best performance obtained was highlighted in bold.

Table 5. Performance in the metrics evaluated when forecasting the energy consumption of the HVAC, using different time lags and HPs, applying the C2 configuration.

Time Lags	Evaluation Metrics	S50	S60	S70	T50	T60	T70
1	R^2 (%)	0.800	0.793	0.805	0.813	0.820	0.813
	RMSE (kWh)	23.60	23.92	23.30	22.82	22.39	22.85
	CVRMSE (%)	0.260	0.263	0.256	0.251	0.246	0.251
2	R^2 (%)	0.736	0.807	0.792	0.804	0.806	0.806
	RMSE (kWh)	26.84	23.14	24.03	23.39	23.25	23.22
	CVRMSE (%)	0.295	0.255	0.264	0.257	0.256	0.255
3	R^2 (%)	0.817	0.817	0.832	0.826	0.840	0.823
	RMSE (kWh)	22.60	22.56	21.65	22.06	21.15	22.23
	CVRMSE (%)	0.249	0.248	0.238	0.243	0.233	0.245
4	R^2 (%)	0.829	0.835	0.844	0.858	0.855	0.849
	RMSE (kWh)	21.72	21.44	20.84	19.88	20.12	20.55
	CVRMSE (%)	0.239	0.236	0.229	0.219	0.221	0.226
5	R^2 (%)	0.857	0.850	0.840	0.866	0.863	0.862
	RMSE (kWh)	19.97	20.49	21.07	19.37	19.54	19.65
	CVRMSE (%)	0.220	0.225	0.232	0.213	0.215	0.216
6	R^2 (%)	0.850	0.856	0.857	0.873	0.874	0.875
	RMSE (kWh)	20.37	20.01	20.01	18.82	18.76	18.71
	CVRMSE (%)	0.224	0.220	0.220	0.207	0.206	0.206
7	R^2 (%)	0.862	0.785	0.847	0.866	**0.877**	0.872
	RMSE (kWh)	19.61	23.82	20.68	19.36	**18.56**	18.89
	CVRMSE (%)	0.216	0.262	0.227	0.213	**0.204**	0.208

The best performance obtained was highlighted in bold.

Table 6. Performance in the metrics evaluated when forecasting the energy consumption of the HVAC, using different time lags and HPs, applying the C3 configuration.

Time Lags	Evaluation Metrics	S50	S60	S70	T50	T60	T70
1	R^2 (%)	0.819	0.821	0.823	0.832	0.832	0.832
	RMSE (kWh)	22.48	22.35	22.23	21.69	21.67	21.68
	CVRMSE (%)	0.247	0.246	0.244	0.239	0.238	0.238
2	R^2 (%)	0.824	0.831	0.825	0.838	0.838	0.837
	RMSE (kWh)	22.20	21.71	22.09	21.28	21.27	21.32
	CVRMSE (%)	0.244	0.239	0.243	0.234	0.234	0.235
3	R^2 (%)	0.849	0.852	0.848	0.855	0.853	0.858
	RMSE (kWh)	20.57	20.32	20.62	20.15	20.26	19.92
	CVRMSE (%)	0.226	0.224	0.227	0.222	0.223	0.219
4	R^2 (%)	0.855	0.859	0.842	0.865	0.868	0.869
	RMSE (kWh)	20.13	19.83	20.96	19.39	19.17	19.15
	CVRMSE (%)	0.221	0.218	0.230	0.213	0.211	0.211
5	R^2 (%)	0.860	0.781	0.872	0.882	0.880	0.879
	RMSE (kWh)	19.76	23.33	18.93	18.19	18.28	18.36
	CVRMSE (%)	0.217	0.257	0.208	0.200	0.201	0.202
6	R^2 (%)	0.863	0.871	0.860	0.882	0.882	0.877
	RMSE (kWh)	19.51	19.01	19.72	18.16	18.16	18.53
	CVRMSE (%)	0.215	0.209	0.217	0.200	0.200	0.204
7	R^2 (%)	0.724	0.873	0.869	**0.886**	0.883	0.885
	RMSE (kWh)	24.85	18.83	19.09	**17.88**	18.10	17.89
	CVRMSE (%)	0.273	0.207	0.210	**0.197**	0.199	0.197

The best performance obtained was highlighted in bold.

Table 7. Performance in the metrics evaluated when forecasting the energy consumption of the HVAC, using different time lags and HPs, applying the C4 configuration.

Time Lags	Evaluation Metrics	S50	S60	S70	T50	T60	T70
1	R^2 (%)	0.757	0.766	0.763	0.828	0.817	0.805
	RMSE (kWh)	25.90	25.38	25.63	21.94	22.59	23.28
	CVRMSE (%)	0.285	0.279	0.282	0.241	0.248	0.256
2	R^2 (%)	0.812	0.790	0.787	0.814	0.828	0.792
	RMSE (kWh)	22.93	24.10	24.27	22.77	21.89	23.94
	CVRMSE (%)	0.252	0.265	0.267	0.250	0.241	0.263
3	R^2 (%)	0.759	0.786	0.804	0.823	0.829	0.817
	RMSE (kWh)	25.32	24.32	23.33	22.25	21.86	22.61
	CVRMSE (%)	0.278	0.268	0.257	0.245	0.240	0.249
4	R^2 (%)	0.829	0.787	0.821	0.857	0.853	0.847
	RMSE (kWh)	21.83	24.22	22.30	19.96	20.23	20.65
	CVRMSE (%)	0.240	0.266	0.245	0.219	0.223	0.227
5	R^2 (%)	0.842	0.826	0.829	0.871	0.866	0.858
	RMSE (kWh)	20.95	21.99	21.83	18.99	19.36	19.91
	CVRMSE (%)	0.230	0.242	0.240	0.209	0.213	0.219
6	R^2 (%)	0.803	0.842	0.825	0.878	0.878	0.876
	RMSE (kWh)	22.67	20.94	22.01	18.47	18.48	18.59
	CVRMSE (%)	0.249	0.230	0.242	0.203	0.203	0.204
7	R^2 (%)	0.845	0.775	0.827	**0.879**	0.878	0.872
	RMSE (kWh)	20.77	24.88	21.92	**18.37**	18.49	18.91
	CVRMSE (%)	0.228	0.274	0.241	**0.202**	0.203	0.208

The best performance obtained was highlighted in bold.

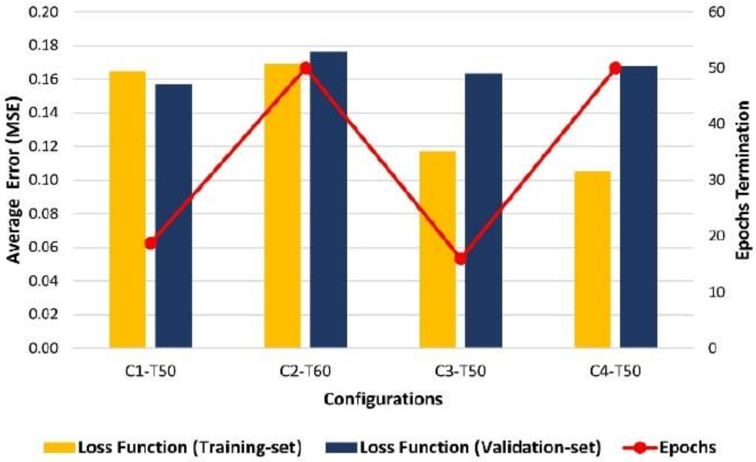


Figure 7. Evaluation of training and validation for the best configurations obtained.

In addition, to verify the robustness and applicability of the configurations of the previously trained models, we compared the performance of the daily forecast provided by the model with respect to unseen data from the testing-set. In Figure 8, some samples are shown with the forecast of the energy consumption of the building's heat pump obtained for one day with these configurations.

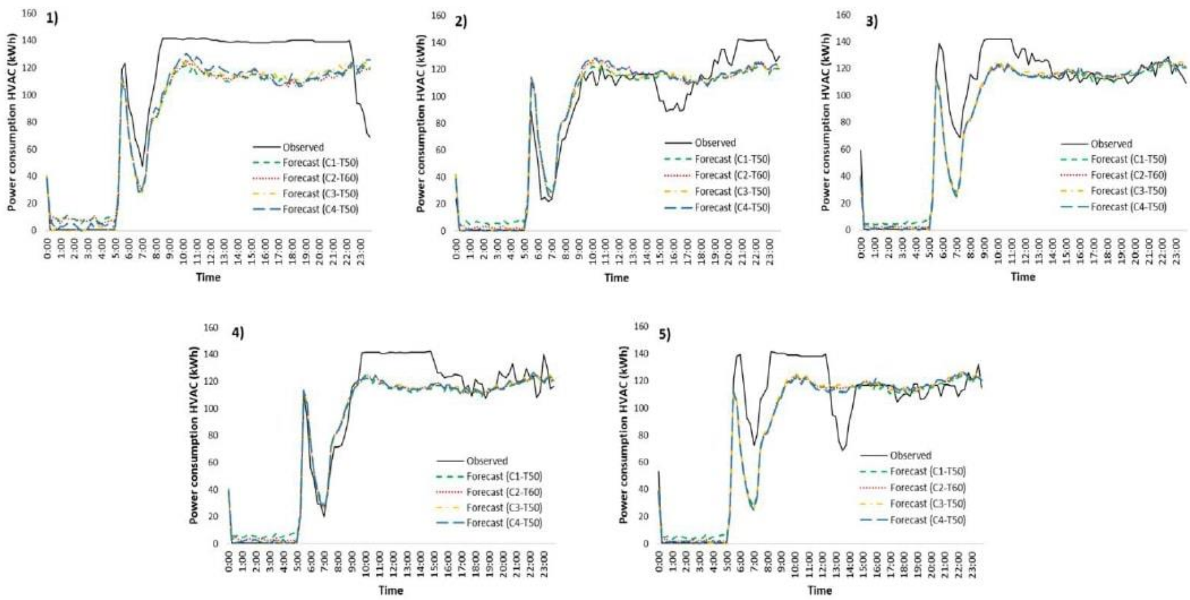


Figure 8. Comparison of the observed values with those predicted by the configurations obtained (C1-T50, C2-T60, C3-T50, C4-T50) for the daily energy consumption of the heat pump.

For these forecasted samples, we have also evaluated and calculated the relative error of the average energy consumption and the relative error of the maximum energy consumption of the building's heat pump. Tables 8–11 show each configuration of the model obtained.

Table 8. Calculation of the relative error of the daily energy consumption (average and maximum) with respect to the testing-set (model: C1-T50).

Forecast Day	Avg. EC * by Day (Y *)	Avg. EC * by Day (Ŷ *)	Relative Error (Avg. EC * by Day)	Max. EC * by Day (Y *)	Max. EC * by Day (Ŷ *)	Relative Error (Max. EC * by Day)
1	99.24	84.44	14.92%	141.600	124.988	11.73%
2	82.15	84.82	3.25%	142.075	124.824	12.14%
3	92.41	84.90	8.12%	142.080	125.786	11.47%
4	89.93	84.76	5.74%	142.400	124.654	12.46%
5	90.87	84.85	6.62%	141.189	125.379	11.20%
Avg.	90.92	84.753	7.73%	141.869	125.126	11.80%

Avg. EC *: Average energy consumption (kW); Max. EC *: Maximum energy consumption (kW); Y *: Expected value; Ŷ *: Forecasted value.

Table 9. Calculation of the relative error of the daily energy consumption (average and maximum) with respect to the testing-set (model: C2-T60).

Forecast Day	Avg. EC * by Day (Y *)	Avg. EC * by Day (Ŷ *)	Relative Error (Avg. EC * by Day)	Max. EC * by Day (Y *)	Max. EC * by Day (Ŷ *)	Relative Error (Max. EC * by Day)
1	99.24	84.39	14.96%	141.60	126.08	10.96%
2	82.15	85.20	3.71%	142.07	130.55	8.11%
3	92.41	85.02	8.00%	142.08	125.93	11.37%
4	89.93	84.85	5.64%	142.40	126.01	11.51%
5	90.87	84.87	6.61%	141.19	125.90	10.83%
Avg.	90.92	84.86	7.78%	141.87	126.90	10.55%

Avg. EC *: Average energy consumption (kW); Max. EC *: Maximum energy consumption (kW); Y *: Expected value; Ŷ *: Forecasted value.

Table 10. Calculation of the relative error of the daily energy consumption (average and maximum) with respect to the testing-set (model: C3-T50).

Forecast Day	Avg. EC * by Day (Y *)	Avg. EC * by Day (Ŷ *)	Relative Error (Avg. EC * by Day)	Max. EC * by Day (Y *)	Max. EC * by Day (Ŷ *)	Relative Error (Max. EC * by Day)
1	99.24	85.18	14.16%	141.60	132.07	7.58%
2	82.15	85.36	3.91%	142.07	129.79	8.65%
3	92.41	84.75	8.28%	142.08	128.39	9.64%
4	89.93	84.80	5.70%	142.40	127.69	10.33%
5	90.87	84.56	6.94%	141.19	127.79	9.49%
Avg.	90.92	84.93	7.80%	141.87	129.15	9.14%

Avg. EC *: Average energy consumption (kW); Max. EC *: Maximum energy consumption (kW); Y *: Expected value; Ŷ *: Forecasted value.

Table 11. Calculation of the relative error of the daily energy consumption (average and maximum) with respect to the testing-set (model: C4-T50).

Forecast Day	Avg. EC * by Day (Y *)	Avg. EC * by Day (Ŷ *)	Relative Error (Avg. EC * by Day)	Max. EC * by Day (Y *)	Max. EC * by Day (Ŷ *)	Relative Error (Max. EC * by Day)
1	99.24	84.58	14.77%	141.60	127.40	10.03%
2	82.15	84.61	3.00%	142.07	126.88	10.69%
3	92.41	84.86	8.17%	142.08	127.64	10.16%
4	89.93	84.77	5.73%	142.40	127.36	10.56%
5	90.87	84.82	6.65%	141.19	127.81	9.47%
Avg.	90.92	84.73	7.66%	141.87	127.42	10.18%

Avg. EC *: Average energy consumption (kW); Max. EC *: Maximum energy consumption (kW); Y *: Expected value; Ŷ *: Forecasted value.

6. Results and Discussion

This study was developed with the objective of building a DL model based on LSTM that allows forecasting of the daily energy consumption of the HVAC system in buildings. For this, it has been applied specifically to a heat pump in a historic building, the Teatro Real in Spain. Another purpose was to evaluate the best configuration of this model by using different time lags, various HP tuning, and implementation of the dropout and early stopping regularization techniques during the training of the model.

Analyzing the performance obtained, it can be seen in Tables 4–7, that all the configurations showed good precision in forecasting the short-term energy consumption of the building's heat pump, given that they obtained values within the range of 0.197–0.285 of CVRMSE, the metric used by ASHRAE, IPMVP, and FEMP to validate the accuracy of a forecast model [93–96]. Although the results show that any configuration of the proposed model can be used, this does not imply that the models generalize correctly. For this reason, we analyzed and evaluated all the configurations that provided the best results in all the accuracy evaluation metrics (see Equations (7)–(9)), to identify and obtain the model with the best values of time lags, HPs tuning, and goodness of the regularization techniques.

Taking into account these considerations, as seen in Tables 4–7, all configurations for a time lag of seven days obtained good results according to the evaluation consideration. Furthermore, for the C1, C3, and C4 configurations, better results were obtained with a *Tanh* activation function of 50 neurons (T50). C1 proposed configuration obtained R^2, RMSE and CVRMSE values of 0.876, 18.61 and 0.205, respectively; C3 configuration obtained R^2, RMSE, and CVRMSE values of 0.886, 17.88 and 0.197, respectively; C4 configuration obtained R^2, RMSE, and CVRMSE values of 0.879, 18.37, and 0.202, respectively. Finally, C2 configuration obtained better results with a *Tanh* activation function of 60 neurons (T60), in which it achieved R^2, RMSE, and CVRMSE values of 0.877, 18.56, and 0.204, respectively.

From the results shown above we can highlight several interesting findings. First, to perform a daily forecast in this case study, it was necessary to apply seven-day historical time lags. The HVAC system of a house only uses the historical behavior of the previous day to forecast the next day as there are few temporal variations throughout a week [4]. In contrast, a complex building such as the one studied here presents many disturbances during the week (e.g., occupant behavior, building use, work time, theatrical performances, events, etc.). The use of seven days of history offers the opportunity to capture these disturbances, making it feasible to anticipate the energy consumption of the HVAC system for the next day with adequate accuracy and favors the implementation of applications such as optimal control, preventive maintenance and fault detection, and diagnosis [57–59]. Regarding the second finding, the configurations obtained (C1-C4) achieved a good performance in the CVRMSE metric, which was between 0.095 and 0.103, less than that proposed of 0.30 by international agencies.

Other significant findings of the results obtained were the activation function and the number of neurons for the LSTM layer. In the first finding, the best results obtained using the activation function *Tanh* versus *SeLu* are highlighted. One justification for this is that, while *SeLu* has the ability to avoid vanishing or exploding gradient problems, it is not able to hold the gradient for a long period of time before it reaches zero, a property that *Tanh* can solve by calculating its second derivative [29,84,99]. In the second finding, the number of neurons evaluated in this study for the LSTM layer was 50, 60, and 70 neurons, in which three model configurations, C1, C3, and C4, performed well with only 50 neurons, whereas the C2 configuration performed better with 60 neurons.

From the previous results, we have taken the best set of configurations, including the activation function and number of neurons obtained, defined as follows: C1-T50, C2-T60, C3-T50, and C4-T50. To understand and justify the difference in the number of neurons required for the LSTM layer, we proceeded to analyze the results of the loss function (LF) and the number of epochs used by these configurations with the training-set and validation-set data.

Taking into account the above and analyzing the results observed in Figure 7, the C1-T50 configuration showed the best results with training and validation sets since the use of regularization techniques for dropout and early stopping prevented it from presenting any tendency of overfitting in LF-training and LF-validation. It is also appreciated that an average of less than 20 epochs was required to achieve good results, so this model has enough computational capacity to capture temporal variations and make a daily forecast of the energy consumption of the building´s heat pump. On the other hand, the C2-T60 configuration, although it employs a greater number of neurons and implements dropout, generated the greatest errors in LF-training and LF-validation. In addition, it had a tendency to overfit since it does not implement early stopping to monitor LF-validation during training, which continued until reaching the established maximum.

There are several possible explanations for these results: (i) the established dropout rate was not optimal for this number of neurons; (ii) the influence of other selected HPs such as learning rate, epochs, optimizer, etc. From this perspective, the use of optimization techniques focused on these HPs could improve the model accuracy [3,47,100]. Meanwhile, C3-T50 and C4-T50 configurations showed a clear trend to overfit, as LF-training decreased dramatically compared to LF-validation. On the one hand, the C3-T50 configuration, while making use of the early stopping technique, was not enough to prevent overfitting over the training-set. On the other hand, the C4-T50 configuration, by not using regularization techniques, was not able to correctly adapt to the data, causing the model to overfit. Thus, we have observed that using only the early stopping technique or any of the regularization techniques would not guarantee that the model is robust and offers good forecast accuracy in data with high temporal variation (chaotic series), as in the case of complex building HVAC systems. Therefore, the use of the early stopping technique will depend on the data, the size of the data, and the context to which it is directed.

On the other hand, to test the robustness and applicability of all the resulting C1-T50, C2-T60, C3-T50, and C4-T50 configurations, we evaluated each configuration of the model with some samples from the testing-set and obtained the corresponding daily forecasts, to compare the accuracy when making the forecast of a day, as well as determine the relative errors obtained in the forecast of the average and maximum daily energy consumption of the heat pump. In this sense, the results observed in Figure 8 show good accuracy of the configurations in the forecast and in detection of the daily temporal variation of the energy consumption of the heat pump. However, in accordance with the aforementioned results and what is observed in Tables 8–11, the C2-T60, C3-T50, and C4-T50 configurations obtained values of the relative error of the average daily energy consumption of (7.78%, 7.80%, 7.66%) and the maximum daily energy consumption of (10.55%, 9.14%, 10.18%), respectively. The C1-T50 configuration obtained a relative error of daily average energy consumption of 7.73% and daily maximum energy consumption of 11.80%.

According to the previous results, the relative errors of the energy consumption (average and maximum) of the C2-T60, C3-T50, and C4-T50 configurations were quite low; however, this can be misleading because, by implementing one or none of the regularization techniques, these configurations cannot reliably generalize over the test-sets, as they exhibit overfitting. For the C1-T50 configuration, although the relative errors are not so low, the model could generalize better to the data since the implemented regularization techniques allowed capturing of the temporal variations.

7. Conclusions and Future Works

In this study, we developed an LSTM-based model aimed at forecasting the daily energy consumption of the HVAC system in buildings, specifically a heat pump at the Teatro Real in Spain. In particular, we focused on determining the time lags that best suits the need. In addition to identifying the best tuning of hyperparameters (HPs) for the LSTM layer, we analyzed the implementation of dropout and early stopping regularization techniques during learning the proposed model.

From this point of view, we have compared different model configurations with respect to the actual data provided by the building's BMS, in which each model configuration implemented one, two, or none of the above regularization techniques. Subsequently, we evaluated, for all these model configurations, multiple executions of several time lags (one to seven days), with HPs as the activation function (*SeLu* or *Tanh*) and a number of 50, 60, and 70 neurons. As a result, we analyzed and evaluated the forecast accuracy of each model configuration with respect to the various results obtained in the evaluation metrics (R^2, RMSE, CVRMSE).

From the experiments conducted, we identified that the CVRMSE measurement results obtained by the model configurations were within the acceptable accuracy range (<0.30) for the global uncertainty in the prediction of energy use according to the criteria indicated by the ASHRAE, IPMVP, and FEMP guidelines to validate an HVAC calibrated model [93–96]. Additionally, for all configurations, we determined that the best time lag to make a daily forecast is the previous seven days. This suggests that models aimed at forecasting the energy consumption of the HVAC system should consider a wide time lag to capture all the patterns or temporal variations of these systems and the building [5,18,26,27], so it could be a baseline for researchers studying similar complex cases.

In addition, we determined the HPs of the C1 to C4 configurations with the best performance in the metrics. The *Tanh* activation function was more accurate in these configurations compared with the *SeLu* activation function, while the best number of neurons was 50 for the C1-T50, C3-T50, and C4-T50 configurations and 60 neurons for the C2-T60 configuration. Based on this evidence, we consider that the activation functions and the number of neurons will depend on many factors, such as data size and variations, model complexity, HPs, use of regularization techniques, or the context in which it is studied. Therefore, the application of optimization techniques for some of these factors would improve the accuracy for model prediction [3], although this would imply a high computational cost [36,62].

We also observed that dropout and early stopping regularization techniques can work together in analogous scenarios, as they easily capture volatility, nonlinearity, and hidden patterns during DNN training. In fact, we have demonstrated that the configuration of model C1-T50, which implemented these regularization techniques, was able to determine the relative error of the forecast of average and maximum daily energy consumption by 7.73% and 11.80%, respectively. In addition, this model has a lower performance in the CVRMSE metric of 0.095 than that proposed by the ASHRAE, IPMVP, and FEMP guidelines. Therefore, it generates generalizable forecasts and is capable of capturing the temporal variations of a complex building such as the one studied.

Taking into account the above, this HVAC system energy consumption forecast model could be part of a broader framework, in which, through the analysis of the forecast generated, intelligent decisions can be made, such as performing preventive maintenance, fault detection and diagnosis, optimization of operation modes, etc.

On the other hand, it should be noted that the proposed model can be improved by adjusting other HPs, such as the learning rate, optimizer type, number of epochs, dropout rate, and number of delay or "patience" for early stopping. Therefore, it would be interesting to investigate whether there is an improvement in the forecasting accuracy of the model when tuning some of these HPs [4,62].

Finally, in future work, we will compare the performance of this model with respect to other Deep Learning models, under the implementation of an intelligent microservices environment within a cloud-connected ecosystem, in order to improve the BMS architecture and extend its functionalities [32] since microservices allow the decoupling of services that require higher computational power (GPU or parallel computing) [101].

Author Contributions: Conceptualization, J.M.G.-P. and M.V.-L.; Methodology, L.M.-P., H.C.-G., J.M.G.-P. and M.V.-L.; Software, L.M.-P. and H.C.-G.; Validation, L.M.-P., J.M.G.-P., M.V.-L., C.S.d.B. and J.L.C.-S.; Formal analysis, L.M.-P., J.M.G.-P., M.V.-L. and J.L.C.-S.; Investigation, L.M.-P., H.C.-G. and J.M.G.-P.; Data curation, L.M.-P., H.C.-G. and J.M.G.-P.; Writing—original draft preparation, L.M.-P., H.C.-G., J.M.G.-P. and M.V.-L.; Writing—review and editing, L.M.-P., H.C.-G., J.M.G.-P. and M.V.-L.; Visualization, L.M.-P., H.C.-G. and J.L.C.-S.; Supervision, J.M.G.-P. and M.V.-L.; Project administration, J.M.G.-P. and C.S.d.B.; Funding acquisition, J.M.G.-P. and C.S.d.B. All authors have read and agreed to the published version of the manuscript.

Funding: This study forms part of the "Smart Energy" Campus of International Excellence (CIE), promoted by the Universidad de Alcalá and the Universidad Rey Juan Carlos, for the financial support received by granting the authors the Second prize of the 2019 Awards Call for the research project: "Intelligent Management System for Optimizing Energy Consumption in Building Air Conditioning". In addition, this work forms part of the project "Smart Tool for the Design of Residential Projects and Energetically Sustainable Buildings project" funding from the FONDO I + D + I SENACYT, 2018.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing is not applicable to this article.

Acknowledgments: We would like to extend special thanks to the management committee of Teatro Real (Madrid, Spain) who contributed with the dataset of the building HVAC System. In addition, we would like to thank the company "ACIX ENERGIA Y MEDIO AMBIENTE S.L." for his technical-scientific advice on the concepts related to HVAC systems in singular buildings. Finally, we thank the National Secretariat for Science, Technology and Innovation (SENACYT, Panama) and the Institute for the Training and Use of Human Resources (IFARHU), for the doctoral study scholarship awarded within the IFARHU-SENACYT doctoral scholarship program, 2018.

Conflicts of Interest: The authors declare no conflict of interest.

Nomenclature

Acronyms	**Description**
HVAC	Heating, ventilating, and air conditioning
BMS	Building management system
ML	Machine learning
DL	Deep learning
DNN	Deep neural network
LSTM	Long short-term memory
RNN	Recurrent neural network
MLR	Multiple-linear regression
ARIMA	Autoregressive integrated moving average
DT	Decision Tree
RF	Random Forest
SVM	Support vector machine
ANN	Artificial neural network
MLP	Multilayer perceptron
HPs	Hyperparameters
SGD	Stochastic gradient descent
AdaGrad	Adaptive gradient algorithm
RMSprop	Root mean square propagation
Adam	Adaptive moment estimation
ReLu	Rectified linear unit
Leaky ReLu	Leaky rectified linear unit
ELU	Exponential linear unit
SeLu	Scaled exponential linear unit
Tanh	Hyperbolic tangent
LF	Loss function
MSE	Mean squared error

MAE	Mean absolute error
MAPE	Mean average percentage error
WFV	Walk-forward validation
R^2	Coefficient of determination
RMSE	Root mean square error
CVRMSE	Coefficient of variation of root mean square error
ASHRAE	American Society of Heating, Refrigerating and Air-Conditioning Engineers
IPMVP	International Performance Measurement and Verification Protocol
FEMP	Federal Energy Management Program

References

1. Burcin, B.-G.; Ioannis, B.; Omar, E.-A.; Nora, E.-G.; Tarek, M.; Shuai, L. Civil Engineering Grand Challenges: Opportunities for Data Sensing, Information Analysis, and Knowledge Discovery. *J. Comput. Civ. Eng.* **2014**, *28*, 4014013. [CrossRef]
2. European Commission. Energy Performance of Buildings Directive. Available online: https://ec.europa.eu/energy/topics/energy-efficiency/energy-efficient-buildings/energy-performance-buildings-directive_en (accessed on 7 March 2021).
3. Hwang, J.K.; Yun, G.Y.; Lee, S.; Seo, H.; Santamouris, M. Using deep learning approaches with variable selection process to predict the energy performance of a heating and cooling system. *Renew. Energy* **2020**, *149*, 1227–1245. [CrossRef]
4. Sendra-Arranz, R.; Gutiérrez, A. A long short-term memory artificial neural network to predict daily HVAC consumption in buildings. *Energy Build.* **2020**, *216*, 109952. [CrossRef]
5. Yildiz, B.; Bilbao, J.I.; Sproul, A.B. A review and analysis of regression and machine learning models on commercial building electricity load forecasting. *Renew. Sustain. Energy Rev.* **2017**, *73*, 1104–1122. [CrossRef]
6. Kusiak, A.; Xu, G. Modeling and optimization of HVAC systems using a dynamic neural network. *Energy* **2012**, *42*, 241–250. [CrossRef]
7. Aguilar, J.; Garcés-Jiménez, A.; Gallego-Salvador, N.; Gutiérrez de Mesa, J.A.; Gomez-Pulido, J.M.; García-Tejedor, Á.J. Autonomic Management Architecture for Multi-HVAC systems in Smart Buildings. *IEEE Access* **2019**, *7*, 123402–123415. [CrossRef]
8. Sun, Y.; Haghighat, F.; Fung, B.C.M. A review of the-state-of-the-art in data-driven approaches for building energy prediction. *Energy Build.* **2020**, *221*, 110022. [CrossRef]
9. Spandagos, C.; Ng, T.L. Equivalent full-load hours for assessing climate change impact on building cooling and heating energy consumption in large Asian cities. *Appl. Energy* **2017**, *189*, 352–368. [CrossRef]
10. Mocanu, E.; Nguyen, P.H.; Gibescu, M.; Kling, W.L. Deep learning for estimating building energy consumption. *Sustain. Energy Grids Netw.* **2016**, *6*, 91–99. [CrossRef]
11. Kuster, C.; Rezgui, Y.; Mourshed, M. Electrical load forecasting models: A critical systematic review. *Sustain. Cities Soc.* **2017**, *35*, 257–270. [CrossRef]
12. Gonzalez-Romera, E.; Jaramillo-Moran, M.A.; Carmona-Fernandez, D. Monthly Electric Energy Demand Forecasting Based on Trend Extraction. *IEEE Trans. Power Syst.* **2006**, *21*, 1946–1953. [CrossRef]
13. Friedrich, L.; Afshari, A. Short-term Forecasting of the Abu Dhabi Electricity Load Using Multiple Weather Variables. *Energy Procedia* **2015**, *75*, 3014–3026. [CrossRef]
14. Ahmad, T.; Chen, H.; Guo, Y.; Wang, J. A comprehensive overview on the data driven and large scale based approaches for forecasting of building energy demand: A review. *Energy Build.* **2018**, *165*, 301–320. [CrossRef]
15. Mohandes, S.R.; Zhang, X.; Mahdiyar, A. A comprehensive review on the application of artificial neural networks in building energy analysis. *Neurocomputing* **2019**, *340*, 55–75. [CrossRef]
16. Chou, J.-S.; Ngo, N.-T. Time series analytics using sliding window metaheuristic optimization-based machine learning system for identifying building energy consumption patterns. *Appl. Energy* **2016**, *177*, 751–770. [CrossRef]
17. Deb, C.; Zhang, F.; Yang, J.; Lee, S.E.; Shah, K.W. A review on time series forecasting techniques for building energy consumption. *Renew. Sustain. Energy Rev.* **2017**, *74*, 902–924. [CrossRef]
18. Amasyali, K.; El-Gohary, N.M. A review of data-driven building energy consumption prediction studies. *Renew. Sustain. Energy Rev.* **2018**, *81*, 1192–1205. [CrossRef]
19. Bourdeau, M.; Zhai, X.Q.; Nefzaoui, E.; Guo, X.; Chatellier, P. Modeling and forecasting building energy consumption: A review of data-driven techniques. *Sustain. Cities Soc.* **2019**, *48*, 101533. [CrossRef]
20. Liu, T.; Xu, C.; Guo, Y.; Chen, H. A novel deep reinforcement learning based methodology for short-term HVAC system energy consumption prediction. *Int. J. Refrig.* **2019**, *107*, 39–51. [CrossRef]
21. Shao, X.; Pu, C.; Zhang, Y.; Kim, C.S. Domain Fusion CNN-LSTM for Short-Term Power Consumption Forecasting. *IEEE Access* **2020**, *8*, 188352–188362. [CrossRef]
22. Fan, C.; Xiao, F.; Wang, S. Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques. *Appl. Energy* **2014**, *127*, 1–10. [CrossRef]
23. Seyedzadeh, S.; Rahimian, F.P.; Glesk, I.; Roper, M. Machine learning for estimation of building energy consumption and performance: A review. *Vis. Eng.* **2018**, *6*. [CrossRef]
24. Chou, J.-S.; Truong, D.-N. Multistep energy consumption forecasting by metaheuristic optimization of time-series analysis and machine learning. *Int. J. Energy Res.* **2021**, *45*, 4581–4612. [CrossRef]

25. Zhou, C.; Fang, Z.; Xu, X.; Zhang, X.; Ding, Y.; Jiang, X.; Ji, Y. Using long short-term memory networks to predict energy consumption of air-conditioning systems. *Sustain. Cities Soc.* **2020**, *55*, 102000. [CrossRef]
26. Walter, T.; Price, P.N.; Sohn, M.D. Uncertainty estimation improves energy measurement and verification procedures. *Appl. Energy* **2014**, *130*, 230–236. [CrossRef]
27. Somu, N.; Raman M R, G.; Ramamritham, K. A hybrid model for building energy consumption forecasting using long short term memory networks. *Appl. Energy* **2020**, *261*, 114131. [CrossRef]
28. Zhang, C.; Li, J.; Zhao, Y.; Li, T.; Chen, Q.; Zhang, X. A hybrid deep learning-based method for short-term building energy load prediction combined with an interpretation process. *Energy Build.* **2020**, *225*, 110301. [CrossRef]
29. Géron, A. *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*; O'Reilly Media: Newton, MA, USA, 2019; ISBN 1492032611.
30. Proyecto CEI "Energía Inteligente". Available online: https://www.campusenergiainteligente.es/en/ (accessed on 7 April 2021).
31. Mendoza-Pittí, L.; Garcés-Jiménez, A.; Aguilar, J.; Gómez-Pulido, J.M.; Vargas-Lombardo, M. Proposal of Physical Models of Multi-HVAC Systems for Energy Efficiency in Smart Buildings. In Proceedings of the 2019 7th International Engineering, Sciences and Technology Conference (IESTEC), Panama, Panama, 9–11 October 2019; pp. 641–646.
32. Mendoza-Pitti, L.; Calderón-Gómez, H.; Vargas-Lombardo, M.; Gómez-Pulido, J.M.; Castillo-Sequera, J.L. Towards a Service-Oriented Architecture for the Energy Efficiency of Buildings: A Systematic Review. *IEEE Access* **2021**, *9*, 26119–26137. [CrossRef]
33. Aguilar, J.; Garcés-Jiménez, A.; Gómez-Pulido, J.M.; R-Moreno, M.D.; Gutiérrez-de-Mesa, J.-A.; Gallego, N. Autonomic Management of a Building's multi-HVAC System Start-Up. *IEEE Access* **2021**. [CrossRef]
34. Le Cun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
35. Mabrouk, A.; Redondo, R.P.D.; Kayed, M. Deep Learning-Based Sentiment Classification: A Comparative Survey. *IEEE Access* **2020**, *8*, 85616–85638. [CrossRef]
36. Torres, J.F.; Hadjout, D.; Sebaa, A.; Martínez-Álvarez, F.; Troncoso, A. Deep Learning for Time Series Forecasting: A Survey. *Big Data* **2020**, *9*, 3–21. [CrossRef]
37. Elhariri, E.; Taie, S.A. H-Ahead Multivariate microclimate Forecasting System Based on Deep Learning. In Proceedings of the 2019 International Conference on Innovative Trends in Computer Engineering (ITCE), Aswan, Egypt, 2–4 February 2019; pp. 168–173.
38. Chandramitasari, W.; Kurniawan, B.; Fujimura, S. Building Deep Neural Network Model for Short Term Electricity Consumption Forecasting. In Proceedings of the 2018 International Symposium on Advanced Intelligent Informatics (SAIN), Yogyakarta, Indonesia, 29–30 August 2018; pp. 43–48.
39. Hadri, S.; Naitmalek, Y.; Najib, M.; Bakhouya, M.; Fakhri, Y.; Elaroussi, M. A Comparative Study of Predictive Approaches for Load Forecasting in Smart Buildings. In Proceedings of the Procedia Computer Science, Coimbra, Portugal, 4–7 November 2019; Volume 160, pp. 173–180.
40. Kim, T.-Y.; Cho, S.-B. Predicting residential energy consumption using CNN-LSTM neural networks. *Energy* **2019**, *182*, 72–81. [CrossRef]
41. Alden, R.E.; Gong, H.; Ababei, C.; Ionel, D.M. LSTM Forecasts for Smart Home Electricity Usage. In Proceedings of the 2020 9th International Conference on Renewable Energy Research and Application (ICRERA), Glasgow, UK, 27–30 September 2020; pp. 434–438.
42. Moon, J.; Park, S.; Rho, S.; Hwang, E. A comparative analysis of artificial neural network architectures for building energy consumption forecasting. *Int. J. Distrib. Sens. Netw.* **2019**, *15*. [CrossRef]
43. Rahman, A.; Srikumar, V.; Smith, A.D. Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks. *Appl. Energy* **2018**, *212*, 372–385. [CrossRef]
44. Alawadi, S.; Mera, D.; Fernández-Delgado, M.; Alkhabbas, F.; Olsson, C.M.; Davidsson, P. A comparison of machine learning algorithms for forecasting indoor temperature in smart buildings. *Energy Syst.* **2020**. [CrossRef]
45. Kim, Y.; Son, H.; Kim, S. Short term electricity load forecasting for institutional buildings. *Energy Rep.* **2019**, *5*, 1270–1280. [CrossRef]
46. Kuo, P.-H.; Huang, C.-J. A High Precision Artificial Neural Networks Model for Short-Term Energy Load Forecasting. *Energies* **2018**, *11*, 213. [CrossRef]
47. Kumar, S.; Hussain, L.; Banarjee, S.; Reza, M. Energy Load Forecasting using Deep Learning Approach-LSTM and GRU in Spark Cluster. In Proceedings of the 2018 Fifth International Conference on Emerging Applications of Information Technology (EAIT), Kolkata, India, 12–13 January 2018; pp. 1–4.
48. Fan, C.; Wang, J.; Gang, W.; Li, S. Assessment of deep recurrent neural network-based strategies for short-term building energy predictions. *Appl. Energy* **2019**, *236*, 700–710. [CrossRef]
49. Wang, L.; Lee, E.W.M.; Yuen, R.K.K. Novel dynamic forecasting model for building cooling loads combining an artificial neural network and an ensemble approach. *Appl. Energy* **2018**, *228*, 1740–1753. [CrossRef]
50. Roy, S.S.; Samui, P.; Nagtode, I.; Jain, H.; Shivaramakrishnan, V.; Mohammadi-ivatloo, B. Forecasting heating and cooling loads of buildings: A comparative performance analysis. *J. Ambient Intell. Humaniz. Comput.* **2020**, *11*, 1253–1264. [CrossRef]
51. Cho, J.S.; Hu, Z.; Sartipi, M. A/C Load Forecasting Using Deep Learning. In Proceedings of the 2017 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 14–16 December 2017; pp. 1840–1841.
52. Machida, Y.; Honoki, H.; Kawano, H.; Sato, F.; Ishikawa, J. Power Consumption Estimation for Building Air Conditioning Systems Using Recurrent Neural Network. In Proceedings of the 2020 IEEE/SICE International Symposium on System Integration (SII), Honolulu, HI, USA, 12–15 January 2020; pp. 854–861.

53. Ellis, M.J.; Chinde, V. An encoder–decoder LSTM-based EMPC framework applied to a building HVAC system. *Chem. Eng. Res. Des.* **2020**, *160*, 508–520. [CrossRef]
54. Hwang, I.; Cho, H.; Ji, Y.; Kim, H. Estimating Power Consumption of Air-conditioners Using a Sequence-to-sequence Model. In Proceedings of the 2019 IEEE 9th International Conference on Consumer Electronics (ICCE-Berlin), Berlin, Germany, 8–11 September 2019; pp. 295–300.
55. Mtibaa, F.; Nguyen, K.-K.; Azam, M.; Papachristou, A.; Venne, J.-S.; Cheriet, M. LSTM-based indoor air temperature prediction framework for HVAC systems in smart buildings. *Neural Comput. Appl.* **2020**, *32*, 17569–17585. [CrossRef]
56. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
57. Jeong, J.; Hong, T.; Ji, C.; Kim, J.; Lee, M.; Jeong, K.; Koo, C. Development of a prediction model for the cost saving potentials in implementing the building energy efficiency rating certification. *Appl. Energy* **2017**, *189*, 257–270. [CrossRef]
58. Gao, D.; Sun, Y.; Lu, Y. A robust demand response control of commercial buildings for smart grid under load prediction uncertainty. *Energy* **2015**, *93*, 275–283. [CrossRef]
59. Xue, X.; Wang, S.; Sun, Y.; Xiao, F. An interactive building power demand management strategy for facilitating smart grid optimization. *Appl. Energy* **2014**, *116*, 297–310. [CrossRef]
60. Qian, F.; Gao, W.; Yang, Y.; Yu, D. Potential analysis of the transfer learning model in short and medium-term forecasting of building HVAC energy consumption. *Energy* **2020**, *193*, 116724. [CrossRef]
61. Pérez-Lombard, L.; Ortiz, J.; Pout, C. A review on buildings energy consumption information. *Energy Build.* **2008**, *40*, 394–398. [CrossRef]
62. Almalaq, A.; Zhang, J.J. Evolutionary Deep Learning-Based Energy Consumption Prediction for Buildings. *IEEE Access* **2019**, *7*, 1520–1531. [CrossRef]
63. Mellouli, N.; Akerma, M.; Hoang, M.; Leducq, D.; Delahaye, A. Deep Learning Models for Time Series Forecasting of Indoor Temperature and Energy Consumption in a Cold Room. In *Computational Collective Intelligence*; Nguyen, N.T., Chbeir, R., Exposito, E., Aniorté, P., Trawiński, B., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 133–144, ISBN 978-3-030-28374-2.
64. Yu, Y.; Si, X.; Hu, C.; Zhang, J. A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures. *Neural Comput.* **2019**, *31*, 1235–1270. [CrossRef] [PubMed]
65. Zhu, J.; Ge, Z.; Song, Z.; Gao, F. Review and big data perspectives on robust data mining approaches for industrial process modeling with outliers and missing data. *Annu. Rev. Control* **2018**, *46*, 107–133. [CrossRef]
66. Pal, B.; Tarafder, A.K.; Rahman, M.S. Synthetic Samples Generation for Imbalance Class Distribution with LSTM Recurrent Neural Networks. In Proceedings of the International Conference on Computing Advancements; Association for Computing Machinery, New York, NY, USA, 10–12 January 2020; pp. 1–5.
67. Chawla, N.V.; Lazarevic, A.; Hall, L.O.; Bowyer, K.W. SMOTEBoost: Improving Prediction of the Minority Class in Boosting. In *European Conference on Principles of Data Mining and Knowledge Discovery*; Lavrač, N., Gamberger, D., Todorovski, L., Blockeel, H., Eds.; Springer Berlin Heidelberg: Berlin/Heidelberg, Germany, 2003; pp. 107–119.
68. Lepot, M.; Aubin, J.-B.; Clemens, F.H.L.R. Interpolation in Time Series: An Introductive Overview of Existing Methods, Their Performance Criteria and Uncertainty Assessment. *Water* **2017**, *9*, 796. [CrossRef]
69. Fritsch, F.N.; Carlson, R.E. Monotone Piecewise Cubic Interpolation. *SIAM J. Numer. Anal.* **1980**, *17*, 238–246. [CrossRef]
70. Panapongpakorn, T.; Banjerdpongchai, D. Short-Term Load Forecast for Energy Management Systems Using Time Series Analysis and Neural Network Method with Average True Range. In Proceedings of the 2019 First International Symposium on Instrumentation, Control, Artificial Intelligence, and Robotics (ICA-SYMP), Bangkok, Thailand, 16–18 January 2019; pp. 86–89.
71. Chou, J.-S.; Tran, D.-S. Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders. *Energy* **2018**, *165*, 709–726. [CrossRef]
72. Chae, Y.T.; Horesh, R.; Hwang, Y.; Lee, Y.M. Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings. *Energy Build.* **2016**, *111*, 184–194. [CrossRef]
73. Tsay, R.S. *Multivariate Time Series Analysis: With R and Financial Applications*; John Wiley & Sons: Hoboken, NJ, USA, 2013.
74. Vafaeipour, M.; Rahbari, O.; Rosen, M.A.; Fazelpour, F.; Ansarirad, P. Application of sliding window technique for prediction of wind velocity time series. *Int. J. Energy Environ. Eng.* **2014**, *5*, 105. [CrossRef]
75. Paoli, C.; Voyant, C.; Muselli, M.; Nivet, M.-L. Forecasting of preprocessed daily solar radiation time series using neural networks. *Sol. Energy* **2010**, *84*, 2146–2160. [CrossRef]
76. Gasparin, A.; Lukovic, S.; Alippi, C. Deep learning for time series forecasting: The electric load case. *arXiv* **2019**, arXiv:1907.09207.
77. Somu, N.; Raman M R, G.; Ramamritham, K. A deep learning framework for building energy consumption forecast. *Renew. Sustain. Energy Rev.* **2021**, *137*, 110591. [CrossRef]
78. Rätz, M.; Javadi, A.P.; Baranski, M.; Finkbeiner, K.; Müller, D. Automated data-driven modeling of building energy systems via machine learning algorithms. *Energy Build.* **2019**, *202*, 109384. [CrossRef]
79. Choi, D.; Shallue, C.J.; Nado, Z.; Lee, J.; Maddison, C.J.; Dahl, G.E. On empirical comparisons of optimizers for deep learning. *arXiv* **2019**, arXiv:1910.05446.
80. Okewu, E.; Adewole, P.; Sennaike, O. Experimental Comparison of Stochastic Optimizers in Deep Learning. In Proceedings of the Computational Science and Its Applications—ICCSA 2019, Saint Petersburg, Russia, 1–4 July 2019; Misra, S., Gervasi, O., Murgante, B., Stankova, E., Korkhov, V., Torre, C., Rocha, A.M.A.C., Taniar, D., Apduhan, B.O., Tarantino, E., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 704–715.

81. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. In Proceedings of the 3rd International Conference on Learning Representations (ICLR), San Diego, CA, USA, 7–9 May 2015.
82. Abbasimehr, H.; Shabani, M.; Yousefi, M. An optimized model using LSTM network for demand forecasting. *Comput. Ind. Eng.* **2020**, *143*, 106435. [CrossRef]
83. Hu, Y.-L.; Chen, L. A nonlinear hybrid wind speed forecasting model using LSTM network, hysteretic ELM and Differential Evolution algorithm. *Energy Convers. Manag.* **2018**, *173*, 123–142. [CrossRef]
84. Rasamoelina, A.D.; Adjailia, F.; Sinčák, P. A Review of Activation Function for Artificial Neural Network. In Proceedings of the 2020 IEEE 18th World Symposium on Applied Machine Intelligence and Informatics (SAMI), Herlany, Slovakia, 23–25 January 2020; pp. 281–286.
85. Klambauer, G.; Unterthiner, T.; Mayr, A.; Hochreiter, S. Self-normalizing neural networks. In Proceedings of the Advances in Neural InformationProcessing Systems (NIPS), Long Beach, CA, USA, 4–9 December 2017; Curran Associates Inc.: Long Beach, CA, USA; pp. 972–981.
86. Kuan, L.; Yan, Z.; Xin, W.; Yan, C.; Xiangkun, P.; Wenxue, S.; Zhe, J.; Yong, Z.; Nan, X.; Xin, Z. Short-term electricity load forecasting method based on multilayered self-normalizing GRU network. In Proceedings of the 2017 IEEE Conference on Energy Internet and Energy System Integration (EI2), Beijing, China, 26–28 November 2017; pp. 1–5.
87. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
88. Aggarwal, C.C. *Teaching Deep Learners to Generalize BT—Neural Networks and Deep Learning: A Textbook*; Aggarwal, C.C., Ed.; Springer International Publishing: Cham, Switzerland, 2018; pp. 169–216. ISBN 978-3-319-94463-0.
89. Goodfelow, I.; Bengio, Y.; Courville, A. *Deep Learning (Adaptive Computation and Machine Learning Series)*; MIT Press: Cambridge, MA, USA, 2016; Volume 10.
90. Aggarwal, C.C. *An Introduction to Neural Networks BT—Neural Networks and Deep Learning: A Textbook*; Aggarwal, C.C., Ed.; Springer International Publishing: Cham, Switzerland, 2018; pp. 1–52. ISBN 978-3-319-94463-0.
91. Bergmeir, C.; Benítez, J.M. On the use of cross-validation for time series predictor evaluation. *Inf. Sci.* **2012**, *191*, 192–213. [CrossRef]
92. Udeh, K.; Wanik, D.W.; Bassill, N.; Anagnostou, E. Time Series Modeling of Storm Outages with Weather Mesonet Data for Emergency Preparedness and Response. In Proceedings of the 2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), New York, NY, USA, 10–12 October 2019; pp. 499–505.
93. ASHRAE. *ASHRAE Guideline 14-2014—Measurement of Energy, Demand and Water Savings*; American Society of Heating, Refrigeration and Air Conditioning Engineers: Atlanta, GA, USA, 2014.
94. Runge, J.; Zmeureanu, R. Forecasting Energy Use in Buildings Using Artificial Neural Networks: A Review. *Energies* **2019**, *12*, 3254. [CrossRef]
95. ASHRAE. *Ashrae Guideline 14: Measurement of Energy and Demand Savings*; American Society of Heating, Refrigeration and Air Conditioning Engineers: Atlanta, GA, USA, 2002; Volume 35.
96. Ruiz, G.R.; Bandera, C.F. Validation of Calibrated Energy Models: Common Errors. *Energies* **2017**, *10*, 1587. [CrossRef]
97. Chollet, F. Others Keras. 2015. Available online: https://keras.io (accessed on 5 April 2021).
98. Dustin, F. Jetson Nano Brings AI Computing to Everyone. Available online: https://developer.nvidia.com/blog/jetson-nano-ai-computing/ (accessed on 5 April 2021).
99. Sarowar, G.; Naser, M.; Nizamuddin, S.M.; Imtiaz Bin Hamid, N.; Mahmud, A. Enhancing Bengali character recognition process applying heuristics on Neural Network. *Int. J. Comput. Sci. Netw. Secur.* **2009**, *9*, 154.
100. Djenouri, D.; Laidi, R.; Djenouri, Y.; Balasingham, I. Machine Learning for Smart Building Applications: Review and Taxonomy. *ACM Comput. Surv.* **2019**, *52*. [CrossRef]
101. Calderon-Gomez, H.; Mendoza-Pitti, L.; Vargas-Lombardo, M.; Gomez-Pulido, J.M.; Castillo-Sequera, J.L.; Sanz-Moreno, J.; Sencion, G. Telemonitoring System for Infectious Disease Prediction in Elderly People Based on a Novel Microservice Architecture. *IEEE Access* **2020**, *8*, 118340–118354. [CrossRef]

2.4 Publicación 4: Ubiquitous and ambient-assisted living eHealth platforms for Down's syndrome and palliative care in the Republic of Panama: A systematic review

2.4.1 Resumen de la publicación

Este artículo presenta dos plataformas de eHealth que muestran cómo los sistemas sanitarios están migrando de una arquitectura cliente-servidor a un paradigma basado en la web y ubicuo. En primer lugar, utilizando la vida asistida por el entorno y la computación ubicua, donde el trabajo realizado permite mejorar la forma en que se prestan los servicios de cuidados paliativos a los ancianos y a los enfermos terminales, facilitando así el trabajo de los médicos, las enfermeras y otros agentes sanitarios. Además, se aplican métodos de aprendizaje automático y un repositorio de resultados de pacientes centrado en datos y ubicuo con la finalidad de mejorar el proceso de estimación del riesgo de síndrome de Down con predicciones más precisas basadas en los parámetros locales de la paciente embarazada.

2.4.2 Contribuciones significativas de la publicación

En esta cuarta publicación, el doctorando realizó diferentes contribuciones significativas a la comunidad científica: (i) se diseñaron y desarrollaron dos plataformas de eHealth para dotarlas de funcionalidades ubicuas; (ii) las plataformas incorporan servicios avanzados mediante técnicas de aprendizaje automático e interoperabilidad; (iii) las plataformas permiten recolectar de manera ordenada la información propia del cuidado, tratamiento y seguimiento que se le da al paciente, a fin de mejorar la calidad de vida tanto del paciente como la de sus familiares; (iv) se consigue realizar una estimación de la probabilidad de riesgo de que el infante nazca con síndrome de Down; (v) el estudio realizado permite al personal de salud brindar una mejor orientación y tratamiento para el cuidado de los pacientes.

2.4.3 Índices de calidad de la publicación

Autores: J.J., Saldaña Barrios, L. Mendoza**,** E. Pitti & M. Vargas.

Título: Ubiquitous and ambient-assisted living eHealth platforms for Down's syndrome and palliative care in the Republic of Panama: A systematic review

Fecha de publicación: diciembre 2018

Health informatics journal, SAGE PUBLICATIONS INC, 2018

Volumen 24, Número 4

Páginas: 356 - 367

DOI: 10.1177/1460458216671560

Ranking JCR 2018: Health Care Sciences & Services (Q2) y Medical Informatics (Q3)

Factor de impacto (JCR 2018): 2.297

Ranking SJR 2018: Health Informatics (Q2).

Factor de impacto (SJR 2018): 0.62

2.4.4 Publicación

Special Issue: UCAmI-IWAAL-AmIHEALTH conference

Health Informatics Journal

Health Informatics Journal
2018, Vol. 24(4) 356–367

DOI: 10.1177/1460458216671560
journals.sagepub.com/home/jhi
SAGE

Ubiquitous and ambient-assisted living eHealth platforms for Down's syndrome and palliative care in the Republic of Panama: A systematic review

Juan Jose Saldaña Barrios, Luis Mendoza, Edgardo Pitti and Miguel Vargas
Technological University of Panama, Panama

Abstract

In this work, the authors present two eHealth platforms that are examples of how health systems are migrating from client-server architecture to the web-based and ubiquitous paradigm. These two platforms were modeled, designed, developed and implemented with positive results. First, using ambient-assisted living and ubiquitous computing, the authors enhance how palliative care is being provided to the elderly patients and patients with terminal illness, making the work of doctors, nurses and other health actors easier. Second, applying machine learning methods and a data-centered, ubiquitous, patient's results' repository, the authors intent to improve the Down's syndrome risk estimation process with more accurate predictions based on local woman patients' parameters. These two eHealth platforms can improve the quality of life, not only physically but also psychologically, of the patients and their families in the country of Panama.

Keywords

ambient-assisted living, Down's syndrome, eHealth, palliative care, ubiquitous computing

Introduction

As mentioned in Saldaña and Vargas-Lombardo[1] and Tran et al.,[2] in Panama, a lot of medical information are still being record in paper. The information systems related to health are not developed with standards that help to manage the patient's clinical information. Down's syndrome and palliative care (PC) are some examples of these affected areas.

The information provided by the finance and economic ministry[3] shows that 1 in every 100 births presents Down's syndrome, and around 15,000 cases were registered by 2012 in Panama.

Corresponding author:
Juan Jose Saldaña, Software Engineering Department, Technological University of Panama, Chiriqui, Panama.
Email: juan.saldana@utp.ac.pa

Senior population aged 60 years in Panama is increasing because advances in medicine have achieved an increase in life expectancy, but most people also arise likely to have advanced disease, involving prioritization attentions at the end of life. That is why providing control pain in patients with terminal disease takes more importance everyday in the health sector as a humanitarian necessity is not a medical obligation.

Structuring a platform capable of bringing the management of patients receiving PC in Panama is necessary to help achieve the main goal of PC, providing a better quality of life for patients and their families.

To research and improve the Down's syndrome risk estimation process and how PCs are provided in Panama, it is necessary to collect, organize and share information using artificial intelligent methods and ubiquitous computing. In this article, we present two case studies of eHealth platforms designed and developed based on ubiquity, machine learning techniques and interoperability.

The rest of the document is organized as follows: section "Ubiquity" describes what ubiquitous is. Section 3 explains the origins of PC. Section "NB" explains briefly what naive Bayes (NB) method is. Section "PC in Panama" provides information of PC in the country of Panama. Section "Down's syndrome" resumes what is Down's syndrome and the impact in the country of Panama. Section "Ubiquitous eHealth platform design, development and implementation" explains the main point of the analysis, design, development and implementation of both platforms. Section "Conclusion" presents the conclusion.

Ubiquity

Ubiquity is the quality of ubiquitous and it refers to the ability to have presence everywhere. Initially, this term was used as a reference to God who is capable of being everywhere.

Mark Weiser describes in his work "The Computer for the Twenty-First Century"[4] the impact that the communication and information technologies would have on the everyday life of the human being. He developed a program in the late 1980s that he called Ubicomp (Ubiquitous Computing). In this model, the communications' capacity was beyond what was expected at the time, so it opened the next generation of computing with information technology accessible wherever and whenever.

Weiser thought that Ubicomp was opposite to virtual reality because virtual reality puts people on a computer-generated world, while Ubicomp places computers at the service of people in the real world. Based on that, Weiser expected to create an environment where devices regardless of the size and functionality could interconnect and manage information, making it more accessible and consistent with the people's daily activities.

Ubicomp has many areas of research and application, with healthcare being one of them,[5] which gives rise to the term Ubicomp in the area of health or pervasive healthcare. It aims to provide technology services to the health sector of Ubicomp allowing access to information inside and outside the medical facilities.

Ubicomp has become notorious in recent years with several projects. The telemonitoring service offered by telemedicine is the result of one such project, which allows specialists to perform remote and real-time monitoring on older patients or PC patients.[6]

Ubiquitous System Patient Medical Records or SUHPC is another example of ubiquitous projects created based on Ubicomp, which allows to manage the patient record remotely. The information can be accessed in real-time in diverse institutions based on the health information requirements.[7]

NB

NB[8–10] is a probabilistic classifier and a machine learning technique that uses the Bayes' theorem, but at the same time assumes a "naive" strong independence between the variables which are independent of each other. NB is a technique that requires first to learn using a training set of classified data. After one introduces the selected data as a representative sample of the population, a model is created. This model will receive the non-classified data to be analyzed and classify it based on the rules of the model.

The advantages of using NB are as follows:

- It is not complicated to implement.
- It provides accurate results although the sample of data for training is small.

Some of the disadvantages are as follows:

- If the variables to analyze present some dependencies, it would reduce considerably the results of the test.

Using NB, we intent to provide a method to add an extra evaluation layer to the prediction process already presented in Saldaña and Vargas-Lombardo.[1]

PC in Panama

Studies conducted in 2012 indicated that 8 percent of the world population are more than 65 years, and it is estimated that within 20 years this percentage will increase to 20 percent.[11] This increase in older people is due to the great strides we have today in medicine, as it provides improvements in the treatment of various infectious diseases and other innovations. However, this increase in life expectancy involves chronic degenerative diseases in the patient, which also affects the families of the patient.

According to the 2010 census of Panama, adult population aged 60 years and older is about 9.7 percent, and it is estimated that by 2020 this will be around 12.4 percent,[12] indicating that this increase will involve a great impact on the health sector in the country, bringing with it the need to ensure greater emphasis on these people.

These home care services were given the emergence of HOSPES Association for Palliative Care in 1992, this being the first in the country to offer care in home mode. Three years later in 1995, the Program for Palliative Care and Pain Relief was created within the premises of the National Cancer Institute (ION), allowing it to provide the care in outpatient and inpatient modes. In 2003, law 68 arose, which required all health facilities in the country to provide the PC with professionals within their facilities. In the years 2006–2007, the PC was provided inside the country, covering every type of care. Finally, on 21 June 2010 under Resolution 499, the National Palliative Care Program of Panama was created.

The hospice has been providing in Panama for over 20 years, which has evolved over time but still the information is not electronically saved and many times the PC is not applied to the patient when it is necessary. In some cases, the PC arrived the patient's home after the patient had died.

Down's syndrome

Trisomy 21 also known as Down's syndrome is an aneuploidy where the fetus shows a genetic alteration having three chromosome 21.[13] This trisomy is one of the major causes of deficiencies or physical disabilities in children and premature deaths that take place before birth, situation that many mothers are unaware of. This chromosomal disorder causes various physical deformities, hearth defects, organ malformations, mental retardation, thyroid disorder and diseases such as Alzheimer's. Trisomy 21 responsible for Down's syndrome is the most frequent aneuploidy. How is the Down's syndrome currently calculated? As mentioned in Saldaña et al.,[14] screening is a probabilistic technique applied to a population to calculate the risk or probability that the fetus suffers a particular disease.

In the screening, first serum and biochemical markers are established and compared with historical median reference values of the population. When the test results of the patient and the multiple of median (MoM) of the markers have different values, the test is considered positive.

The screening methods for Down's syndrome are performed in the second and first trimesters, the first trimester being the most difficult to execute. One of the main barriers of this test is the lack of sampling data to perform the test.

For the first trimester trisomy 21 screening, it is necessary to perform a more effective detection taking into consideration some of the characteristics of the mother such as her weight, her ethnicity, whether she has diabetes or whether she smokes. These factors could affect the result of the test so it needs to be corrected.

Ubiquitous eHealth platform design, development and implementation

Ubiquitous palliative healthcare platform

The first step was to make a state of the art of PC in Panama. All the requirements, resources, actors, process and current issue were gathered from the specialist, current documents, final user interviews and patients. All this information was analyzed to develop a ubiquitous platform to provide an improved PC to the patients in the country. Figure 1 shows the general use cases and actors of the system that help us understand the context of the ambulatory care.

In order to cover all the steps in which the patients need to receive PC, the treatment has been divided into three types: home care, ambulatory care and hospital care.

- *Home care*. This attention mode is very important because the patient will not be in the health institutions. It lets the patient to share with his family at home in his last stage.
- *Ambulatory care*. This mode has two ways to perform. In the first case, the patient has the ability to attend the institution for care. In the second case, the presence of a person (family or friend) is necessary to ensure the care of the patient. This person is known as the primary caregiver and he needs to receive the ongoing training on how should give care to the sick.
- *Hospital care*. This last method is applied when the patient cannot remain at home or the caregiver no longer has the professional skills to care for the sick. The patient has increased suffering caused by the disease, thus requiring more treatments onerous for each symptom and pain relieving suffering of the patient.

Evolution notes. In PC, it is really important to record the current status of the patient in each stage to evaluate the evolution of the illness with each applied treatment. The evolutions notes are divided into four sections, called SOAP or Subjective, Objective, Assessment Plan:

- *S (Subjective)*. This section records all the information provided by the patient, such as symptoms and pains. The subjective impressions of the specialist are also included.

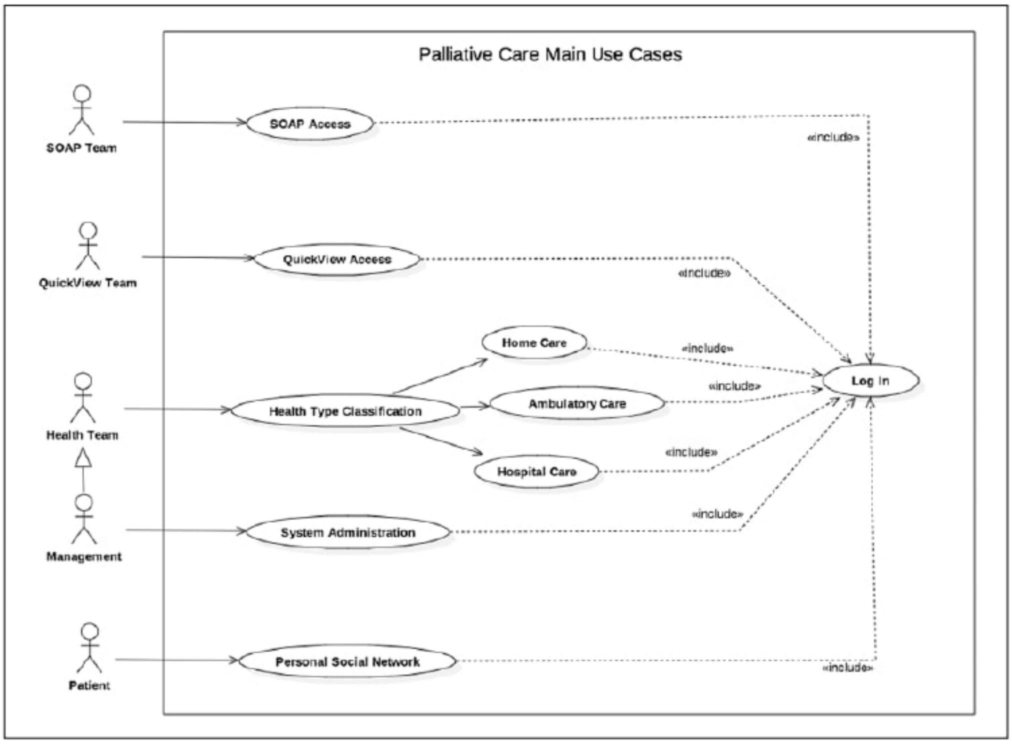


Figure 1. Palliative care use case diagram.

- *O (Objective)*. In this section, the vital signs, physical and complementary examinations of the patients are recorded.
- *A (Assessment)*. In this section, the specialist evaluates the status and its evolution.
- *P (Plan)*. This section modifies the plan applied previously according to the patient's new tests and evaluation.

The main classes of the platforms are presented in Figure 2. The usability was a very important factor in the design of the platform and it was based on goal-oriented design by Allan Cooper.[15,16]

The interaction with the platform was designed based on the usability and specific goal that each type of user will have with the platform.

The system includes more classes, but the intention of the diagram is to show only the classes that are related to the attribute, operations and functional requirements that are specific to the context of PC.

The general architecture of the system can be appreciated in Figure 3 as was proposed in Saldaña and Vargas-Lombardo.[1]

Implementation and testing. For the development, Laravel was used as the development framework, connecting a relational MySQL database to Eloquent ORM and working with Bootstrap to provide a rich user interface.

The platform was tested using white box testing method by two requirement engineers. After all the functionalities were working, the platform was deployed in the cloud of the university and was enabled for the hospital specialist for about 2 months. They used the platform by recording data of

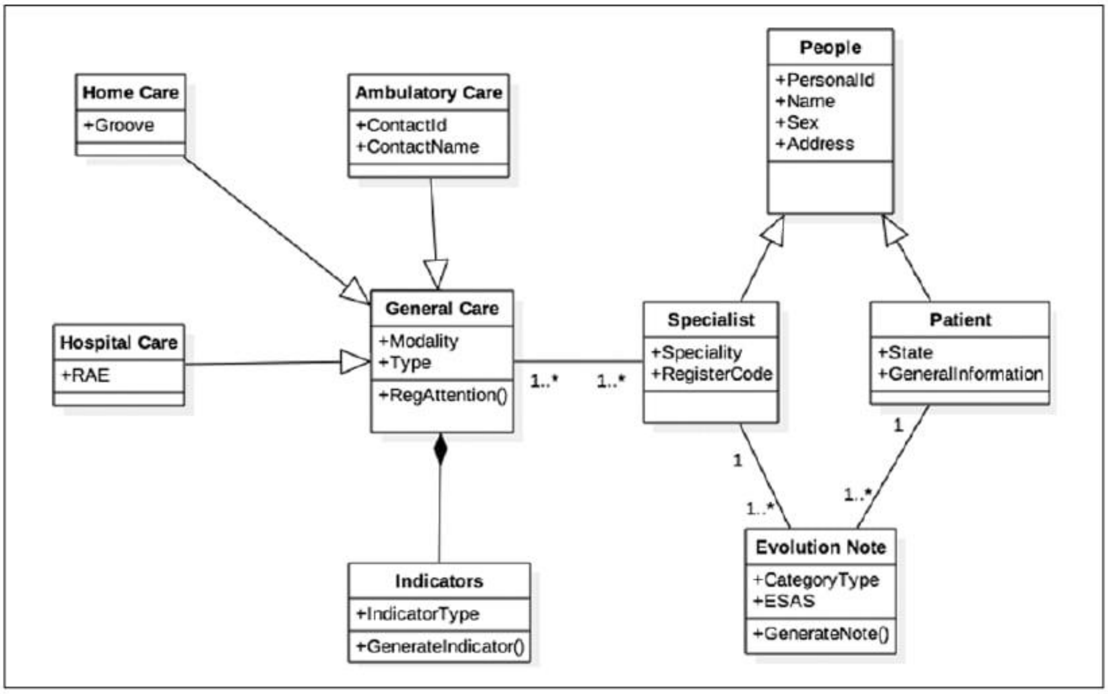


Figure 2. Palliative healthcare class diagram.

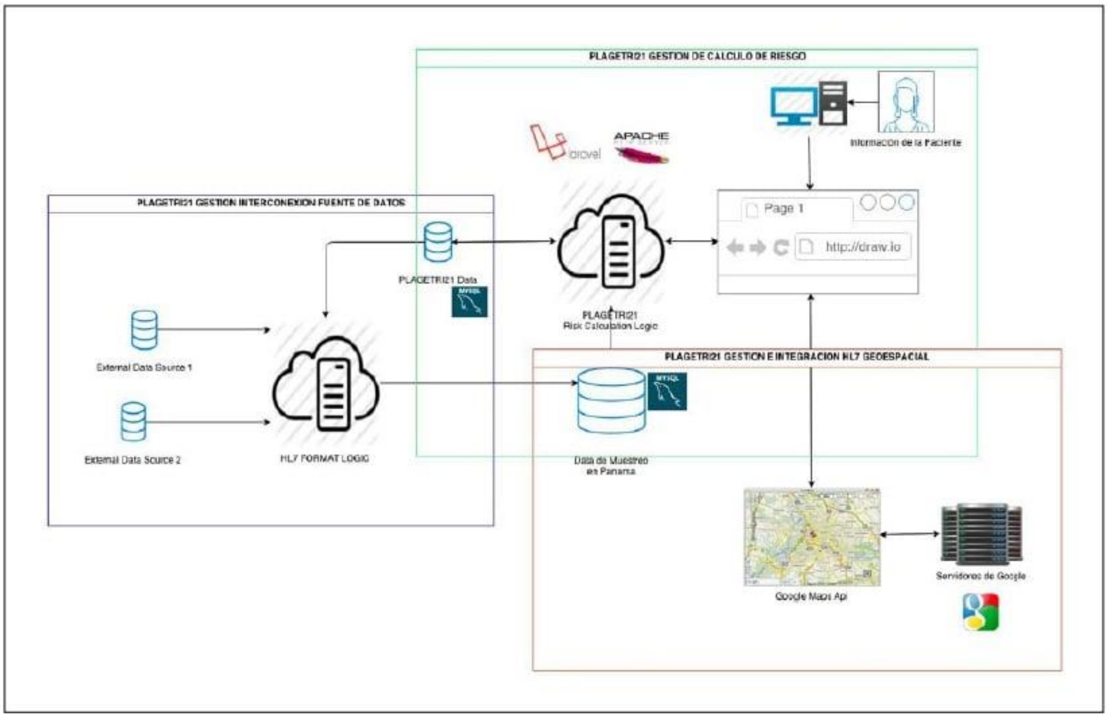


Figure 3. PLAGETRI21 eHealth management platform architecture.

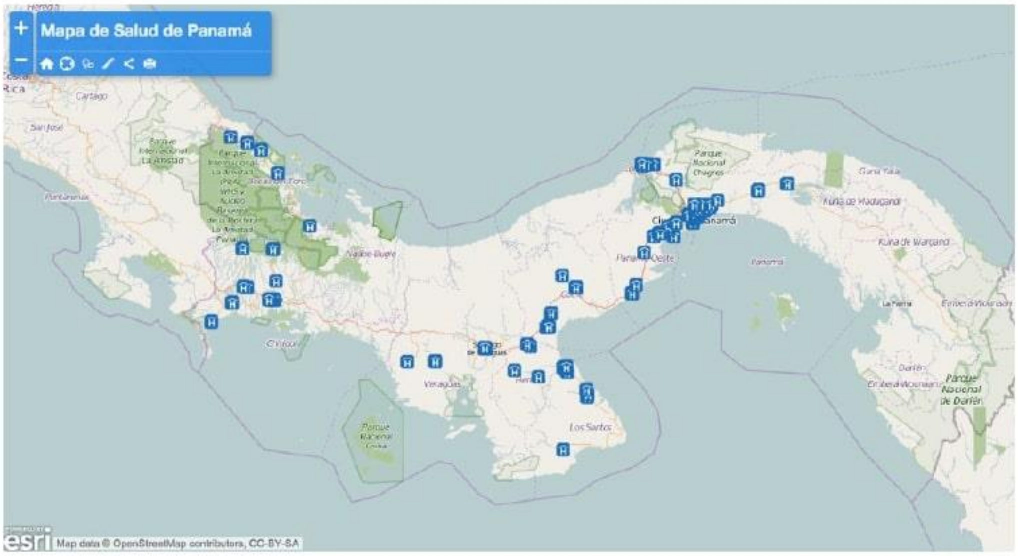


Figure 4. Health institutions that will utilize the eHealth platform for palliative care in the country of Panama.

patients from 1 year ago allowing them to test all the functionalities and provide their feedback. The next step was to sign the term and conditions agreement and deploy the platform to the medical institutions. Figure 4 shows the medical institutions and hospitals of the country that will have access to the platform.

Smart platform for Down's syndrome risk estimation process

The eHealth Management Platform PLAGETRI21[1] is based on the calculation risk method of likelihood, published in Benn,[17] combining the a priori risk for maternal age obtained from the meta-analysis with the likelihood obtained from combining the MoM of the different markers used in each profile. The MoM was calculated using the multivariate normal distribution. This calculation follows a mathematical and statistical process:

- Risk estimation based on the maternal age;
- Markers' standardization;
- MoM calculation;
- Weight and correction factors' adjustments;
- Maternal weight corrections;
- Covariates' corrections;
- Likelihood ratio estimation;
- Risk estimation.

Additional to the normal process, the platform provides two new functionalities. First, it adds an extra layer of analysis applying NB techniques. Using a training set previously selected, the platform generates a model that posteriori receives the not classified data and separates the normal from the abnormal values, predicting whether the test is positive or negative. The following steps are performed to implement the NB classification:

1. Calculating the average $\mu_{F_iC_j}$ of each feature for the classes;
2. Calculation of the variance $\sigma^2_{F_iC_j}$ of the classes;
3. Estimate the probability for each class C_j;
4. Estimate the probability $\sigma^2_{F_iC_j}$ of each feature F_i due to class C_j;
5. Calculate the evidence of the value that it is really the probability of occurrence of all the features $p(F_1, F_2, \ldots, F_i)$;
6. Evaluate the class that presents the higher probability $\hat{p} = \underset{j\in\{1,\ldots,J\}}{argmax}\, p(C_j)\prod_{k=1}^{i} p(F_k \mid C_j)$.

Table 1 shows the implementation of these processes.

Second, the architecture's platform is based on the architecture presented in Barbarito et al.,[18] Esri[19] and Feldmann et al.,[20] which allows the interoperability between many hospital information systems. PLAGETRI21 is able to interact using the HL7 standard with many data source of clinical information. It uses the Clinical Document Architecture (CDA) to save domain-sampling data from diverse sources around the country without the necessity of installing any software at the client side and using laptop and mobile devices. The CDA standard allows the interoperability with other platforms that also implement this standard in their architecture. The message body structured HL7 CDA consists of two parts that are the header and message body as shown in Figure 5.

As an example, to send data about the height and weight of the patient, the tag <entry> is used and is structured as follows:


```
<entry>
<observation classCode="OBS" moodCode="EVN">
<code code="363808001" codeSystem="2.16.840.1.113883.6.96" codeSys-
temName="SNOMED CT" displayName="Peso Corporal"/>
<effectiveTime value="201504071430"/>
<value xsi:type="PQ" value="71.6" unit="kg"/>
</observation>
</entry>
<entry>
<observation classCode="OBS" moodCode="EVN">
<code code="384627007" codeSystem="2.16.840.1.113883.6.96" codeSys-
temName="SNOMED CT" displayName="Estatura"/>
<effectiveTime value="201504071430"/>
<value xsi:type="PQ" value="1.65" unit="m">
</observation>
</entry>
```


The interconnection with other data sources involves security and privacy of the information as presented by Geissbuhler. The data that are extracted from the different data sources include general information of the patient such as age, sex, ethnic, blood type, place of birth and residence. Specific fields such as name, last names and personal ID are not included in the sampling data source. The platform also implements the geospatial interoperability standard proposed in Granell et al.[21] and Ahern.[22] The inclusion of geographical information to the platform enables the clinical information being georeferenced by birth place, residence place and location where the patient receives the medical care, allowing us to research how the location data affect the process.

Table 1. Down's syndrome estimation applying naive Bayes to the data.

	p(Normal| Criterion 1, Criterion 2, Criterion 3)	*p*(Abnormal| Criterion 1, Criterion 2, Criterion 3)	*p*(Criterion 1|Normal)	*p*(Criterion 1|Abnormal)	*p*(Criterion 2|Normal)	*p*(Criterion 2|Abnormal)	*p*(Criterion 3|Normal)	p(Criterion 3|Abnormal)	Evidence
Normal	9.7946E−01	2.0545E−02	1.5860E−01	2.2124E−01	5.2578E−01	1.5711E−01	7.7873E−01	3.7226E−01	5.9983E−02
Abnormal	3.6384E−01	6.3616E−01	2.8865E−01	2.2939E−01	7.1001E−01	1.1676E+00	1.5217E−02	1.9341E−01	7.7554E−03
Normal	7.6056E−01	2.3944E−01	3.7669E−01	2.3326E−01	6.1834E−01	1.0993E+00	1.0290E−01	2.7955E−01	2.8512E−02
Abnormal	8.8520E−08	1.0000E+00	2.8432E−01	2.5526E−01	1.0589E+00	1.2358E+00	2.5672E−11	2.6295E−03	7.8997E−05
Normal	9.5582E−01	4.4182E−02	8.7175E−01	2.4733E−01	1.0241E+00	1.2479E+00	1.7386E−01	2.2083E−01	1.4692E−01
Abnormal	2.2593E−01	7.7407E−01	8.8236E−01	2.6085E−01	9.6253E−01	1.2556E+00	1.4177E−03	1.1965E−01	4.8216E−03
Normal	8.0611E−01	1.9389E−01	1.2441E+00	2.5726E−01	7.8952E−01	3.2329E−01	5.9359E−03	1.6019E−01	6.5442E−03
Normal	8.1843E−01	1.8157E−01	1.7509E−01	2.2254E−01	4.2356E−01	8.9411E−01	3.4393E−01	2.7017E−01	2.8197E−02
Abnormal	1.4100E−01	8.5900E−01	1.1726E−01	2.4911E−01	1.0531E+00	1.2383E+00	7.1843E−03	1.6647E−01	5.6931E−03
Abnormal	4.6286E−01	5.3714E−01	1.1795E+00	2.5465E−01	6.7275E−01	1.1420E+00	5.1809E−03	1.5585E−01	8.0359E−03
Normal	9.1255E−01	8.7448E−02	2.0307E−01	2.2452E−01	1.1375E+00	1.1731E+00	3.3593E−01	2.6823E−01	7.6941E−02
Abnormal	4.2107E−01	5.7893E−01	1.3384E−01	2.1905E−01	8.4016E−02	2.7885E−01	7.1563E−02	1.7207E−01	1.7290E−03
Abnormal	3.2302E−01	6.7698E−01	9.8361E−01	2.4986E−01	1.0582E+00	6.0721E−01	7.7312E−04	1.0560E−01	2.2539E−03
Normal	7.1689E−01	2.8311E−01	9.0362E−01	2.4806E−01	6.7241E−01	1.1418E+00	2.6880E−02	2.1635E−01	2.0613E−02
Abnormal	6.4765E−04	9.9935E−01	2.4634E−05	1.8604E−01	1.1011E+00	6.7738E−01	8.5628E−02	2.7016E−01	3.2445E−03
Abnormal	1.7207E−28	1.0000E+00	8.7490E−28	2.3910E−02	5.0820E−01	9.9355E−01	5.9926E−05	6.1926E−02	1.4010E−04
Normal	9.7666E−01	2.3342E−02	7.0629E−01	2.4339E−01	9.0319E−01	1.2501E+00	7.8031E−01	3.7145E−01	4.6113E−01
Normal	9.5241E−01	4.7592E−02	3.1621E−01	2.3070E−01	1.0963E+00	1.2139E+00	5.2713E−01	3.0977E−01	1.7359E−01
Normal	9.2440E−01	7.5604E−02	3.1600E−01	2.3069E−01	1.0060E+00	1.2518E+00	4.1522E−01	3.5512E−01	1.2919E−01
Normal	9.6669E−01	3.3314E−02	3.3280E−01	2.3144E−01	1.1539E+00	1.1485E+00	7.8171E−01	3.6972E−01	2.8095E−01
Normal	8.2330E−01	1.7670E−01	8.4631E−02	2.1332E−01	1.1584E+00	8.0449E−01	1.9638E−01	2.2872E−01	2.1156E−02
Normal	6.7343E−01	3.2657E−01	2.6284E−01	2.2807E−01	4.1702E−02	1.5791E−01	1.5138E−01	2.1225E−01	2.2292E−03

```
<ClinicalDocument>
      <!-- CDA Header -->
      <!-- CDA Body -->
      <component>
          <structuredBody>
            <component>
               <section>…</section>
               <section>…</section>
            </component>
          </structuredBody>
      </component>
</ClinicalDocument>
```

Figure 5. HL7 CDA structure.

Table 2. Results of the screening test.

Data	Values
Total samples	100
Valid test	98
Test with double verification	6
Valid and approved result tests	92
Doubtful test	4
Acceptance rate for doubtful test	50%–60%
Acceptance test wit 90% of accuracy rate	100%
Acceptance test wit 95% of accuracy rate	92.58%
General acceptance rate	94.3877551

Implementation and testing. The platform was tested initially by the development team and three medicine specialists from the hospital. After all the bugs were fixed, the screening group of the laboratory took a sample data from 100 patients. The chemical tests were introduced and analyzed by the platform and the results can be found in Table 2.

The platform is running on a private cloud inside the Technological University of Panama and the Electronics Health and Supercomputing Research's Group developed it.

Conclusion

The Down's syndrome risk estimation platform enhances the accuracy of the result because it adds an extra layer of data analysis applying machine learning methods to establish smart classifiers extracted from the population sampling and includes valuable geographical information to the procedures, not taken into consideration before.

Thanks to the information that is being captured, organized and shared from diverse sources in the country, the first trimester screening test will be applied allowing to detect any illness earlier in order to provide the treatment in a timely manner.

The PC platform allows, thanks to its ubiquitous properties, to record personal and medical information in real-time even from the patient's home. It also lets specialist of the field to have access to this information and apply adequate medicines and treatment. The platform is a tool that allows the patient's family being in touch with the medical specialist.

Ubicomp, machine learning techniques and ambient-assisted living open to us a broad source of resources that can be applied to improve many areas of healthcare.

Acknowledgements

The Research Group Electronics Health and Supercomputing (GISES) of the Technological University of Panama supported this research as well as Dr Miguel Vargas-Lombardo National Research Investigator award (SNI) from the SENACYT. Both projects are being presented as main topics looking for founds with SENACYT.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

1. Saldaña J and Vargas-Lombardo M. eHealth management platform for screening and prediction of Down's syndrome in the Republic of Panama. *E-Health Telecommun Syst Netw* (Scientific Research Publishing) 2014; 3(3): 33–42, http://www.scirp.org/journal/PaperInformation.aspx?PaperID=49724&#abstract (accessed 27 February 2015).
2. Tran V-A, Johnson N, Redline S, et al. OnWARD: ontology-driven web-based framework for multi-center clinical studies. *J Biomed Inform* 2011; 44 (Suppl. 1): S48–S53, http://www.sciencedirect.com/science/article/pii/S1532046411001468 (accessed 20 February 2015).
3. Domínguez Y. Sindrome de Down Batalla social. *Día a Día*, 2015, http://www.diaadia.com.pa/primerplano/s%C3%ADndrome-de-down-batalla-social-268675 (accessed 1 January 2015).
4. Weiser M. The computer for the 21st century. *IEEE Pervas Comput* 2002; 1(1): 19–25.
5. Pomares ES. *Computación Ubicua: un gran desafío*. Cholula, Mexico: National Institute of Astrophysics, Optics and Electronicsp, 2010, p. 2.
6. Escayola J, Martínez I, Serrano L, et al. Propuesta de una Nueva Arquitectura de Software para uso del Estándar ISO/IEEE 11073 en Dispositivos Médicos de Limitada Capacidad de Procesado y Memoria, 2008, pp. 2–5, http://diec.unizar.es/intranet/articulos/uploads/Propuesta%20de%20una%20Nueva%20Arquitectura%20de%20Software%20para%20uso%20del%20Estandar%20ISO-IEEE%2011073%20en%20Dispositivos%20Medicos%20de%20Limitada%20Capacidad%20de%20Procesado%20y%20Memoria.pdf
7. Rodríguez Robledo G. Sistema Ubicuo de Historia Clínica del Paciente. *Instituto Politécnico Nacional*, 2006, http://www.saber.cic.ipn.mx/cake/SABERsvn/trunk/Repositorios/webVerArchivo/366/1
8. Soria D, Garibaldi JM, Ambrogi F, et al. A "non-parametric" version of the naive Bayes classifier. *Knowledge-Based Syst* 2011; 24(6): 775–84, http://www.sciencedirect.com/science/article/pii/S0950705111000414 (accessed 27 May 2015).
9. Naive Bayes text classification. http://nlp.stanford.edu/IR-book/html/htmledition/naive-bayes-text-classification-1.html (2009, accessed 8 October 2015).
10. Marucci-Wellman HR, Lehto MR and Corns HL. A practical tool for public health surveillance: semi-automated coding of short injury narratives from large administrative databases using naïve Bayes algorithms. *Accident Anal Prev* 2015; 84: 165–76, http://www.sciencedirect.com/science/article/pii/S0001457515300099 (accessed 5 October 2015).
11. Valencia B and Inés M. Revista Colombiana de Anestesiología Envejecimiento de la población. *un reto para la salud pública* 2012; 40(69): 192–194.

12. INEC Contraloría General de la República de Panamá. *Estimaciones y proyecciones de la población en la republica, provincia, comarca indígena por distrito, según sexo y edad; 2010–20*, https://www.contraloria.gob.pa/inec/Publicaciones/Publicaciones.aspx?ID_SUBCATEGORIA=10&ID_PUBLICACION=499&ID_IDIOMA=1&ID_CATEGORIA=3
13. Gyselaers W, Vereecken A, Van Herck E, et al. Screening for trisomy 21 in Flanders: a 10 years review of 40.490 pregnancies screened by maternal serum. *Eur J Obstet Gyn R B* 2004; 115(2): 185–9, http://www.sciencedirect.com/science/article/pii/S0301211503006456 (accessed 26 February 2014).
14. Saldaña BJJ, Rovetto C, Pitti E, et al. Modelado formal de la metodología para la predicción de pacientes con Síndrome de Down en Panamá, 2015, http://www.laccei.org/LACCEI2015-SantoDomingo/RefereedPapers/RP082.pdf
15. Guersenzvaig A. El usuario arquetípico: creación y uso de personajes en el diseño de productos interactivos (Human-computer interact), http://www.alzado.org/imgconts/autor_id3/personajes_alzado2.pdf
16. Honduvilla M, Bernabé Poveda MA and Manrique Sancho MT. La usabilidad de los geoportales: aplicación del Diseño Orientado a Metas (DOM). In: *IV Jornadas Técnicas de las Infraestructuras de Datos Espaciales de España*, Santiago de Compostela, España, 17–19 October 2007.
17. Benn PA. Advances in prenatal screening for Down syndrome: I. general principles and second trimester testing. *Clin Chim Acta* 2002; 323(1–2): 1–16, http://www.sciencedirect.com/science/article/pii/S0009898102001869 (accessed 26 February 2014).
18. Barbarito F, Pinciroli F, Mason J, et al. Implementing standards for the interoperability among healthcare providers in the public regionalized Healthcare Information System of the Lombardy Region. *J Biomed Inform* 2012; 45(4): 736–45, http://www.sciencedirect.com/science/article/pii/S153204641200007X (accessed 20 February 2014).
19. Esri A. HL7 and Spatial interoperability standards for public health and health care delivery, 2011, https://www.esri.com/library/whitepapers/pdfs/hl7-spatial-interoperability.pdf
20. Feldmann RL, Shull F, Denger C, et al. A survey of software engineering techniques in medical device development 2007. In: Joint workshop on high confidence medical devices, software, and systems and medical device plug-and-play interoperability (HCMDSS-MDPnP 2007), Cambridge, MA, 25–27 June, pp. 46–54. New York: IEEE, http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4438163 (accessed 19 July 2013).
21. Granell C, Fernández ÓB and Díaz L. Geospatial information infrastructures to address spatial needs in health: collaboration, challenges and opportunities. *Future Gener Comp Sy* 2014; 31: 213–222, http://www.sciencedirect.com/science/article/pii/S0167739X13000629 (accessed 5 February 2014).
22. Ahern DK. Challenges and opportunities of eHealth research. *Am J Prev Med* 2007; 32(Suppl. 5): S75–S82, http://www.sciencedirect.com/science/article/pii/S0749379707000451 (accessed 26 February 2014).

2.5 Publicación 5: Machine-learning Model to Predict the Intradialytic Hypotension Based on Clinical-Analytical Data

2.5.1 Resumen de la publicación

La hipotensión intradialítica (IDH) es un problema recurrente en pacientes que reciben tratamiento de hemodiálisis (HD). Además, la IDH está relacionada con una gran variedad de factores, por lo que, determinar si un paciente presenta IDH durante el tratamiento de HD es una tarea compleja para los modelos tradicionales (estadísticos). De hecho, los modelos basados en machine-learning (ML) tienen la capacidad para descubrir y aprender los patrones presentes en los datos. En este sentido, este artículo tiene el objetivo de identificar un modelo basado en ML que permita determinar si el paciente presentará o no IDH al inicio del tratamiento de HD. Para ello, se seleccionó una combinación de variables de los registros clínicos y analíticos de sangre del paciente que recibe el tratamiento de HD. Asimismo, se desarrollaron y evaluaron diferentes modelos de ML (Multi-layer perceptron, Logistic Regression, Random Forest y XGBoost), a fin de identificar el que mejor se desempeña realizando esta tarea. Como resultado del estudio y análisis realizado, el modelo XGBoost consiguió un mejor desempeño en las métricas analizadas en comparación al resto de modelos de ML evaluados. Donde, este logró valores del 96.9% y 94.5% para las áreas bajo la curva de ROC y Precision-Recall. En consecuencia, estos resultados sugieren que el modelo XGBoost es capaz de predecir con una alta confiabilidad y robustez la ocurrencia de IDH o no en el paciente y, además, ofrece al personal médico facilidades para el apoyo en la toma de decisiones basadas en la información inferida por el propio modelo.

2.5.2 Contribuciones significativas de la publicación

En esta quinta publicación, el doctorando realizó diferentes contribuciones significativas a la comunidad científica: (i) presenta un modelo de machine learning (ML) que permite determinar si el paciente presentará o no una hipotensión intradialítica (IDH) al inicio de la sesión de hemodiálisis (HD); (ii) se construyó un novedoso conjunto de datos en el que se unifican las variables de dos bases de datos bien diferenciadas (datos clínicos y analíticos); (iii) el modelo obtenido de machine learning consiguió un desempeño superior en diferentes métricas respecto a otros estudios realizados; (iv) este modelo de machine learning es de gran utilidad al momento de tomar una decisión inteligente por parte del personal de salud; (v) proporciona al personal de salud información sobre la identificación de los parámetros clínicos y analíticos más decisivos que afectan a la aparición de IDH durante la sesión de HD.

2.5.3 Índices de calidad de la publicación

Autores: L. Mendoza-Pittí, J. M. Gómez-Pulido, M. Vargas-Lombardo, J. A. Gómez-Pulido, M. -L. Polo-Luque and D. Rodréguez-Puyol

Título: Machine-Learning Model to Predict the Intradialytic Hypotension Based on Clinical-Analytical Data

Fecha de publicación: 07 de julio de 2022

IEEE Access, 2022

Volumen: 10, Páginas: 72065-72079

ISSN: 2169-3536

DOI: 10.1109/ACCESS.2022.3189018

Ranking JCR 2021: Computer science, information systems (Q2)

Factor de impacto (JCR 2021): 3.476

Ranking SJR 2021: Computer Science (Q1)

Factor de impacto (SJR 2021): 0.927

2.5.4 Publicación

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 21 June 2022, accepted 4 July 2022, date of publication 7 July 2022, date of current version 13 July 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3189018

RESEARCH ARTICLE

Machine-Learning Model to Predict the Intradialytic Hypotension Based on Clinical-Analytical Data

LUIS MENDOZA-PITTÍ[1,2], JOSÉ MANUEL GÓMEZ-PULIDO[1], MIGUEL VARGAS-LOMBARDO[2], (Senior Member, IEEE), JUAN A. GÓMEZ-PULIDO[3], (Senior Member, IEEE), MARÍA-LUZ POLO-LUQUE[4], AND DIEGO RODRÍGUEZ-PUYOL[5]

[1] Health Computing and Intelligent Systems Research Group (HCIS), Department of Computer Science, Ramón y Cajal Institute for Health Research (IRYCIS), University of Alcalá, 28805 Alcalá de Henares, Spain
[2] SNI, Senacyt Panama, E-Health and Supercomputing Research Group (GISES), Technological University of Panama, Panama City 0819-07289, Panama
[3] Department of Technologies of Computers and Communications, University of Extremadura, 10003 Cáceres, Spain
[4] Department of Nursing and Physiotherapy, University of Alcalá, 28805 Alcalá de Henares, Spain
[5] Department of Medicine and Medical Specialties, Research Foundation of the University Hospital Príncipe de Asturias, IRYCIS, University of Alcalá, Alcalá de Henares, 28805 Madrid, Spain

Corresponding authors: Miguel Vargas-Lombardo (miguel.vargas@utp.ac.pa) and José Manuel Gómez-Pulido (jose.gomez@uah.es)

This work was supported in part by the clinical team of the Nephrology Section of the Hospital Universitario Príncipe de Asturias, in part by the National Secretariat for Science, Technology and Innovation (SENACYT, Panama) and the Institute for the Training and Use of Human Resources (IFARHU), for the doctoral study scholarship, in part by the SENACYT through the National Research System (SNI-SENACYT) National Research System of which one of the authors is a member, and in part by the Fresenius Medical Care AG company.

ABSTRACT Predicting whether patients will experience intradialytic hypotension (IDH) during hemodialysis (HD) is not an easy task. IDH is associated with multiple risk factors, meaning that traditional statistical models are unable to find the relationships that affect it. In this context, the use of models based on machine learning (ML) can allow the discovery of complex relationships, since they can solve problems without being explicitly programmed. In this work we developed, evaluated and identified an ML-based model that is capable of predicting at the beginning of the HD session whether a patient will suffer from IDH during its prolonged development. To develop the ML models, we used the hold-out and cross-validation methods; while, to evaluate the performance of the models we used the metrics F1-score, Matthews Correlation Coefficient, areas under the receiver operating characteristic (AUROC) and precision-recall curve (AUPRC). In this sense, we selected and used a reduced combination of variables from clinical records and blood analytics, which have proven to be decisive for the occurrence of IDH. The predictive results obtained through our work confirmed that the best ML model was based on the XGBoost model, achieving values of 0.969 and 0.945 for AUROC and AUPRC respectively. Therefore, our study suggests that the XGBoost model has a very high predictive capacity for the appearance of an IDH in HD patients and presents great versatility and flexibility in terms of supporting informed decision-making by medical staff.

INDEX TERMS Clinical-analytical data, hemodialysis, intradialytic hypotension, machine learning, predicting model, XGBoost.

I. INTRODUCTION

Chronic kidney disease (CKD) is the sixth fastest growing cause of death, affecting approximately 11–13% of the population [1]. It is estimated that 850 million people worldwide suffer from kidney disease, and it is responsible for at least 2.4 million deaths per year. Furthermore, the incidence of CKD is expected to increase due to aging of the population, since although it can develop at any age, its prevalence is higher among older adults. One in five men and one in four women aged 65–74 suffer from CKD, and half of people aged 75 or older have CKD to some degree. The prevalence

The associate editor coordinating the review of this manuscript and approving it for publication was Turgay Celik.

of CKD in Spain is approximately 15%, whereas in the United States it is approximately 13%, and it can therefore be observed that these rates are quite similar [2]. Furthermore, Spain has one of the highest rates of CKD compared to other European countries [3].

Dialysis or renal transplantation are treatments for advanced CKD, which occurs in approximately 1% of people with CKD. They are also one of the most expensive chronic disease treatments [2], [4]. In this case, patients may be treated by hemodialysis (HD).

HD is a treatment that replaces two of the kidney's main functions: periodic filtering of the blood and elimination of excess fluids from the body. To perform HD, artificial filters called dialyzers are used in which a large amount of blood is circulated through the filter via vascular access. However, fluid removal is one of the main causes of hypotension in patients [5].

Hypotension due to dialysis, known as intradialytic hypotension (IDH), is one of the most frequent complications in clinical practice. IDH has a prevalence ranging from 5–40% [6], [7]. This large range of values is because there is no clear consensus and uniformity in many investigations on the definition of the onset of IDH [8]–[10]. Nevertheless, several studies have used one or more of the following criteria to define IDH; (i) a threshold/nadir in systolic blood pressure (SBP); (ii) an absolute reduction in SBP; and (iii) a requirement for intervention [9], [10]. Furthermore, identifying the factors that cause the occurrence of IDH is a complex task, since multiple factors are involved, such as the modality of dialysis, patient type or medical criteria [7], [11]. Therefore, identification of the factors that most influence its occurrence would improve decision-making by clinical staff (doctors, nurses, and assistants, among others).

HD is usually carried out in sessions that last approximately four hours and are performed on alternate days. Large amounts of data are generated during these HD sessions, often consisting of different records (e.g., demographic data, laboratory records or medical prescriptions). In this sense, the use of specialized techniques such as big data and artificial intelligence (AI) are of great use in the analysis of variabilities and intrinsic relationships in patient registries with and without IDH.

Recent developments in the field of AI have stimulated its application in daily clinical practice [12], [13], because it is useful for processing massive and complex sets of big data and can facilitate diagnostic and therapeutic decisions, promote medical innovation, and reduce costs, among its other advantages. In view of this, AI-based models such as machine learning (ML) have received increasing interest in recent years, since they can significantly contribute to disease detection, early diagnosis, and prediction and/or automatic classification of diseases [1], [14]. This means that medical staff have more information for decision-making, which consequently improves medical care and patient outcomes.

Although the application of ML models is relatively mature in other specialties, this has not been the case in the field of nephrology, as the lack of evidence and the limited scope of research in kidney disease have not allowed this specialty to benefit from these technologies [15], [16]. Therefore, the development of intelligent ML-based models may support nephrology medical staff in the context of identifying the occurrence of IDH in patients receiving HD.

Some of the ML models that are most used in the different specialties of medicine are: Logistic Regression, Random Forest, Multi-layer Perceptron and Extreme Gradient Boosting (XGBoost),

For this reason, the present research focuses on developing, evaluating and identifying an intelligent ML model that allows predictions regarding whether a patient will have IDH during an HD session. In addition, this work aims to determine the most relevant parameters associated with the occurrence of IDH based on the massive analysis of data relating to an HD session, including the clinical parameters measured at the beginning of the session and the most recent parameters of the blood analytics available at the time session. Data were provided by nephrology staff of the Hospital Príncipe de Asturias in Madrid, Spain. Different techniques were applied to this dataset (explained Section III), in order to obtain a useful dataset for the development of ML models that are able to determine whether or not the patient will have IDH at the beginning of HD treatment.

In this research study, different ML models (Logistic Regression, Random Forest, Multi-layer Perceptron and eXtreme Gradient Boosting [XGBoost]) were developed and studied. These ML models are widely used in fields such as medicine, energy, agriculture, among others [17]–[19] and, in addition, they are good at performing classification tasks; therefore, we decided to evaluate their performance in our case study. The models were developed by applying hold-out and cross-validation techniques (explained Section III). The evaluation criteria used to compare these models were the F1-score, Matthews Correlation Coefficient (MCC), areas under the receiver operating characteristic (AUROC) and precision-recall curve (AUPRC) metrics, as these are very useful when evaluating the performance of dichotomous classification models [20], [21].

After evaluating the different models, the results showed that the model based on the XGBoost demonstrated the best performance, obtaining 0.86 and 0.81 for F1-score and MCC respectively. In addition, obtaining an area under the curve (AUC) for the receiver operating characteristic (ROC) of 0.969 and an AUC for precision-recall (PR) of 0.945. Consequently, the results of this study suggest that the XGBoost model is feasible for its application as a predictor of IDH in HD sessions, due to its versatility and flexibility in terms of supporting informed decision-making by medical staff. It also provides medical staff with insight into the identification of the most decisive clinical and analytical parameters that affect the occurrence of IDH during an HD session.

The article is organized as follows. In Section II, we review and describe some prior work related to the context of our study. In Section III, we briefly detail the methodological procedure employed in the development of the ML models. In Section IV, we evaluate and present our results for several different metrics for each of the models developed. In Section V, we analyze and discuss the most important findings. Finally, in Section VI, we highlight our main conclusions and suggestions from the study and describe future research work

II. LITERATURE REVIEW

According to several studies [19], [22]–[25] ML-based models are widely used for prediction/classification in different specialties of medicine, such as neurology, cardiology, and pulmonology; in contrast, the study and development of ML is in its early stages in the nephrology field [15], [16]. However, the technological boom and the enormous volume of stored data can offer great benefits to this medical specialty [16], [26].

IDH is one of the most common complications in patients with CKD [7], [27]–[29] It is characterized by hemodynamic changes during HD treatment, which are due to multiple factors [7]. The implementation of intelligent models that are able to predict/classify the occurrence of IDH would allow health care personnel to provide early and effective treatments, and to make more informed decisions. Some of the most relevant work in this context is described below.

In [29], the authors predicted IDH using photoplethysmography signal fingerprinting (PPG), and subsequently employed a genetic algorithm for the extraction of the most important features and compared it with other algorithms for automatic IDH classification. They found that using their proposed method with the Adaboost algorithm showed good performance in classifying IDH and pre-IDH patients, where they obtained 90.6%, 86% and 93% accuracy, sensitivity and specificity respectively. Although these researchers obtained good results in terms of IDH classification, their work did not incorporate the analysis with other parameters (such as biochemical or analytical factors) that influence the occurrence of IDH.

The authors of [30], [31], studied the influence of heart rate variation and/or decreased oxygen saturation in patients with a view to predicting IDH. In the first work they reported satisfactory results, where they achieved an AUC of 0.63. Additionally, in the second study they indicated good results when predicting IDH with 80% accuracy. Therefore, they suggested that these parameters were useful for predicting IDH during HD; however, these researchers require justification of the effectiveness and consistency of these models.

Other researchers conducted a prospective cohort study [32] in which they analyzed heart rate variability together with some clinical parameters and patient laboratory results to predict IDH one month in advance. Their multivariate negative binomial model showed a significant ability to predict IDH using these parameters, where the model obtained an AUC of 0.804.

In [33] the authors proposed an intelligent model that was capable of alerting about Blood Pressure irregularities during HD. They posed the need to predict the Systolic Blood Pressure (SBP) of the following HD sessions, as well as to improve the quality of life of patients receiving HD. To do this, they used a dataset with 9,245 records, which presented 248 registered patients with their clinical variables (i.e., age, dialysis duration, blood flow, etc.). They compared different ML models (i.e., linear model, random forest, support vector regression, XGBoost and LASSO regression) and ensemble method. They noted that the random forest model ($R2=0.95$, $RMSE=6.6$) and XGBoost ($R2=1.0$, $RMSE=1.83$) obtained comparable performance in the training phase; however, in the test phase the random forest model ($R2=0.49$, $RMSE=16.2$) performed better. The ensemble method ($R2=0.50$, $RMSE=16.01$) performed best for the next SBP prediction.

Similarly, the study in [34] presented an intelligent early warning system that was able to predict IDH by checking the next Blood Pressure (BP) reading. They used a database of 653 patients, of whom underwent 55,516 HD sessions, leaving 285,705 valid BP records. The authors applied time-dependent logistic regression analysis to build predictive models, using 13 different patient clinical variables (i.e., age, genre, dry weight, dialysis duration, etc.). Their models obtained favorable results of 86% and 81% in terms of sensitivity and specificity for both nadir systolic BP (SBP) of <90 mmHg and <100 mmHg. In addition, their model obtained results of 64% and 64% in terms of sensitivity and specificity for SBP of ≥ 20 mmHg.

In [35], a deep neural network (DNN) model was proposed with the potential to determine the clinical factors that are related to the occurrence of IDH during an HD session. The researchers collected demographic data, HD clinical variables and laboratory data to identify factors associated with IDH. They tested different ML models (i.e., support vector machine, artificial neural network, random forest, decision tree, k-nearest neighbor, naïve Bayes) and showed that the proposed DNN-model was superior to the alternative models. They found that the 4-factor locus achieved great performance in the evaluation metrics (accuracy of 64.97% and sensitivity of 87.97%). However, the results obtained were not sufficiently satisfactory for this method to be employed in a medical setting during an HD session.

Likewise, in [36], a model based on a recurrent neural network (RNN) was developed to give real-time predictions of the risk of a patient presenting with IDH. The authors used different datasets where defined intradialytic hypotension when nadir systolic blood pressure (SBP) was <90 mmHg (IDH-1) or when a decrease in SBP ≥ 20 mmHg and/or a decrease in mean arterial pressure ≥ 10 mmHg on the basis of the initial SBPs (IDH-2) or prediction time SBPs (IDH-3) occurred within 1 hour. They compared the evaluation metrics resulting from the RNN model and other ML models (i.e., logistic regression, multi-layer perceptron and LightGBM), and they found that the RNN model had the best values in the different datasets. The RNN model achieved AUROCs of 0.94, 0.87 and 0.79 for IDH-1, IDH-2 and IDH-3 respectively.

In [37], ML algorithms were applied to develop models predicting hypotension after initiating continuous renal replacement therapy (CRRT). The authors defined hypotension as a reduction in mean arterial pressure of 20 mmHg from baseline within six hours. They tested different ML models including a support vector machine (SVM), a deep neural network (DNN), a light gradient boosting machine (LightGBM) and an extreme gradient boosting machine (XGBoost). Of these, the XGBoost model presented the best performance with an AUROC of 0.828. The authors noted that the use of ML algorithms could improve the predictability of hypotension after initiating CRRT.

The study in [38], used time-series differencing to extract the characteristics that most influenced the occurrence of IDH. This was called the time-relevant difference and was calculated based on the current time and the previous three IDH occurrence values. The researchers tested the generated features with non-time-series algorithms and showed that there was an improvement in the performance of the different algorithms when using this type of approach. For example, the LightGBM model achieved a performance of 86%, 66%, 88.9%, 85.1% and 0.946 for accuracy, precision, sensitivity, specificity and AUROC respectively. Therefore, they considered that this approach would be useful in small hospitals or settings where data were collected in a shorter time, as it gave acceptable values.

The aforementioned studies presented several models to predict or classify the occurrence of IDH in patients receiving HD. Each of these studies used a different approach or methods to treat the data arising from HD sessions. However, most of the related studies have sought to predict IDH while the patient is receiving HD, meaning that medical staff need to constantly monitor or receive alerts on the likelihood of IDH occurring in the patient. This can cause stress to nephrology staff, which may cause them to provide erroneous treatments to the patient.

Moreover, it is well known that the datasets in the nephrological domain are massive and heterogeneous, since they include a large number of variables (such as demographic data, laboratory results, analytical variables, etc.) and are collected during or after each patient's HD session, which means that the datasets have a high degree of dimensionality. In this sense, it would not be feasible to develop ML-based models with this large number of variables, as it would affect the performance and efficiency of each of these models [39].

For these reasons, our aim in this study is to provide new insight in predicting the occurrence of IDH in patients receiving HD. We developed and evaluated an intelligent ML-based model that will work at the start of the HD session, and it will allow medical staff to know whether the patient will have or will not have IDH during the HD session. To achieve this, we consider a combination of known clinical and analytical variables at the beginning of the HD session that can be associated with the occurrence of IDH through massive data analysis, which is described in the next section.

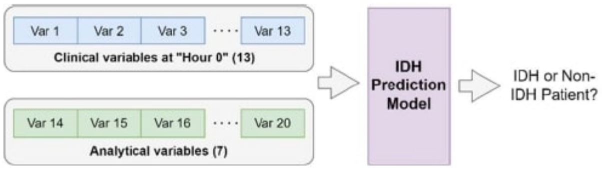


FIGURE 1. Model to predict IDH at the beginning of the HD session.

In the future, our proposed model will be implemented and deployed as a service in a real second-generation microservices ecosystem.

III. METHODOLOGY

In this section, the methodological process performed in the study is described and the data source is discussed.

A. IDH PREDICTIVE MODEL CONCEPT

The objective of the model (classifier) is to detect at the beginning of an HD session whether a patient is susceptible to IDH. For this purpose, a model built from data corresponding to multiple HD patients' sessions has been developed, including a set of analytical variables, from laboratory tests, close to the date of the HD session and the values of another set of clinical variables recorded at the beginning of the HD session in the so-called "hour 0" and of the detection of a drop in SBP throughout the dialysis session as indicated in the following Subsection. In this way, the predictive model will be fed at the beginning of the HD session by the data of the analytical variables available for the patient and the clinical values recorded at the beginning of said session at the so-called hour 0 (see Fig. 1).

B. DATA SOURCE

The dataset was obtained from the Hospital Príncipe de Asturias in Madrid, Spain, and contained records relating to clinical variables of patients who had received HD treatment between January 2016 and October 2019. These records had ethical approval from the hospital and the database was completely anonymized. We used as a starting point, the dataset of clinical variables that was analyzed and transformed in our previous study [40].

Unlike the previous dataset, in this work we defined IDH as a decrease in systolic blood pressure (SBP) of 20 mmHg or more if any of the SBPs measured at "Hour 1", "Hour 2", "Hour 3", and "Hour 4" was less than the SBP measured at "Hour 0" [41] that is illustrated in Fig. 2. Additionally, as a novel contribution, this work has incorporated new data on analytical variables extracted from the hospital laboratory.

In view of the above, the dataset used for this study was obtained by merging two different databases, as shown in Fig. 3. The first was drawn from the records of dialysis sessions and the variables were automatically recorded by the dialyzer itself in the HD sessions, while the second was based on the variables arising from a blood analysis performed on the patient. The two databases were merged in such a way that each HD session was linked to both the data from that HD

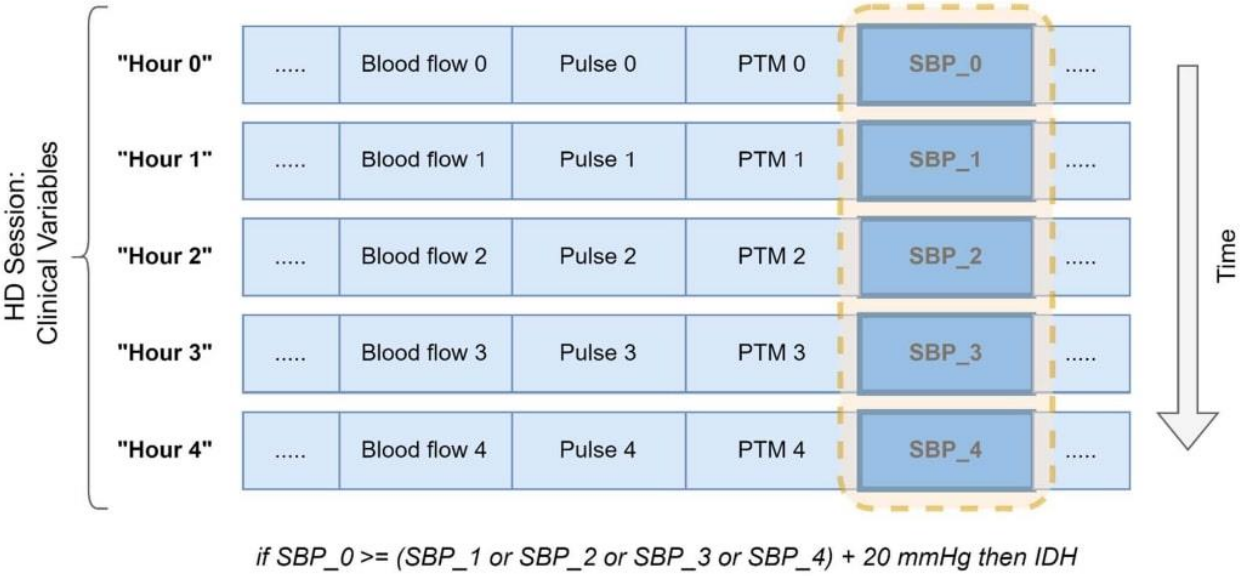


FIGURE 2. Identification and coding of an IDH during an HD session.

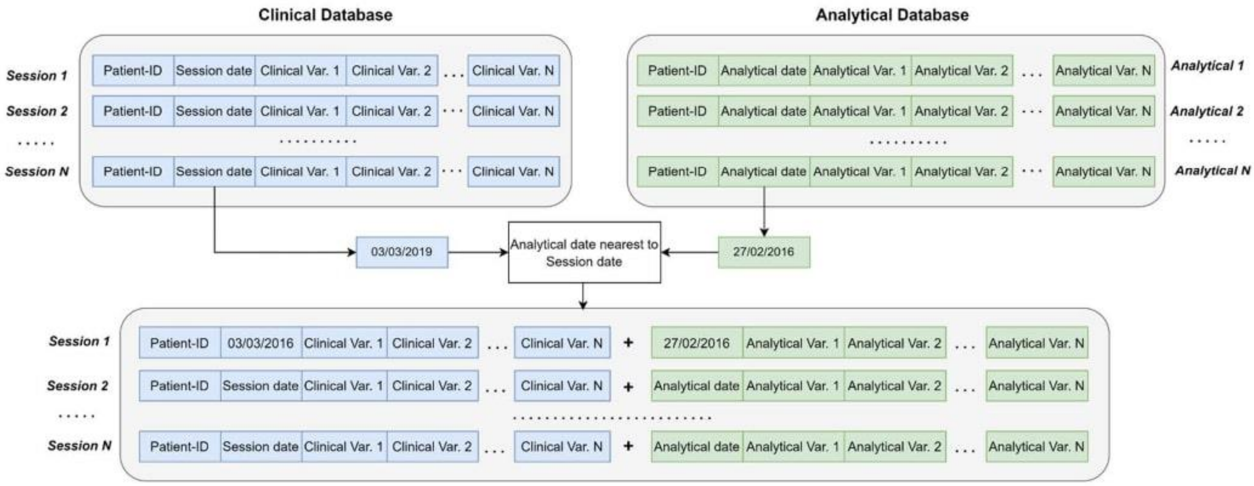


FIGURE 3. Procedure performed to merge the clinical and analytical databases.

session and the blood test carried out closest to that session. For example, if the patient underwent a quarterly blood test in an HD session, the quarterly blood test closest in time to that session was linked.

1) VARIABLE SELECTION

Once merged, the dataset contained more than 200 variables reflecting the clinical and analytical data related to the HD session. However, it is necessary to reduce this number of variables (parameter optimization), since the development of clinical models with a large number of variables is not advisable or practical. In fact, the selection of variables is an important and costly process in the development of clinical prediction models, so, if these variables are excessive or insignificant and do not improve the model, the medical staff, under their experience and clinical knowledge, determine which are the most important and influential variables for the development of the model [42], [43].

In this sense, we perform a second optimization process by reducing the dimensionality of the parameters. For this, we use ensemble trees [44] as a data mining technique and we consider the criteria and domain knowledge of the expert medical staff to identify the relevance of clinical and analytical variables. This left 20 variables (13 clinical and seven analytical), which are shown in Table 1 in the following subsection.

2) STUDY POPULATION

The generated dataset contained information from 22,234 sessions with 299 patients, in which 80% (17,793) of the

TABLE 1. Clinical and analytical variables from patients' hemodialysis sessions.

Type of variable	Variables	Counts (%)	Mean (± SD)
Clinical Variables Recorder at "Hour 0"	**(1) Age (years)**		66.4 (15.7)
	(2) Sex: n (%)		
	Male	16629 (74.8 %)	
	Female	5605 (25.2 %)	
	(3) Weight gain/Weight (before) (kg/kg)		0.024 (0.015)
	(4) Blood flow (mL/min)		379 (61.1)
	(5) Bath conductivity (mS/cm)		14 (0.9)
	(6) PTM (mmHg)		144.1 (55.9)
	(7) SBP (mmHg)		131.7 (28.1)
	(8) Pulse (bpm)		70.3 (13.7)
	(9) Ultrafiltration (mL)		147.1 (81.1)
	(10) Body temperature (ºC)		36.5 (0.4)
	(11) Dialyzers: n (%)		
	Evodial	390 (1.8 %)	
	Sureflux 2.1	452 (2.0 %)	
	ELISIO 210	5181 (23.3 %)	
	FILTRYZER NF-2.1H	582 (2.6 %)	
	FX 100	4212 (18.9 %)	
	FX 80	929 (4.2 %)	
	FX CorDiax 800	8450 (38.0 %)	
	Solacea 21H	2038 (9.2 %)	
	(12) Baths: n (%)		
	ACF 3A1	1558 (7.0 %)	
	ACF 3A12	2517 (11.3 %)	
	ACF 3A2	1583 (7.1 %)	
	ACF 3A5	15103 (67.9 %)	
	ACF 3A6	1473 (6.6 %)	
	(13) Dialyzer techniques: n (%)		
	HD – High flow	426 (1.9 %)	
	HD – Low flow	5 (0.0 %)	
	HDF - On line	21803 (98.1 %)	
Analytical Variables prior to the session	**(14) Leukocyte (x10e3/µL)**		6.7 (2.6)
	(15) Urea (mg/dL)		108.4 (37.5)
	(16) Potassium (mmol/L)		4.7 (0.9)
	(17) Phosphate (mg/dL)		4.6 (1.4)
	(18) Ionized calcium (mmol/L)		1.2 (0.1)
	(19) PH venous blood (ND)[1]		7.4 (0.1)
	(20) PCR (mg/L)		18.4 (31.9)

[1]Not Dimensional.

patient sessions did not involve IDH and 20% (4,441) did report the occurrence of IDH during the HD session.

All of the clinical and analytical variables used in this study are detailed in Table 1. It should be noted that categorical variables are presented as a quantity (percentage), and numerical variables are presented as a mean (±standard deviation (SD)).

C. DATA PRE-PROCESSING

Data pre-processing was performed in three phases, to determine how well the model performed in terms of predicting the occurrence or non-occurrence of IDH in patients' HD sessions. These three pre-processing phases are described below:

1) DATA CLEANING

Some variables with null values and several outliers were found in the records of the patient sessions, which were eliminated based on the domain knowledge provided by the expert medical staff involved in this study.

TABLE 2. Encoding categorial variables of the dataset.

Categorical Variables	Encoding
Sex	$\{s \ldots S\}_{s=1}^{S}$, S = 2
Dializers	$\{d \ldots D\}_{d=1}^{D}$, D = 8
Baths	$\{b \ldots B\}_{b=1}^{B}$, B = 5
Dializer techniques	$\{t \ldots T\}_{t=1}^{T}$, T = 3

2) ENCODING AND STANDARDIZATION OF DATA

Once the dataset was cleaned, we proceeded to encode the categorical variables in the dataset. Since our goal in this process was to avoid unnecessarily increasing the dimensionality of the dataset, we limited the number of dummy variables created from the categorical values [45]. Therefore, we applied the Label Encoding approach, since it allows us to identify the existing values of the categorical variables and replace them with a numerical value. An example of these encoded categorical variables is shown in Table 2.

We then developed a neural network-based model, specifically a multi-layer perceptron (MLP), for the prediction of

IDH and non-IDH patients. It is well-known that it is not recommended to train this type of model on raw data, so we applied standard normalization to the numerical variables of the dataset for the MLP.

3) DATA BALANCING

The dataset was unbalanced with respect to the IDH and non-IDH classes. To remedy this situation and improve the classification performance of the ML models, we balanced the data by applying the synthetic minority oversampling technique (SMOTE) and edited nearest neighbors (ENN) [46]–[49]. Using this method, we performed multiple tests to greatly reduce the non-IDH class while maintaining sufficiently high consistency between the proportions of data in both classes.

After applying SMOTE-ENN, the dataset was reduced to 14,803 sessions, of which 10,294 (70%) belonged to the non-IDH class and approximately 4,509 (30%) to the IHD class, meaning that the dataset was better distributed for classification. Although, the generated dataset does not show an optimal balance in the classes, this will not affect the performance of the ML models, since both classes are well represented and do not overlap, therefore, the ML models will be able to correctly classify the studied classes [50]. All ML models developed in this study were trained, evaluated and validated on this dataset.

D. MACHINE LEARNING MODELS

In this work, we developed and evaluated different ML models (classifiers) to predict the occurrence of IDH or non-IDH at the beginning of the HD session. In the following, we briefly describe each of these classifiers.

1) LOGISTIC REGRESSION

Logistic Regression (LR) is one of the most common ML models studied in the literature. This model employs a logistic function to predict the probability of a variable being classified in a two-class or target problem. This resulting variable is considered dichotomous [51]. Although, LR is easy to implement, interpret and efficient to train, it has difficulties in dealing with nonlinear problems and, moreover, the major limitation of LR is the assumption of linearity between the dependent variable and the independent variables [52].

2) RANDOM FOREST

Random forest (RF) is an ML algorithm, which is based on the ensemble method and is usually trained with the bagging (bootstrapping + aggregation) method. The idea behind the bagging method is that it combines multiple decision trees and each of them are individually trained with a different subset of the dataset features, thus generating a more accurate and stable prediction or classification [53], [54]. Some of the goodness of RF is that it performs well for nonlinear problems, is robust to outliers, has good accuracy in classification tasks. However, it is slow to train with large datasets, it is not easy to interpret and has biases when dealing with categorical variables.

3) MULTI-LAYER PERCEPTRON

The Multi-Layer Perceptron (MLP) is a type of artificial neural network, which consists of a series of neurons (nodes) that are fully connected. MLP networks are generally composed of three layers; an input layer, one or more hidden layers and an output layer. In addition, they employ nonlinear functions, called activation function, at the nodes of the hidden and output layers [54], [55]. This type of network allows distinguishing data that are not linearly separable, the prediction is very fast when the model is trained, it has the ability to learn in real time. On the contrary, MLP networks require the adjustment of several hyperparameters, it is not easy to know the influence of the independent variables on the dependent one, because they are a black-box and, in addition, they are sensitive to the scaling of the variables.

4) EXTREME GRADIENT BOOSTING

The eXtreme Gradient Boosting (XGBoost) model is an ensemble method based on decision trees, that uses a gradient boosting framework [33], [56]. The XGBoost incorporates techniques (i.e., regularization, sampling, pruning) as a measure to avoid the overfitting problem present in the gradient boosting (GB) algorithm. It can be run in parallel and using multiple cores and can therefore make use of the high computational power available in graphics cards and can even be run on server clusters. Another of the advantages of XGBoost is that it does not require scaling or normalization the data and it also has the ability to efficiently handle missing values [55].

5) HYPERPARAMETERS OF MODELS

Hyperparameter optimization or hyperparameter tuning consists of objectively search different values for model hyperparameters and choose a subset that results in a model that achieves the best performance on a given dataset. In this sense, the hyperparameters were tuned to each model using random search cross-validate algorithm, which allows testing different ranges of hyperparameter values and thus generating models with reliable and stable performance. It should be noted that each model has its own set of hyperparameters, therefore, we selected those hyperparameters that most influence the performance and accuracy of the model [54]. The best values of the hyperparameters of the different models are shown in Table 3.

E. MODEL DEVELOPMENT

In this work, each of the models studied were developed using the Python programming language (v3.6). The focus of the classifier models is to predict whether a patient will have IDH or non-IDH at the beginning of the HD session. We considered combining hold-out and cross-validation methods, as they allow to evaluate the predictive ability of the model, prevent overfitting and generate robust and reliable performances [57]–[60].

TABLE 3. Hyperparameter values obtained in the models.

Model	Hyperparameter values
Logistic Regression	Penalty = L2; max iterations = 50.
Random Forest	Max depth = 7, max features = auto, number of estimators = 30.
Multi-layer Perceptron	Size hidden layers=[64,32,1], activation= [relu and sigmoid], batch size= 64, epochs= 50, optimizer= Adam, dropout= 0.2.
XGBoost	Max depth = 6, number of estimators = 46, learning rate = 0.2, subsample = 0.75, grow policy = lossguide.

Within this framework, we first applied the hold-out method [61], where we divided the data into an 80% training-set and a 20% testing-set. With the training-set, the different ML models were evaluated by applying the Stratified K-fold cross validation method [62]–[64], which allows us to evaluate the performance and robustness of the models handling this data set. Subsequently, testing-set was used to evaluate the actual ability of the models to generalize from this unseen data, as well as to avoid introducing biases in the model training phase [65], [66]. The performance of the models was calculated using different metrics described in the following subsection. The entire development process is detailed in the pseudocode shown in Algorithm 1.

Algorithm 1 Pseudo-Code for the Development and Evaluation of Models

1. Begin
2. Load data source (Clinical-Analytical Data)
3. Apply pre-processing methods in the Clinical-Analytical Data.
4. Data splitting in training and testing
5. Apply Stratified K-fold for each ML model using training data
6. Validate each ML model using testing data
7. Calculate performance metrics
8. End

F. MODEL EVALUATION AND VALIDATION

We used Stratified K-fold cross-validation to evaluate models [62]–[64]. This consisted of dividing the training-set into k folds while maintaining the ratio between the classes (IDH and non-IDH). We applied five folds to each model due to the number of samples in the training-set [57]; at each iteration, a different fold was selected to train the model, and it was tested on the rest of the folds.

Although, the Stratified K-fold cross-validation allowed us to have a first evaluation of the performance of the models, these were validated with the testing-set, since this is a data set not used by the models and, therefore, allows us to validate the robustness and reliability of each one.

1) PERFORMANCE METRICS

Because there are a wide variety of metrics to evaluate models, we consider combining several metrics and interpreting their results in a holistic manner. [66]. For this reason, for each of the models, we evaluated several performance metrics, which include accuracy, precision, recall, F1-score and Matthews Correlation Coefficient (MCC) [20], [67], [68]. Equations for each of these metrics are (1)–(5), as shown at the bottom of the next page, where TP, TN, FP and FN represent true positives, true negatives, false positives and false negatives, respectively.

We also generated other evaluation and validation measures that are important to our study and in the medical context are of great importance [69]. These are Receiver Operating Characteristic curve (ROC) and Precision-Recall curve (PRC), since they allow determining the predictive performance of the model, provide readers with a general understanding of the utility of the model, and also allow evaluating the performance of the model when the data present some imbalance [21], [60], [70]. Likewise, for ROC and PRC, the areas under the curve (AUC) of both metrics were calculated and denoted as AUROC and AUPRC respectively.

To interpret and calculate the AUROC and AUPRC curves, we used the DeLong test [71]. This involved comparing the values obtained by the ML models, where the p-values were bilateral and those less than 0.05 were considered to be statistically significant. All comparisons were performed using the testing-set.

2) RECEIVER OPERATING CHARACTERISTIC CURVE (ROC)

This is one of the most widely used evaluation criteria and information on the ability of a model to distinguish between classes (in this case, IDH and non-IDH patients) [54], [67]. It does this by comparing the true positive rate (TPR) with the false positive rate (FPR). As TPR (recall or sensitivity) has already been defined above, the FPR can be defined as shown below:

$$FPR = \frac{FP}{FP + TN} \tag{6}$$

3) PRECISION-RECALL CURVE (PRC)

The PRC represents the performance of each of the different ML models against different thresholds, rather than a single metric (such as accuracy or F1-score). A classifier will show good performance if it has high accuracy and high recall. The equations for this curve were previously defined in (2) and (3).

It is important to note that, for our study we consider the F1-score, AUROC, AUPRC and MCC metrics as the main measures of comparison of the studied models, since they are useful to evaluate the performance of dichotomous classification models and to validate the robustness of the models with respect to a dataset with some imbalance as the one studied here [21].

IV. RESULTS

ML models have been developed in the Python programming language (v3.6) and implemented on a machine with an NVIDIA Jetson Nano Developer Kit [72]. To avoid biases in the comparison of the results obtained, all models studied

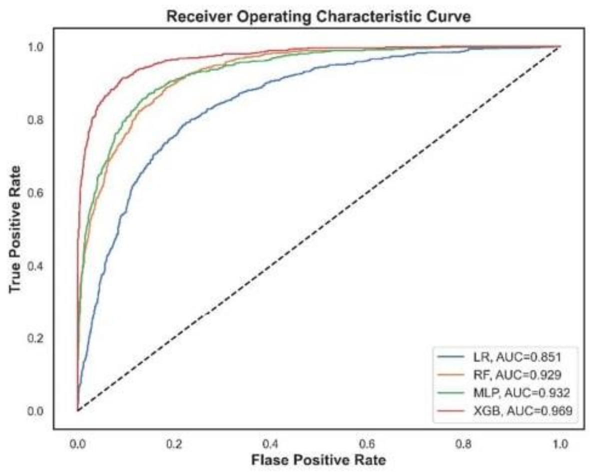


FIGURE 4. Receiver operation characteristic curve (ROC) and the area under the curve (AUC) for the different classification models developed here: LR, logistic regression; RF, random forest; MLP, multi-layer perceptron; XGB, XGBoost.

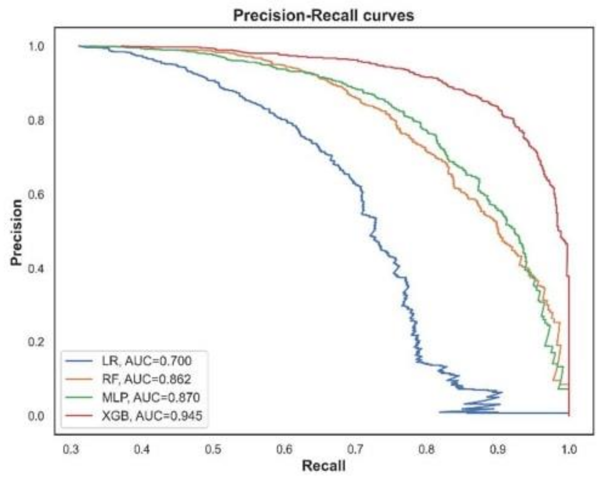


FIGURE 5. Precision-recall curve and area under the curve for the different classification models developed here. LR, logistic regression; RF, random forest; MLP, multi-layer perceptron; XGB, XGBoost.

in this work were also implemented using the same hardware and language resources as the model finally proposed.

To evaluate the performance of the models, we first used the training-set with the Stratified K-fold cross-validation method (Section III-E). During the model training process, the different metrics described in Section III-F were calculated. The results obtained by each model in predicting whether the patient will have a class (IDH or non-IDH) are shown in Table 4. Importantly, the values were calculated as the mean ($\pm$SD) for the metrics of accuracy, precision, recall, F1-score and MCC for both classes (in this case, IDH and non-IDH patients).

As a follow-up of this activity, we applied on the trained models the data remaining from applying the hold-out method (testing-set), as these were never seen or used by the models. Therefore, the results obtained by the models on the metrics applied in this work are shown in Table 5.

Once the models were trained and evaluated, we proceeded to perform a more exhaustive analysis of the models using the test-sets. In this sense, we plotted the ROC and PRC curves shown in Figs. 4 and 5. These plots visualize the ability of the models to predict whether the patient will have IDH or non-IDH at the beginning of the HD session.

As a complement, we calculated and compared the AUROC and AUPRC metrics using the DeLong test of the pRoc library [73]. The results of comparing these metrics in the models allowed us to interpret the ability, robustness and reliability of each of the models to discriminate between IDH and non-IDH classes. The results for the ROC and PRC curves for each model are shown in Table 6, with the AUC (AUROC and AUPRC) values for each model and their 95% confidence intervals (CI). Similarly, the p-value was obtained when comparing the ROC curve of the XGBoost model using the DeLong test, since this model showed the best performance.

Comparing the data in Tables 4, 5 and 6, it can be seen that the XGBoost model has the best predictive capacity when determining whether the patient will have IDH or non-IDH at the start of the HD session. However, it is important to verify that this resulting model has the capacity to generalize with records of future HD patients and, in addition, to verify that the model does not present overfitting. Therefore, Fig. 6 shows the learning curve of the XGBoost model, where

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1 - score = \frac{2 * (Precision * Recall)}{Precision + Recall} \tag{4}$$

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}} \tag{5}$$

TABLE 4. Classification results of models using Stratified k-fold cross-validation method (training-set).

Models	IDH Patient			Non-IDH Patient			Accuracy	MCC
	Precision	Recall	F1-score	Precision	Recall	F1-score		
LR	0.72 (0.002)	0.56 (0.015)	0.63 (0.017)	0.82 (0.003)	0.90 (0.007)	0.86 (0.004)	0.80 (0.005)	0.49 (0.013)
RF	0.87 (0.008)	0.61 (0.019)	0.71 (0.015)	0.85 (0.005)	0.96 (0.005)	0.90 (0.003)	0.85 (0.006)	0.63 (0.009)
MLP	0.82 (0.012)	0.74 (0.012)	0.78 (0.006)	0.88 (0.007)	0.938 (0.01)	0.91 (0.002)	0.87 (0.009)	0.70 (0.016)
XGB	**0.90 (0.013)**	**0.82 (0.014)**	**0.86 (0.013)**	**0.92 (0.006)**	**0.96 (0.005)**	**0.94 (0.004)**	**0.92 (0.007)**	**0.81 (0.017)**

LR, logistic regression; RF, random forest; MLP, multi-layer perceptron; XGB, XGBoost.

TABLE 5. Classification results of models using testing-set.

Models	IDH Patient			Non-IDH Patient			Accuracy	MCC
	Precision	Recall	F1-score	Precision	Recall	F1-score		
LR	0.71	0.59	0.64	0.83	0.89	0.86	0.80	0.51
RF	0.85	0.60	0.71	0.84	0.95	0.90	0.85	0.62
MLP	0.84	0.72	0.77	0.88	0.94	0.91	0.87	0.68
XGB	**0.90**	**0.83**	**0.86**	**0.93**	**0.96**	**0.94**	**0.92**	**0.81**

LR, logistic regression; RF, random forest; MLP, multi-layer perceptron; XGB, XGBoost.

TABLE 6. Results of areas under the curve (AUCs) to classify intradialytic hypotension of ML models.

Models	AUROC (95% CI)	p-value[1]	AUPRC (95% CI)
XGB	0.969 (0.963 to 0.975)		0.945 (0.929 to 0.958)
MLP	0.932 (0.922 to 0.941)	< 0.0001	0.87 (0.847 to 0.890)
RF	0.929 (0.920 to 0.938)	< 0.0001	0.862 (0.838 to 0.890)
LR	0.851 (0.837 to 0.866)	< 0.0001	0.70 (0.669 to 0.728)

[1]Comparison with the ROC curve of the XGBoost model using DeLong test.

TABLE 7. Comparison of our obtained model (XGBoost) with other studies.

Studies	Variables		Metrics				
	Clinical	Analytical	Acc.	F1	MCC	AUROC	AUPRC
S. Park et al., [32]	12	25	-	-	-	0.80	-
J. Chen et al., [35]	12	13	0.65	-	0.19	0.65	-
H. Lee et al., [36]	15	11	-	0.69	-	0.87	0.78
M. W. Kang et al., [37]	9	> 60	0.76	0.66	0.49	0.828	-
J.-Y. Yang et al., [38]	28	-	0.86	-	-	0.94	-
Our model	**13**	**7**	**0.92**	**0.86**	**0.81**	**0.97**	**0.94**

the graph shows the learning curve of the LogLoss error metric. This metric is based on probability and is used to measure the performance of the model.

Additionally, we compared the results of the XGBoost model obtained with the most recent works related to our research (Section II and III-B). In this sense, Table 7 shows the different studies and their respective results for the Accuracy (Acc.), F1-score (F1), MCC, AUROC and AUPRC.

Finally, we plotted the importance of each of the features according to the XGBoost model obtained in this study. In this way, the feature importance plot allowed us to determine the utility or value of each of the features (variables) involved in the construction of the model [74]. The calculation of the score for each feature was determined by its impact (gain) on the performance measure: the higher the value, the greater the impact of the feature (variable) on the performance of the model. Fig. 7 shows the importance of each feature according to the XGBoost model.

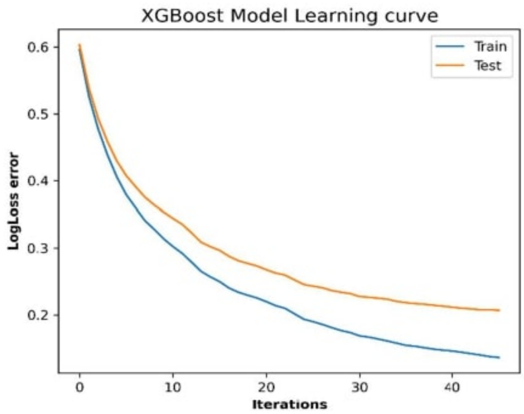


FIGURE 6. XGBoost model learning curve.

V. DISCUSSION

The purpose of this research was to develop and evaluate several ML models (classifiers) that would allow us to predict whether a patient will suffer from IDH or be non-IDH during an HD session based on a combination of clinical and analytical variables. Another objective of this research was to identify the model that has the best ability to discriminate between these classes.

Based on the performance analysis carried out here, we can observe from Table 4 that the results were acceptable in terms of the metrics evaluated (Section III-F). The XGBoost model achieved higher values by classifying both classes (IDH and non-IDH patients) compared to the MLP, RF and LR models. To use this predictive model in clinical practice, the results for predicting whether the patient will suffer from IDH were the most important. The XGBoost model obtained values of 0.92, 0.90, 0.82, 0.86 and 0.81 for the accuracy, precision, recall, F1-score and MCC, respectively, followed by MLP with 0.87, 0.82, 0.74, 0.78 and 0.70 and RF with 0.85, 0.87, 0.61, 0.71 and 0.63, respectively. The worst performer was LR, with values of 0.80, 0.72, 0.56, 0.63 and 0.49.

In the same way, we performed a validation of the different models by applying the testing-set, with the objective of validating the performance of the models with data that

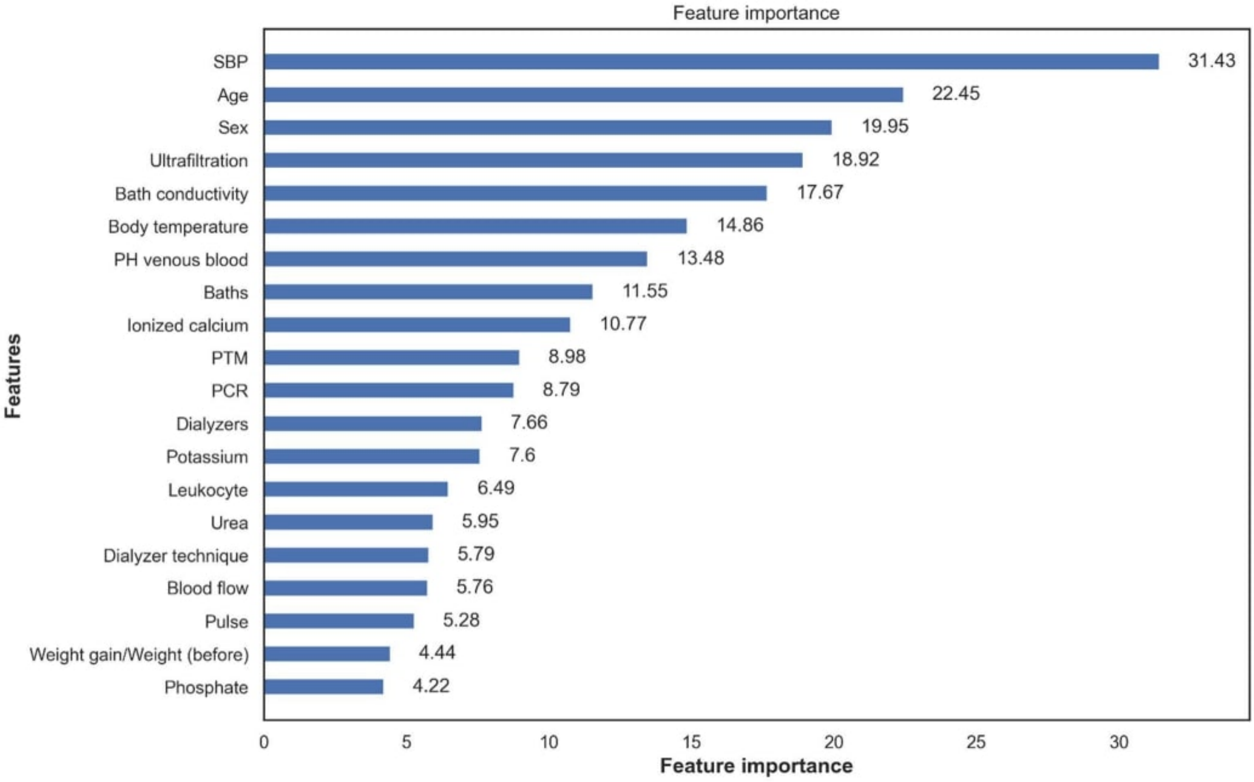


FIGURE 7. **Importance of each feature as determined using the XGBoost model.**

were not used for training. In this sense, it is evident in Table 5 that the XGBoost model also maintains a better performance over the other models when using these records and, in addition, the MCC metric is superior to the rest of the models; therefore, the model has a robustness when classifying patients (IDH or non-IDH). These results present an encouraging view of the XGBoost model in terms of identifying the occurrence of IDH or non-IDH in patients and may contribute to providing additional information in clinical practice. Furthermore, the degree of significance obtained from the different clinical and analytical variables opens possibilities for nephrology clinicians to study the complex task of determining the multiple factors associated with the occurrence of IDH [7], [11].

However, we also considered evaluating the different models by analyzing their ROC and PRC curves, as well as their areas under the AUROC and AUPRC curves. With these considerations in mind, it is evident from the graphs in Figs. 4 and 5, that the XGBoost model presents a good predictive ability when determining whether the patient will suffer IDH or non-IDH at the beginning of the HD session compared to the rest of the models studied. Furthermore, the DeLong test was used to compare and interpret these graphs. The results for the areas under the curves (AUROC and AUPRC) observed in Table 6 show a very good performance of the XGBoost model compared to other developed models. Additionally, there was a statistically significant difference ($p < 0.0001$) in the AUROC, meaning that this model was able to correctly identify and distinguish between the two classes (IDH and non-IDH) [75]–[77].

In addition, it is worth noting that the XGBoost model has excelled in each of the main metrics of this study (Section III-F). Within this context, we observe in Fig. 6 that the model has shown favorable performance for both training-set and testing-set and, furthermore, it is apparent from the LogLoss error curves that the model does not overfit, as the XGBoost model itself employs mechanisms to avoid this [55], [56], [78]. Therefore, this model would be able to generalize over records of future patients receiving HD.

We can highlight some important findings from these results. First, the XGBoost model achieved more accurate results than the rest of the ML models. This model also achieved better results for all metrics compared to previous work [40], where only the clinical variables of the patient HD were used, and presented better results than other studies [33], [35]–[37]. A possible explanation for these good results may lie in the novel incorporation of the clinical and analytical variables that were most influential, according to the applied ensemble-trees and the domain knowledge of nephrology experts in the creation of the ML predictor model. Reducing the dimensionality of the variables (clinical and analytical) provided a significant improvement and increased robustness in the model's ability to classify the data into classes (IDH and non-IDH). The developed predictor will be able to help medical staff to treat a patient at the beginning of the HD session to avoid the probable occurrence of IDH.

The second finding was that the XGBoost model demonstrated superior and favorable performance compared to other works using more complex models [33], [35], [36]. Although DNN-based models have a complex and abstract structure, which gives them the ability to learn/capture the different hidden patterns in the information provided. However, they have certain disadvantages compared to models such as the one based on the XGBoost, as they are sensitive to the scale of the data, dependent on the structure and size of the data, require a large computational load to adjust the large number of parameters of the model, and in the case of binary classification, require well balanced data [79]. Furthermore, DNN-based models are currently considered to be black-box models, meaning that their internal behavior is unknown and it is not possible to identify which variables or features contribute most to a given decision [35], [36].

In the third finding, Table 7 shows that our obtained model (XGBoost) achieves a formidable performance compared to the methods applied in other recent studies. It is important to note that the different works used a greater number of variables (clinical and/or analytical) compared to ours; therefore, not necessarily dealing with more variables generates better performance. In addition, using a smaller set of variables allows to obtain a more useful model, the medical staff to better identify the influence of the variables and, in addition, the collection of patient records is more agile.

Another significant finding can be observed from Fig. 7, which shows the variables or features, according to the XGBoost model, that most influenced the classification between patients who did or did not suffer from IDH during an HD session. As discussed above, the ability of this model to determine the patient features that contribute most to a potential case of IDH is important for decision-making by medical staff [80], [81].

Although the results obtained here are encouraging, there are some limitations to this study. This was a retrospective cohort study, and the ability of the model to classify IDH from new patient records would need to be validated, meaning that prospective validation is required. It would also be desirable to test the model' s capability with patient records from other hospital centers and therefore significantly expand the number of HD sessions. Finally, it is important to emphasize that the model obtained may be more robust when applying other optimization techniques.

VI. CONCLUSION AND FUTURE WORKS

In this study, we developed and evaluated ML models with the ability to predict whether a patient will suffer from IDH during an HD session from the available analytical and clinical values recorded at the beginning of the HD session. In particular, we have focused on data provided by the nephrology staff at the Hospital Príncipe de Asturias, Madrid, Spain. Our aim was to develop a flexible, robust model to support decision-making by hospital medical staff, to anticipate and treat the occurrence of IDH in patients receiving HD in a timely manner. We based the approach on a combination of variables (clinical and analytical) that were strongly determinant for the occurrence of IDH during an HD session.

From this predictive point of view, the XGBoost model achieved the best performance compared to the rest of the ML models, obtaining very reliable results of 0.92, 0.90, 0.82, 0.86 and 0.81 for accuracy, precision, recall, F1-score and MCC respectively. It also showed a favorable performance in the area under the curve (AUC) of receiver operating characteristic of 0.969 (95% confidence intervals [CI], 0.963 to 0.975), which was the highest compared to the other ML models ($p < 0.0001$). Similarly, the AUC of precision-recall was 0.945 (95% CI, 0.929 to 0.958). This suggests that models based on the XGBoost may be useful for application in medical specialties such as nephrology due to their ability to be implemented and deployed within different computational paradigms (such as the cloud, fog or edge) [82] and their versatility in terms of utilizing the high computational power offered by graphics cards. In addition, this model provides information on the way in which the importance of features can influence the occurrence of IDH during the HD session; therefore, this is a key advantage over other black-box models such as DNNs [79].

Finally, in future work, we will improve the performance of the model obtained after applying other optimization techniques. Also, we will implement and deploy this model within a second-generation microservices ecosystem that focuses on the telemonitoring of patients receiving HD, to take advantage of all the benefits and opportunities offered by this novel software architecture [83]. We will also build different models from another point of view, with the aim of classifying the occurrence of IDH and the range of SBP decrease that is encountered (e.g., mild, moderate or severe). This could provide improvements in the actions taken by nephrology staff in reference to the treatment that needs to be applied to HD patients. As a consequence, personalized methods or mechanisms of patient care could be developed.

CONFLICTS OF INTEREST

No conflict of interest exists. The authors wish to confirm that, there are no known conflicts of interest associated with this publication.

ACKNOWLEDGMENT

The authors thank the Clinical Team of the Nephrology Section of the Hospital Universitario Príncipe de Asturias for their advice and transfer of the clinical and analytical data. They also would like to thank the National Secretariat for Science, Technology and Innovation (SENACYT, Panama) and the Institute for the Training and Use of Human Resources (IFARHU), for the Doctoral Study Scholarship awarded within the IFARHU-SENACYT Doctoral Scholarship Program, in 2018. They also thank the Fresenius Medical

Care AG company for their collaboration in carrying out this work.

REFERENCES

[1] E. Dovgan, A. Gradišek, M. Luštrek, M. Uddin, A. A. Nursetyo, S. K. Annavarajula, Y.-C. Li, and S. Syed-Abdul, "Using machine learning models to predict the initiation of renal replacement therapy among chronic kidney disease patients," *PLoS ONE*, vol. 15, no. 6, Jun. 2020, Art. no. e0233976, doi: 10.1371/journal.pone.0233976.

[2] M. Gorostidi, M. Sánchez-Martínez, L. M. Ruilope, A. Graciani, J. J. de la Cruz, R. Santamaría, M. D. del Pino, P. Guallar-Castillón, F. de Álvaro, F. Rodríguez-Artalejo, and J. R. Banegas, "Prevalencia de enfermedad renal crónica en españa: Impacto de la acumulación de factores de riesgo cardiovascular," *Nefrología*, vol. 38, no. 6, pp. 606–615, Nov. 2018, doi: 10.1016/j.nefro.2018.04.004.

[3] K. Brück *et al.*, "CKD prevalence varies across the European general population," *J. Amer. Soc. Nephrol.*, vol. 27, no. 7, pp. 2135–2147, Jul. 2016, doi: 10.1681/ASN.2015050542.

[4] J. Arrieta, A. Rodríguez-Carmona, C. Remón, M. Pérez-Fontán, F. Ortega, J. A. Sánchez-Tomero, and R. Selgas, "Cost comparison between haemodialysis and peritoneal dialysis outsourcing agreements," *Nefrología*, vol. 32, no. 2, pp. 247–248, 2012, doi: 10.3265/Nefrologia.pre2011.Dec.11311.

[5] M. Kitamura and A. Saito, "Dialysis hypotension: A review of recent studies of causative factors," *Nephrology*, vol. 6, no. 3, pp. 109–112, Jun. 2001, doi: 10.1046/J.1440-1797.2001.00053.X.

[6] J. A. Chou, K. Kalantar-Zadeh, and A. T. Mathew, "A brief review of intradialytic hypotension with a focus on survival," *Seminars Dialysis*, vol. 30, no. 6, pp. 473–480, Nov. 2017, doi: 10.1111/SDI.12627.

[7] M. Kanbay, L. A. Ertuglu, B. Afsar, E. Ozdogan, D. Siriopol, A. Covic, C. Basile, and A. Ortiz, "An update review of intradialytic hypotension: Concept, risk factors, clinical implications and management," *Clin. Kidney J.*, vol. 13, no. 6, pp. 981–993, Dec. 2020, doi: 10.1093/CKJ/SFAA078.

[8] P. B. Reeves and F. R. M. Causland, "Mechanisms, clinical implications, and treatment of intradialytic hypotension," *Clin. J. Amer. Soc. Nephrol.*, vol. 13, no. 8, pp. 1297–1303, Aug. 2018, doi: 10.2215/CJN.12141017.

[9] D. F. Keane, J. G. Raimann, H. Zhang, J. Willetts, S. Thijssen, and P. Kotanko, "The time of onset of intradialytic hypotension during a hemodialysis session associates with clinical parameters and mortality," *Kidney Int.*, vol. 99, no. 6, pp. 1408–1417, Jun. 2021, doi: 10.1016/j.kint.2021.01.018.

[10] M. M. Assimon and J. E. Flythe, "Definitions of intradialytic hypotension," *Seminars Dialysis*, vol. 30, no. 6, pp. 464–472, Nov. 2017, doi: 10.1111/SDI.12626.

[11] C. Barbieri, I. Cattinelli, L. Neri, F. Mari, R. Ramos, D. Brancaccio, B. Canaud, and S. Stuard, "Development of an artificial intelligence model to guide the management of blood pressure, fluid volume, and dialysis dose in end-stage kidney disease patients: Proof of concept and first clinical assessment," *Kidney Diseases*, vol. 5, no. 1, pp. 28–33, 2019, doi: 10.1159/000493479.

[12] J. Saez-Rodriguez, M. M. Rinschen, J. Floege, and R. Kramann, "Big science and big data in nephrology," *Kidney Int.*, vol. 95, no. 6, pp. 1326–1337, Jun. 2019, doi: 10.1016/j.kint.2018.11.048.

[13] J. U. Becker, D. Mayerich, M. Padmanabhan, J. Barratt, A. Ernst, P. Boor, P. A. Cicalese, C. Mohan, H. V. Nguyen, and B. Roysam, "Artificial intelligence and machine learning in nephropathology," *Kidney Int.*, vol. 98, no. 1, pp. 65–75, Jul. 2020, doi: 10.1016/j.kint.2020.02.027.

[14] K. Shailaja, B. Seetharamulu, and M. A. Jabbar, "Machine learning in healthcare: A review," in *Proc. 2nd Int. Conf. Electron., Commun. Aerosp. Technol. (ICECA)*, Mar. 2018, pp. 910–914, doi: 10.1109/ICECA.2018.8474918.

[15] A. Verma, V. C. Chitalia, S. S. Waikar, and V. B. Kolachalama, "Machine learning applications in nephrology: A bibliometric analysis comparing kidney studies to other medicine subspecialities," *Kidney Med.*, vol. 3, no. 5, pp. 762–767, Sep. 2021, doi: 10.1016/j.xkme.2021.04.012.

[16] Q. Li, Q.-L. Fan, Q.-X. Han, W.-J. Geng, H.-H. Zhao, X.-N. Ding, J.-Y. Yan, and H.-Y. Zhu, "Machine learning in nephrology: Scratching the surface," *Chin. Med. J.*, no. 6, pp. 687–698, Mar. 2020, doi: 10.1097/CM9.0000000000000694.

[17] L. Benos, A. C. Tagarakis, G. Dolias, R. Berruto, D. Kateris, and D. Bochtis, "Machine learning in agriculture: A comprehensive updated review," *Sensors*, vol. 21, no. 11, p. 3758, May 2021, doi: 10.3390/S21113758.

[18] E. García-Martín, C. F. Rodrigues, G. Riley, and H. Grahn, "Estimation of energy consumption in machine learning," *J. Parallel Distrib. Comput.*, vol. 134, pp. 75–88, Dec. 2019, doi: 10.1016/j.jpdc.2019.07.007

[19] S. Panicker, "Use of machine learning techniques in healthcare: A brief review of cardiovascular disease classification," in *Proc. 2nd Int. Conf. Commun. Inf. Process. (ICCIP)*, 2020, pp. 1–12, doi: 10.2139/SSRN.3681833.

[20] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 1, Dec. 2020, doi: 10.1186/S12864-019-6413-7.

[21] T. Saito and M. Rehmsmeier, "The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets," *PLoS ONE*, vol. 10, no. 3, Mar. 2015, Art. no. e0118432, doi: 10.1371/journal.pone.0118432.

[22] A. Mujumdar and V. Vaidehi, "Diabetes prediction using machine learning algorithms," *Proc. Comput. Sci.*, vol. 165, pp. 292–299, Jan. 2019, doi: 10.1016/j.procs.2020.01.047.

[23] P. Khan, M. F. Kader, S. M. R. Islam, A. B. Rahman, M. S. Kamal, M. U. Toha, and K.-S. Kwak, "Machine learning and deep learning approaches for brain disease diagnosis: Principles and recent advances," *IEEE Access*, vol. 9, pp. 37622–37655, 2021, doi: 10.1109/ACCESS.2021.3062484.

[24] J. Lee, J. Woo, A. R. Kang, Y.-S. Jeong, W. Jung, M. Lee, and S. H. Kim, "Comparative analysis on machine learning and deep learning to predict post-induction hypotension," *Sensors*, vol. 20, no. 16, p. 4575, Aug. 2020, doi: 10.3390/S20164575.

[25] N. Caballé-Cervigón, J. L. Castillo-Sequera, J. A. Gómez-Pulido, J. M. Gómez-Pulido, and M. L. Polo-Luque, "Machine learning applied to diagnosis of human diseases: A systematic review," *Appl. Sci.*, vol. 10, no. 15, p. 5135, Jul. 2020, doi: 10.3390/APP10155135.

[26] P. Panwong and N. Iam-On, "Predicting transitional interval of kidney disease stages 3 to 5 using data mining method," in *Proc. 2nd Asian Conf. Defence Technol. (ACDT)*, Jan. 2016, pp. 145–150, doi: 10.1109/ACDT.2016.7437659.

[27] P. Reach, M. Touzot, Y. Lombardi, C. Maheas, E. Sacco, A. Fels, H. Beaussier, P. Ureña-Torres, G. Chatellier, C. Ridel, and M. Zuber, "Electrochemical skin conductance by Sudoscan®: A new tool to predict intradialytic hypotension," *Nephrol. Dialysis Transplantation*, vol. 36, no. 8, pp. 1511–1518, Jul. 2021, doi: 10.1093/ndt/gfab183.

[28] B. Sars, F. M. van der Sande, and J. P. Kooman, "Intradialytic hypotension: Mechanisms and outcome," *Blood Purification*, vol. 49, nos. 1–2, pp. 158–167, 2020, doi: 10.1159/000503776.

[29] M. Shahabi, V. R. Nafisi, and F. Pak, "Prediction of intradialytic hypotension using PPG signal features," in *Proc. 22nd Iranian Conf. Biomed. Eng. (ICBME)*, Nov. 2015, pp. 399–404, doi: 10.1109/ICBME.2015.7404178.

[30] C. Perazzini, P. G. Bolasco, L. Corazza, M. Tramonti, E. Mancini, A. Santoro, and S. Severi, "Prediction of intradialytic hypotension based on oxygen saturation variations," in *Proc. Comput. Cardiol.*, 2013, pp. 1223–1226.

[31] V. K. Y. Chan, L. W. C. Chan, and D. H. K. Chow, "Oxygen saturation and heart rate variations as predictors of intradialytic hypotension," *Nephrol. Nursing J.*, vol. 45, no. 1, pp. 53–61, 2018.

[32] S. Park, W.-J. Kim, N.-J. Cho, C.-Y. Choi, N. H. Heo, H.-W. Gil, and E. Y. Lee, "Predicting intradialytic hypotension using heart rate variability," *Sci. Rep.*, vol. 9, no. 1, p. 2574, Dec. 2019, doi: 10.1038/S41598-019-39295-Y.

[33] J.-C. Huang, Y.-C. Tsai, P.-Y. Wu, Y.-H. Lien, C.-Y. Chien, C.-F. Kuo, J.-F. Hung, S.-C. Chen, and C.-H. Kuo, "Predictive modeling of blood pressure during hemodialysis: A comparison of linear model, random forest, support vector regression, XGBoost, LASSO regression and ensemble method," *Comput. Methods Programs Biomed.*, vol. 195, Oct. 2020, Art. no. 105536, doi: 10.1016/j.cmpb.2020.105536.

[34] C.-J. Lin, C.-Y. Chen, P.-C. Wu, C.-F. Pan, H.-M. Shih, M.-Y. Huang, L.-H. Chou, J.-S. Tang, and C.-J. Wu, "Intelligent system to predict intradialytic hypotension in chronic hemodialysis," *J. Formosan Med. Assoc.*, vol. 117, no. 10, pp. 888–893, Oct. 2018, doi: 10.1016/j.jfma.2018.05.023.

[35] J.-B. Chen, K.-C. Wu, S.-H. Moi, L.-Y. Chuang, and C.-H. Yang, "Deep learning for intradialytic hypotension prediction in hemodialysis patients," *IEEE Access*, vol. 8, pp. 82382–82390, 2020, doi: 10.1109/ACCESS.2020.2988993.

[36] H. Lee, D. Yun, J. Yoo, K. Yoo, Y. C. Kim, D. K. Kim, K.-H. Oh, K. W. Joo, Y. S. Kim, N. Kwak, and S. S. Han, "Deep learning model for real-time prediction of intradialytic hypotension," *Clin. J. Amer. Soc. Nephrol.*, vol. 16, no. 3, pp. 396–406, Mar. 2021, doi: 10.2215/CJN.09280620.

[37] M. W. Kang, S. Kim, Y. C. Kim, D. K. Kim, K.-H. Oh, K. W. Joo, Y. S. Kim, and S. S. Han, "Machine learning model to predict hypotension after starting continuous renal replacement therapy," *Sci. Rep.*, vol. 11, no. 1, p. 17169, Dec. 2021, doi: 10.1038/S41598-021-96727-4.

[38] J.-Y. Yang, H. W. Hu, C.-H. Liu, K.-Y. Chen, C.-H. Un, C.-C. Huang, C.-C. Chen, C.-C.-K. Lin, H. Chang, and H.-M. Lin, "Differencing time series as an important feature extraction for intradialytic hypotension prediction using machine learning," in *Proc. IEEE 3rd Eurasia Conf. Biomed. Eng., Healthcare Sustainability (ECBIOS)*, May 2021, pp. 19–20, doi: 10.1109/ECBIOS51820.2021.9510749.

[39] S. S. Thakur, S. S. Abdul, H.-Y. Chiu, R. B. Roy, P.-Y. Huang, S. Malwade, A. A. Nursetyo, and Y.-C. Li, "Artificial-intelligence-based prediction of clinical events among hemodialysis patients using non-contact sensor data," *Sensors*, vol. 18, no. 9, p. 2833, Aug. 2018, doi: 10.3390/S18092833.

[40] J. A. Gómez-Pulido, J. M. Gómez-Pulido, D. Rodríguez-Puyol, M.-L. Polo-Luque, and M. Vargas-Lombardo, "Predicting the appearance of hypotension during hemodialysis sessions using machine learning classifiers," *Int. J. Environ. Res. Public Health*, vol. 18, no. 5, p. 2364, Feb. 2021, doi: 10.3390/IJERPH18052364.

[41] K. Workgroup, "K/DOQI clinical practice guidelines for cardiovascular disease in dialysis patients," *Amer. J. Kidney Diseases*, vol. 45, no. 3, pp. S1–S153, Apr. 2005.

[42] M. Z. I. Chowdhury and T. C. Turin, "Variable selection strategies and its importance in clinical prediction modelling," *Family Med. Community Health*, vol. 8, no. 1, Feb. 2020, Art. no. e000262, doi: 10.1136/FMCH-2019-000262.

[43] E. W. Steyerberg and Y. Vergouwe, "Towards better clinical prediction models: Seven steps for development and an ABCD for validation," *Eur. Heart J.*, vol. 35, no. 29, pp. 1925–1931, Aug. 2014, doi: 10.1093/eurheartj/ehu207.

[44] G. Mehrpoor, M. M. Azimzadeh, and A. Monfared, "Data mining: A novel outlook to explore knowledge in health and medical sciences," *Int. J. Travel Med. Global Health*, vol. 2, no. 2, pp. 87–90, 2014. [Online]. Available: http://www.ijtmgh.com/article_33289.html

[45] C. Seger, "An investigation of categorical variable encoding techniques in machine learning: Binary versus one-hot and feature hashing," School Elect. Eng. Comput. Sci. (EECS), KTH, Stockholm, Sweden, Tech. Rep. 2018:596, 2018, p. 34.

[46] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Dec. 2002.

[47] M. Beckmann, N. F. F. Ebecken, and B. S. L. P. de Lima, "A KNN undersampling approach for data balancing," *J. Intell. Learn. Syst. Appl.*, vol. 7, no. 4, p. 104, 2015.

[48] B. R. Manju and A. R. Nair, "Classification of cardiac arrhythmia of 12 lead ECG using combination of SMOTEENN, XGBoost and machine learning algorithms," in *Proc. 9th Int. Symp. Embedded Comput. Syst. Design (ISED)*, Dec. 2019, pp. 1–7, doi: 10.1109/ISED48680.2019.9096244.

[49] V. López, A. Fernández, S. García, V. Palade, and F. Herrera, "An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics," *Inf. Sci.*, vol. 250, pp. 113–141, Nov. 2013, doi: 10.1016/j.ins.2013.07.007.

[50] B. Krawczyk, "Learning from imbalanced data: Open challenges and future directions," *Prog. Artif. Intell.*, vol. 5, no. 4, pp. 221–232, 2016, doi: 10.1007/S13748-016-0094-0.

[51] K. Shah, H. Patel, D. Sanghvi, and M. Shah, "A comparative analysis of logistic regression, random forest and KNN models for the text classification," *Augmented Hum. Res.*, vol. 5, no. 1, p. 12, Dec. 2020, doi: 10.1007/S41133-020-00032-0.

[52] G. M. Fitzmaurice and N. M. Laird, "Multivariate analysis: Discrete variables (logistic regression)," in *International Encyclopedia of the Social & Behavioral Sciences*, N. J. Smelser and B. S. Baltes, Eds. Oxford, U.K.: Pergamon, 2001, pp. 10221–10228.

[53] K. Kirasich, T. Smith, and B. Sadler, "Random forest vs logistic regression: Binary classification for heterogeneous datasets," *SMU Data Sci. Rev.*, vol. 1, no. 3, p. 9, 2018.

[54] A. Geron, *Hands-on Machine Learning With Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, 2nd ed. Newton, MA, USA: O'Reilly Media, Inc., 2019.

[55] J. Wu, Y. Li, and Y. Ma, "Comparison of XGBoost and the neural network model on the class-balanced datasets," in *Proc. IEEE 3rd Int. Conf. Frontiers Technol. Inf. Comput. (ICFTIC)*, Nov. 2021, pp. 457–461, doi: 10.1109/ICFTIC54370.2021.9647373.

[56] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[57] S. Yadav and S. Shukla, "Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification," in *Proc. IEEE 6th Int. Conf. Adv. Comput. (IACC)*, Feb. 2016, pp. 78–83, doi: 10.1109/IACC.2016.25.

[58] A. Vabalas, E. Gowen, E. Poliakoff, and A. J. Casson, "Machine learning algorithm validation with a limited sample size," *PLoS ONE*, vol. 14, no. 11, Nov. 2019, Art. no. e0224365, doi: 10.1371/journal.pone.0224365.

[59] D. Berrar, "Cross-validation," in *Encyclopedia of Bioinformatics and Computational Biology*, S. Ranganathan, M. Gribskov, K. Nakai, and C. B. Schönbach, Eds. Oxford, U.K.: Elsevier, 2019, pp. 542–545.

[60] G. S. Handelman, H. K. Kok, R. V. Chandra, A. H. Razavi, S. Huang, M. Brooks, M. J. Lee, and H. Asadi, "Peering into the black box of artificial intelligence: Evaluation metrics of machine learning methods," *Amer. J. Roentgenol.*, vol. 212, no. 1, pp. 38–43, Jan. 2019, doi: 10.2214/AJR.18.20224.

[61] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proc. IJCAI*, 1995, vol. 14, no. 2, pp. 1137–1145.

[62] S. Arlot and A. Celisse, "A survey of cross-validation procedures for model selection," *Statist. Surv.*, vol. 4, pp. 40–79, Jan. 2010, doi: 10.1214/09-SS054.

[63] K. R. Lakshmi, Y. Nagesh, and M. V. Krishna, "Performance comparison of three data mining techniques for predicting kidney dialysis survivability," *Int. J. Adv. Eng. Technol.*, vol. 7, no. 1, p. 242, 2014.

[64] J. G. Moreno-Torres, J. A. Saez, and F. Herrera, "Study on the impact of partition-induced dataset shift on k-fold cross-validation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 8, pp. 1304–1312, Aug. 2012, doi: 10.1109/TNNLS.2012.2199516.

[65] J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan, and A. Saboor, "Heart disease identification method using machine learning classification in E-Healthcare," *IEEE Access*, vol. 8, pp. 107562–107582, 2020, doi: 10.1109/ACCESS.2020.3001149.

[66] S. A. Hicks, I. Strümke, V. Thambawita, M. Hammou, M. A. Riegler, P. Halvorsen, and S. Parasa, "On evaluation metrics for medical applications of artificial intelligence," *Sci. Rep.*, vol. 12, no. 1, p. 5979, Dec. 2022, doi: 10.1038/S41598-022-09954-8.

[67] D. M. W. Powers, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation," *Int. J. Mach. Learn. Technol.*, vol. 2, no. 1, pp. 37–63, 2011.

[68] A. U. Haq, J. P. Li, M. H. Memon, S. Nazir, and R. Sun, "A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms," *Mob. Inf. Syst.*, vol. 2018, Dec. 2018, Art. no. 3860146, doi: 10.1155/2018/3860146.

[69] R. Kumar and A. Indrayan, "Receiver operating characteristic (ROC) curve for medical researchers," *Indian Pediatrics*, vol. 48, no. 4, pp. 277–287, 2011, doi: 10.1007/S13312-011-0055-4.

[70] C. J. Chiew, N. Liu, T. H. Wong, Y. E. Sim, and H. R. Abdullah, "Utilizing machine learning methods for preoperative prediction of postsurgical mortality and intensive care unit admission," *Ann. Surg.*, vol. 272, no. 6, pp. 1133–1139, Dec. 2020, doi: 10.1097/SLA.0000000000003297.

[71] E. R. DeLong, D. M. DeLong, and D. L. Clarke-Pearson, "Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach," *Biometrics*, vol. 44, pp. 837–845, Sep. 1988, doi: 10.2307/2531595.

[72] F. Dustin. (2019). *Jetson Nano Brings AI Computing to Everyone*. Nvidia. Accessed: Sep. 23, 2021. [Online]. Available: https://developer.nvidia.com/blog/jetson-nano-ai-computing/

[73] X. Robin, N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J.-C. Sanchez, and M. Müller, "pROC: An open-source package for R and S+ to analyze and compare ROC curves," *BMC Bioinf.*, vol. 12, no. 1, p. 77, Mar. 2011, doi: 10.1186/1471-2105-12-77.

[74] M. Li, X. Fu, and D. Li, "Diabetes prediction based on XGBoost algorithm," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 768, p. 72093, 2020, doi: 10.1088/1757-899x/768/7/072093.

[75] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology*, vol. 143, no. 1, pp. 29–36, Apr. 1982, doi: 10.1148/RADIOLOGY.143.1.7063747.

[76] M. H. Zweig and G. Campbell, "Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine," *Clin. Chem.*, vol. 39, no. 4, pp. 561–577, 1993.

[77] B. Snider, E. A. McBean, J. Yawney, S. A. Gadsden, and B. Patel, "Identification of variable importance for predictions of mortality from COVID-19 using AI models for Ontario, Canada," *Frontiers Public Health*, vol. 9, p. 798, Jun. 2021, doi: 10.3389/FPUBH.2021.675766.

[78] D. Zhang, L. Qian, B. Mao, C. Huang, B. Huang, and Y. Si, "A data-driven design for fault detection of wind turbines using random forests and XGboost," *IEEE Access*, vol. 6, pp. 21020–21031, 2018, doi: 10.1109/ACCESS.2018.2818678.

[79] O. Niel and P. Bastard, "Artificial intelligence in nephrology: Core concepts, clinical applications, and perspectives," *Amer. J. Kidney Diseases*, vol. 74, no. 6, pp. 803–810, Dec. 2019, doi: 10.1053/j.ajkd.2019.05.020.

[80] M. Hueso, A. Vellido, N. Montero, C. Barbieri, R. Ramos, M. Angoso, J. M. Cruzado, and A. Jonsson, "Artificial intelligence for the artificial kidney: Pointers to the future of a personalized hemodialysis therapy," *Kidney Diseases*, vol. 4, no. 1, pp. 1–9, 2018, doi: 10.1159/000486394.

[81] S. Ansari, S. Molaei, K. Oldham, M. Heung, K. R. Ward, and K. Najarian, "An extended Kalman filter with inequality constraints for real-time detection of intradialytic hypotension," in *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2017, pp. 2227–2230, doi: 10.1109/EMBC.2017.8037297.

[82] I. Odun-Ayo, R. Goddy-Worlu, V. Geteloma, and E. Grant, "A systematic mapping study of cloud, fog, and edge/mobile devices management, hierarchy models and business models," *Adv. Sci., Technol. Eng. Syst. J.*, vol. 4, no. 2, pp. 91–101, 2019.

[83] H. Calderon-Gomez, L. Mendoza-Pitti, M. Vargas-Lombardo, J. M. Gomez-Pulido, J. L. Castillo-Sequera, J. Sanz-Moreno, and G. Sencion, "Telemonitoring system for infectious disease prediction in elderly people based on a novel microservice architecture," *IEEE Access*, vol. 8, pp. 118340–118354, 2020, doi: 10.1109/ACCESS.2020.3005638.

LUIS MENDOZA-PITTÍ was born in Panama, in 1992. He received the B.E. degree in engineering systems and computing from the Technological University of Panama, Chiriqui, Panama, in 2015. He is currently pursuing the Ph.D. degree in information and knowledge engineering with the University of Alcalá, Alcalá de Henares, Spain. From 2013 to 2016, he was a Student Research Assistant with the Technological University of Panama. His current research interests include smart health, cloud computing, software engineering, and machine learning models.

JOSÉ MANUEL GÓMEZ-PULIDO was born in Cáceres, Spain. He received the B.S. and M.S. degrees in telecommunications engineering from the University Polytechnic of Madrid, Spain, in 1988 and 1994, respectively, and the Ph.D. degree in telecommunication from the University of Alcalá, Madrid, in 2004. In 1992, he was an Assistant Researcher and a Professor with the University Polytechnic of Madrid. From 1993 to 2004, he was with the Signal Theory and Communications Department, University of Alcalá. He is currently an Assistant Professor with the Computational Science Department, University of Alcalá. He has participated in more than 24 research projects several research projects with Spanish and European companies, related with radio propagation in mobile communications, signal processing in radar, ultrasound and laser applications, EMI–EMC design in on-board satellite instrumentation, radio propagation materials proprieties, and numerical technics applied in telecommunication. His research interests include wireless sensors networks, smart grids, and new technologies in the data mining and artificial intelligence world applied to the ICT.

MIGUEL VARGAS-LOMBARDO (Senior Member, IEEE) received the Ph.D. degree from the Polytechnic University of Madrid. He is currently a member of the National Research System (SNI-SENACYT), Panama. He is also a Professor and a Researcher with the Technological University of Panama.

JUAN A. GÓMEZ-PULIDO (Senior Member, IEEE) received the Ph.D. degree in physics (electronics specialty) from Complutense University, Madrid, Spain, in 1993. He is currently a Full Professor of computer organization and design of processors with the Department of Technology of Computers and Communications, Universidad de Extremadura, Spain. He has authored or coauthored 70 ISI journals, almost 100 book chapters, and 300 peer-reviewed conference proceedings. His main research interests include machine learning applied to big-data analysis, reconfigurable and embedded computing based on FPGA devices, bioinformatics, optimization, and evolutionary computing.

MARÍA-LUZ POLO-LUQUE received the B.S. degree in nursing from the Fundación Jiménez Díaz of Madrid, Spain, in 1979, the M.S. degree in diététique superieure from the Henri Poincaré University of Nancy, France, the M.S. degree in nursing sciences from the University of Alicante, Spain, the M.S. degree in gerontology from the University of Santiago de Compostela, Spain, and the Ph.D. degree in health sciences from the University of Alcalá, Spain, in 2011. She has conducted research at the SU.VI.MAX Center, France, regarding methods of measuring food consumption and data analysis. She has had an extensive career as a Nurse at the Transplant Unit, Hospital Fundación Jiménez Díaz, Hospital Necker, Paris, France. She is currently a Professor of health sciences at the University of Alcalá and participates in the UAH Cooperation Group (ACORDES), which aims to promote the resilience of the most vulnerable communities through social and gender equity. Her main research interests include geriatric nursing, nursing cares in highly complex chronic patients, nursing research in health and social problems, and health computing and intelligent systems in nursing cares.

DIEGO RODRÍGUEZ-PUYOL was born in Madrid. He received the Ph.D. degree from Autonoma University, Madrid. He is currently a Full Professor of medicine at the University of Alcalá and the Chief of nephrology at the University Hospital "Príncipe de Asturias." He performed his medicine studies at Complutense University. He is a Nephrologist with a significant clinical, teaching, and translational research expertise. His work has been devoted to the study of renal disease pathogenesis and progression, synthesis of new drugs useful for treating renal diseases, and big data applied to the solution of clinical problems. He has published more than 170 original manuscripts in different international publications.

3 Otras publicaciones

Durante el desarrollo de la investigación se presentaron otras publicaciones científicas que respaldan el interés de esta tesis doctoral. Además, se realizaron aportaciones como colaborador en otras publicaciones científicas de alto impacto.

3.1 Publicaciones en congresos internacionales

3.1.1 Arquitecturas de software para la eficiencia energética en edificaciones: una revisión sistemática

Publicado en: Revista Ibérica de Sistemas e Tecnologias de Informação (RISTI)

Autores: Luis Mendoza-Pittí, José Manuel Gómez-Pulido, Miguel Vargas-Lombardo.

ISSN: 1646-9895

Título del congreso: I congreso de ciencia de la computación, electrónica e ingeniería industrial.

Fecha de publicación: 01/10/2019

Fecha del congreso: 28/10/2019 – 31/10/2019

Páginas y número: 40-52, E23

Ranking SJR (2019): Computer Science (Q4), 0.145

Organizador: Universidad Técnica de Ambato, Facultad de ingeniería en sistemas, electrónica e industrial.

3.1.2 Microservices Architecture for Efficient eHealth Service Generation in the Cloud.

Congreso: International Conference on Tourism Technology & Systems (ICOTTS), 2021

Autores: Huriviades Calderon-Gomez, Luis Mendoza-Pitti, José Manuel Gómez Pulido, Miguel Vargas-Lombardo.

Fecha de publicación: 24/06/2022

Capítulo de libro en Advances in Tourism, Technology and Systems, Springer.

Ubicación: Universidad de Cartagena, Cartagena de Indias, Colombia

Fecha del congreso: 04/11/2021 – 06/11/2021

Organizador: Universidad de Cartagena

DOI: 10.1007/978-981-19-1040-1_24

3.1.3 Ubiquitous and Ambient Assisted Living eHealth Platforms for the Republic of Panama: Two Cases of Study

Congreso: International Conference on Ambient Intelligence for Health (AmIHEALTH) 2015

Autores: Juan José Saldaña, Luis Mendoza, Edgardo Pitti, Miguel Vargas Lombardo.

Fecha de publicación: 20/01/2015

Capítulo de libro en Lecture Notes in Computer Science, Springer.

Ubicación: Puerto Varas, Chile

Fecha del congreso: 01/12/2015 – 04/12/2015

DOI: 10.1007/978-3-319-26508-7_14

3.1.4 Plataforma Ubicua para la Gestión de Cuidados Paliativos en Panamá

Congreso: Inteligencia Ambiental, Ingeniería de Software y Salud Electrónica & Móvil (AmISEmeH), 2016

Autores: Juan José Saldaña, Luis Mendoza, Miguel Vargas Lombardo.

Fecha de publicación: 2016

Ubicación: Panamá

Fecha del congreso: 21/01/2016 – 22/01/2016

ISBN 978-9962-698-35-7

3.2 Colaboración en otras publicaciones

3.2.1 Telemonitoring system for infectious disease prediction in elderly people based on a novel microservice architecture

Autores: Huriviades Calderón-Gómez, Luis Mendoza-Pittí, Miguel Vargas-Lombardo, José Manuel Gómez-Pulido, José Luis Castillo-Sequera, José Sanz-Moreno, Gloria Sención

Fecha de publicación: 29/06/2020

Revista IEEE ACCESS

Volumen: 8, Páginas: 118340-118354

DOI: 10.1109/ACCESS.2020.3005638

Ranking JCR 2020: Computer science, information systems (Q2

Factor de impacto (JCR 2020): 3.367

Ranking SJR 2020: Computer Science (Q1)

Factor de impacto (SJR 2020): 0.59

3.2.2 Evaluating Service-Oriented and Microservice Architecture Patterns to Deploy eHealth Applications in Cloud Computing Environment

Autores: Huriviades Calderón-Gómez, Luis Mendoza-Pittí, Miguel Vargas-Lombardo, José Manuel Gómez-Pulido, Diego Rodríguez-Puyol, Gloria Sención, María-Luz Polo-Luque

Fecha de publicación: 11/05/2021

Revista MDPI – Applied Science

Volumen: 11, Issue: 10

ISBN: 2076-3417

DOI: 10.3390/app11104350

Ranking JCR 2020: Engineering, multidisciplinary (Q2).

Factor de impacto (JCR 2020): 2.679

Ranking SJR 2020: Computer Science Applications (Q2).

Factor de impacto (SJR 2020): 0.435

4 Experiencia investigadora

Durante el desarrollo de mi estancia de investigación en la Universidad de Alcalá de Henares, el doctorando ha participado en diversos proyectos de investigación. Esto se muestra en la Figura 4.

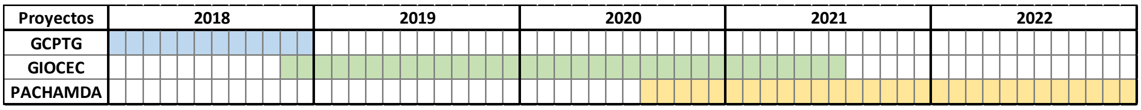


Figura 4. Distribución de la experiencia investigadora del doctorando.

Los proyectos que se aprecian en la figura anterior muestran la experiencia investigadora adquirida mediante la participación en los distintos proyectos llevados durante la estancia de estudios del doctorando. A continuación, se describen brevemente los proyectos en cuestión.

4.1 Gestión de Cuidados Paliativos y Trastornos Genéticos en Panamá (GCPTG)

El proyecto de Gestión de Cuidados Paliativos y Trastornos Genéticos en Panamá (GCPTG) es un proyecto liderado por el Ministerio de Salud (MINSA) y avalado por la Universidad Tecnológica de Panamá.

Este proyecto plantea el desarrollo de una plataforma web que permita gestionar y capturar la información de los pacientes que reciben cuidados paliativos y trastornos genéticos en Panamá. A continuación, se describen los objetivos del proyecto GCPTG:

- Gestionar la atención equipo de salud-paciente, a través de un registro sistematizado de atención.
- Facilitar tanto a los pacientes como a los familiares información sobre el manejo de enfermedades para mejorar la calidad de vida de las personas con enfermedades terminales.
- Proporcionar una historia clínica digitalizada permite al personal médico mantener registros de las enfermedades o patologías de los pacientes, a fin de conocer el estado y/o evolución del paciente.
- Recolectar la información del paciente para la toma de decisiones informadas mediante el uso de avanzadas técnicas de machine learning.

El papel que desempeñó el doctorando durante el desarrollo del proyecto GCPTG fueron los siguientes; (i) desarrollar las plataformas que faciliten la gestión de los datos de los pacientes con enfermedades terminales y trastornos genéticos (tal como síndrome de Down); (ii) mejorar la forma en que se brindan cuidados paliativos a los pacientes ancianos y a los pacientes con enfermedades terminales, facilitando el trabajo de los especialistas de la salud.; (iii) mejorar el proceso de estimación del riesgo de síndrome de Down con predicciones más precisas basadas en los parámetros de pacientes mujeres locales. Durante el desarrollo de este proyecto, el doctorando adquirió las competencias necesarias para el campo de investigación.

4.2 Sistema de Gestión Inteligente para la Optimización del Consumo Energético en la Climatización de Edificios (GIOCEC)

El proyecto denominado Sistema de Gestión Inteligente para la Optimización del Consumo Energético en la Climatización de Edificios (GIOCEC) es un proyecto de la Universidad de Alcalá de Henares en colaboración con la Universidad Rey Juan Carlos.

El objetivo de este proyecto es el de desarrollar un sistema autónomo comercialmente viable que permita implantarse en los sistemas de gestión de edificios (BMS) de tal modo que gobierne el funcionamiento de la instalación de climatización (HVAC). Los retos principales del desarrollo son la implementación de un sistema de supervisión inteligente basado en una optimización multi-objetivo (MO) del consumo de energía, el confort, el rendimiento y el costo, mediante modelos de predicción híbridos (Grey Box) construidos sobre una simplificación de la base matemática, cuyos parámetros se ajusten dinámicamente con modelos de datos. Para la construcción del modelo matemático se empleará una aproximación original denominada multi-HVAC que reducirá la complejidad matemática de la no linealidad de estos sistemas. Para el modelo de datos se empleará una red neuronal (ANN) o algún algoritmo de Machine Learning (ML), según se pruebe el rendimiento. Para la optimización se probarán algoritmos de optimización evolutivos. Para la gestión inteligente se aplicarán modelos de datos dinámicos obtenidos del estado supervisado de los equipos de la instalación y de condicionantes como climatología, ocupación y uso del edificio, grado de confort requerido, tarifas eléctricas horarias, etc; de forma que permitan gobernar autónomamente la propia instalación, sino la auto-protección, auto-configuración y auto-optimización del software.

Para lograr estos objetivos el doctorando desarrolló e implementó diferentes técnicas de Soft-Computing, en la cual le permitieron identificar aquellos métodos más relevantes para el cumplimiento de los objetivos planteados. Asimismo, el grado de innovación del proyecto planteado le permitió al doctorando ganar un premio en la convocatoria Premios 2019 – Campus de Excelencia Internacional (CEI) "Energía Inteligente", donde este ganó el premio en la categoría 1 de proyectos de investigación colaborativos bajo el galardón ACCÉSIT CEI. Por otro lado, este proyecto también le permitió al doctorando ganar otro permio CEI (2022) en el cual fue galardonado en la categoría 1: mejor aportación científica con la publicación científica "Developing a Long Short-Term Memory-Based Model for Forecasting the Daily Energy Consumption of Heating, Ventilation, and Air Conditioning Systems in Buildings". Por consiguiente, el doctorando adquirió diferentes capacidades en su labor investigadora durante el desarrollo del proyecto GIOCEC.

4.3 Prediction of the Appearance of Clinical Hemodynamic Complications in Hemodialysis: An Approximation Based on Massive Data Analysis (PACHAMDA)

El proyecto Prediction of the Appearance of Clinical Hemodynamic Complications in Hemodialysis: An Approximation Based on Massive Data Analysis (PACHAMDA) fue desarrollado en colaboración de un equipo de investigación multidisciplinar de la Universidad de Alcalá de Henares. De igual manera, este proyecto fue avalado por el Hospital Príncipe de Asturias de Madrid, España.

El proyecto PACHAMDA planteó la siguiente hipótesis de investigación: Es posible detectar una combinación de parámetros clínicos y analíticos asociados con la aparición de hipotensión o el desarrollo de insuficiencia cardíaca. Si alguno de estos parámetros fuera ajustable, solo o en combinación, más allá del estado actual de la técnica, la incidencia del problema podría reducirse. En este sentido, se establecieron los siguientes objetivos específicos:

- Detectar la combinación mínima de parámetros clínicos y analíticos asociados a la aparición de hipotensiones.
- Detectar la combinación mínima de parámetros clínicos y analíticos asociados al desarrollo de insuficiencia cardíaca.
- Demostrar que la modificación de estos parámetros disminuye la incidencia de hipotensiones e insuficiencia cardíaca

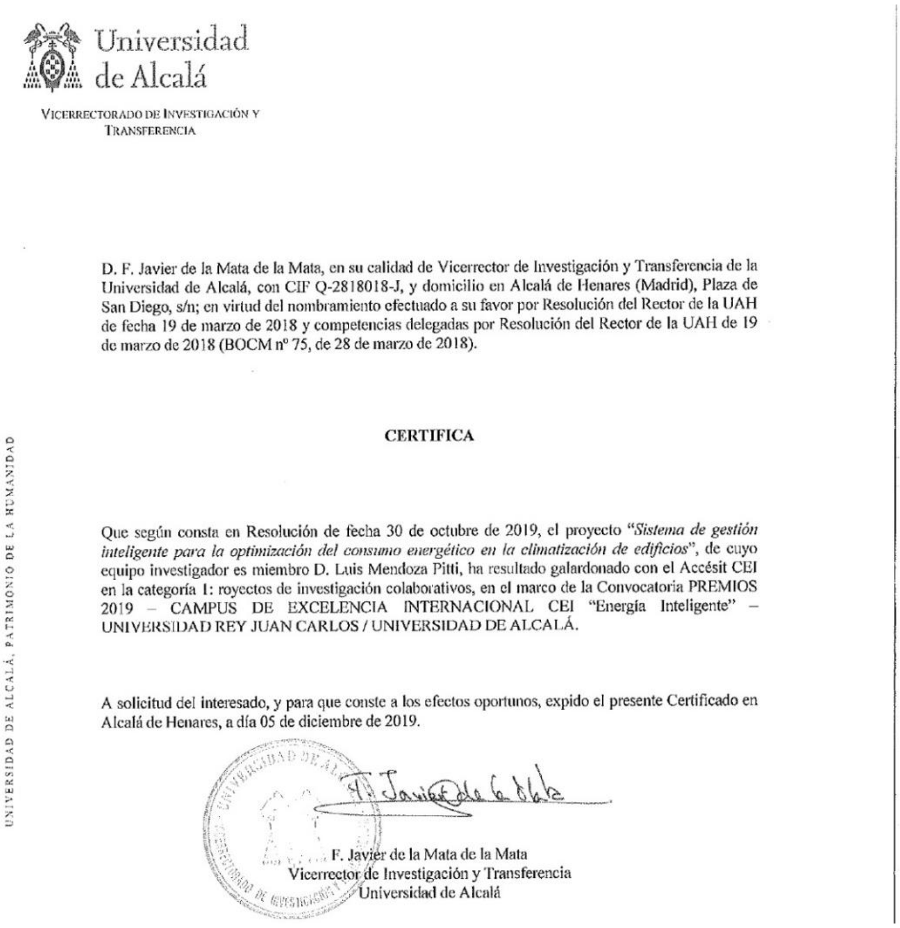
Para conseguir alcanzar los objetivos planteados fue necesario automatizar el proceso de depuración de las bases de datos que mantenían los parámetros clínicos y analíticos asociados a la aparición de hipotensiones. Asimismo, el doctorando desarrollo diferentes modelos de machine learning con la finalidad de identificar aquel modelo que fuera capaz de predecir y clasificar si el paciente padeciese de una hipotensión o no durante toda la sesión de hemodiálisis a realizar. Esto le permitió al doctorando adquirir mayor competencia investigadora en el ámbito de las técnicas de Soft-Computing ideales para este caso de estudio.

5 Méritos alcanzados

5.1 Premios 2019 - Campus de Excelencia Internacional CEI "Energía Inteligente"

El rector de la Universidad Rey Juan Carlos y el Vicerrector de Investigación y Transferencia de la Universidad de Alcalá de Henares, en marco de la premiación al Campus de Excelencia Internacional (CEI) "Energía Inteligente", concedieron el premio ***Accésit CEI*** al proyecto *"Sistema de gestión inteligente para la optimización del consumo energético en la climatización de edificios"* en la categoría 1 de Proyectos de Investigación Colaborativos con una cuantía de 4000 euros.

Este proyecto plantea integrar y desplegar servicios inteligentes autónomos en el BMS del edificio, mediante una arquitectura de software versátil e interoperable que propicie la eficiencia energética de los sistemas HVAC de los edificios. En la Figura 5 se observa el certificado de participación del doctorando firmado por el Vicerrector de Investigación y Transferencia de la Universidad de Alcalá de Henares.



Universidad de Alcalá

VICERRECTORADO DE INVESTIGACIÓN Y TRANSFERENCIA

D. F. Javier de la Mata de la Mata, en su calidad de Vicerrector de Investigación y Transferencia de la Universidad de Alcalá, con CIF Q-2818018-J, y domicilio en Alcalá de Henares (Madrid), Plaza de San Diego, s/n; en virtud del nombramiento efectuado a su favor por Resolución del Rector de la UAH de fecha 19 de marzo de 2018 y competencias delegadas por Resolución del Rector de la UAH de 19 de marzo de 2018 (BOCM nº 75, de 28 de marzo de 2018).

CERTIFICA

Que según consta en Resolución de fecha 30 de octubre de 2019, el proyecto "*Sistema de gestión inteligente para la optimización del consumo energético en la climatización de edificios*", de cuyo equipo investigador es miembro D. Luis Mendoza Pitti, ha resultado galardonado con el Accésit CEI en la categoría 1: royectos de investigación colaborativos, en el marco de la Convocatoria PREMIOS 2019 – CAMPUS DE EXCELENCIA INTERNACIONAL CEI "Energía Inteligente" – UNIVERSIDAD REY JUAN CARLOS / UNIVERSIDAD DE ALCALÁ.

A solicitud del interesado, y para que conste a los efectos oportunos, expido el presente Certificado en Alcalá de Henares, a día 05 de diciembre de 2019.

F. Javier de la Mata de la Mata
Vicerrector de Investigación y Transferencia
Universidad de Alcalá

UNIVERSIDAD DE ALCALÁ, PATRIMONIO DE LA HUMANIDAD

Figura 5. Certificado del premio al Campus de Excelencia Internacional (CEI) "Energía Inteligente" – Accesít CEI.

5.2 Premios 2022 - Campus de Excelencia Internacional CEI "Energía Inteligente"

El rector de la Universidad Rey Juan Carlos y el Vicerrector de Investigación y Transferencia de la Universidad de Alcalá de Henares, en marco de la premiación al Campus de Excelencia Internacional (CEI) "Energía Inteligente", concedieron el premio CEI en la ***categoría 1***: ***mejor aportación científica UAH*** a la publicación llamada *"Developing a Long Short-Term Memory-Based Model for Forecasting the Daily Energy Consumption of Heating, Ventilation, and Air Conditioning Systems in Buildings"* en la cual se otorgó una cuantía de 3000 euros repartidos de forma equitativa a los autores participantes del trabajo. En la Figura 6 se observa la resolución de participación del doctorando en los premios CEI, 2022.

Categoría 1: Mejor aportación científica

PREMIO CEI MEJOR APORTACIÓN CIENTÍFICA UAH: Concedido a la propuesta número **14** "***Developing a Long Short-Term Memory-Based Model for Forecasting the Daily Energy Consumption of Heating, Ventilation, and Air Conditioning Systems in Buildings***" presentada por los Prof. José Luis Castillo Sequera, José Manuel Gómez Pulido, Luis Mendoza Pitti y Clara Simón de Blas, Cuantía del premio de **3.000 €.** Dicho importe se repartirá, a partes iguales, según se indica a continuación:

- José Luis Castillo Sequera: 750 €.
- José Manuel Gómez Pulido: 750 €.
- Luis Mendoza Pitti: 750 €.
- Clara Simón de Bla: 750 €.

Figura 6. Resolución de la mejor aportación científica en los premios CEI 2022.

6 Conclusiones y líneas futuras de investigación

Basándonos en las secciones anteriormente presentados, se aprecian el compendio de publicaciones que comprenden esta tesis doctoral, los cuales justifican los resultados obtenidos tras el desarrollo del trabajo de investigación del doctorando. Estas publicaciones fueron difundidas tanto en revistas como congresos de alto prestigio internacional la cual validan su labor investigadora. A continuación, se describen las principales conclusiones de los trabajos publicados.

6.1 Conclusiones

En la primera publicación se realizó una revisión sistemática de la literatura donde se analizaron las diferentes arquitecturas de software orientada a servicios que propician la gestión de servicios avanzados en las dimensiones claves de la Smart Energy y la Smart Health. De modo que, se logró identificar dos patrones arquitectónicos dirigidos a las dimensiones clave estudiadas. La arquitectura SOA fue identificada como una arquitectura capaz de ofrecer la característica de interoperabilidad para los diferentes componentes del edificio; sin embargo, esta presenta muchas limitaciones al momento de desarrollar e implementar nuevas funcionalidades [75]. De allí, pues que, la arquitectura MSA puede suplir los inconvenientes encontrados en SOA, ya que ofrece diversas ventajas para desarrollar, implementar, gestionar y desplegar servicios (microservicios) y, además, puede emplearse bajo diferentes lenguajes de programación, bases de datos, entre otros recursos computacionales. En consecuencia, estas dos arquitecturas (SOA y MSA) en conjunto proporcionarían un ecosistema más confiable, versátil, interoperable, ágil, escalable y robusto para la implementación de servicios avanzados en las dimensiones claves de la sociedad digital.

La segunda publicación muestra los resultados obtenidos tras desarrollar un sistema que simula el comportamiento y operaciones de los sistemas Multi-HVAC de un edificio en concreto en España a través del uso de modelos físicos simplificados. Esto permitió establecer aquellos puntos o momentos de optimización que mejorarían de manera significativa la eficiencia energética en los sistemas HVAC reales. Por lo que, esta primera fase del trabajo impulsó el desarrollo de servicios más avanzados que mejoren considerablemente la eficiencia energética, así como la disminución del consumo energético en los sistemas HVAC de un edificio real, mediante el uso de técnicas de Soft-Computing novedosas.

La tercera publicación muestra un modelo basado en Deep Learning (DL) dirigido a pronosticar el consumo energético diario de los sistemas HVAC en edificios. Esto se consiguió utilizando el modelo Long Short-Term Memory (LSTM), debido a que tiene una gran capacidad para captar el comportamiento dinámico y no-lineal propio del contexto de los sistemas HVAC [76]. Por lo que, los resultados obtenidos en este estudio destacan la importancia de implementar servicios avanzados basados en técnicas de Soft-Computing, pues estos permiten maximizar la eficiencia energética y reducir el impacto ambiental, debido a que ofrece a los propietarios o administradores de los edificios una forma de tomar de decisiones de manera más inteligente e informada [77].

La cuarta publicación muestra como las plataformas de eHealth están migrando desde un paradigma clásico a uno ubicuo (servicios) [78], [79]. Puesto que, una gestión deficiente de los datos de un paciente conllevaría a múltiples problemas, por lo que, es necesario emplear estándares que faciliten la gestión de la información clínica de los pacientes que reciben cuidados o presenten un trastorno genético. En este sentido, la adopción de las nuevas tecnologías está permitiendo que el entorno de salud implemente nuevas arquitecturas que les facilite la gestión, análisis y procesamiento de los datos masivos generados por los parámetros de los pacientes. En consecuencia, la adopción de las técnicas de Machine Learning (ML), la computación ubicua y la vida asistida por el entorno brindan un amplio acceso al conocimiento generado por estos enfoques, por lo que, los especialistas (médicos o enfermeras) podrán tomar decisiones más informadas, y como consecuencia mejorando los tratamientos y cuidados a los pacientes.

La quinta publicación muestra los resultados obtenidos tras emplear un servicio avanzado mediante un modelo de ML, en el cual este logró predecir si el paciente padecerá o no de hipotensión al inicio de la sesión de hemodiálisis. Para conseguir esto se utilizó una combinación de datos masivos generados por los propios parámetros clínicos y de laboratorio de los pacientes, ya que estos fueron fuertemente determinantes en la aparición de casos de hipotensos durante la sesión de hemodiálisis. En este sentido, ofrecer servicios avanzados en el campo de la medicina, permite que los ciudadanos tengan un mayor acceso a una atención médica confiable y accesible, generando bienestar y una mejor calidad de vida para las personas. Asimismo, estos servicios avanzados ayudan a los especialistas a diagnosticar enfermedades y encontrar aquellos tratamientos adecuados para los pacientes, pues estos les proporciona información para tomar decisiones inteligentes.

Como complemento, cada una de las dimensiones claves de la Smart City aquí estudiados (Smart Energy y Smart Health) son capaces de aprovechar todas las ventajas ofrecidas por las diferentes tecnologías avanzadas analizadas e identificadas en nuestro estudio, pues estas permiten implementar servicios avanzados con un alto valor añadido a la sociedad digital. Por ejemplo, el uso mediante inteligencia artificial para gestionar de manera inteligente y eficaz los diferentes componentes del edificio, logrando así que el edificio sea más sostenible tras reducir tanto las emisiones al medio ambiente como el costo energético producido por los componentes del edificio. Otro ejemplo sería, el desarrollo de servicios avanzados que analicen y predigan el comportamiento patológico y evolutivo de una enfermedad a través de métodos avanzados de inteligencia artificial, por lo que permitiría ofrecer una mejor calidad de vida a los pacientes crónicos o ancianos, buscando siempre el bienestar de los ciudadanos.

En conclusión, esta tesis doctoral ha expuesto distintos hallazgos de gran relevancia científica y que han sido constatados por los expertos científicos internacionales, a través de la difusión de cinco publicaciones de alto impacto que resaltan la importancia de la construcción e implementación de los servicios avanzados, con el fin de mejorar el bienestar de la ciudadanía y la sostenibilidad ambiental en las dimensiones claves de la sociedad digital (Smart Energy y Smart Health). Por tanto, esta investigación aporta un fuerte argumento sobre la validez del desarrollo de servicios avanzados que facilitan la toma de decisiones inteligentes en las dimensiones claves de la sociedad digital.

6.2 Líneas futuras de investigación

Al momento de desarrollar las diferentes investigaciones se lograron identificar oportunidades para mejorar los trabajos futuros. En este sentido, algunas líneas futuras de investigación a desarrollar de este trabajo se encuentran las siguientes: (i) automatizar la selección de características de los datos involucrados en los servicios avanzados, ya que constantemente se generan más datos proporcionados por las diversas fuentes en las dimensiones claves de la sociedad digital aquí estudiadas (Smart Energy y Smart Health), por lo que es necesario automatizar estos procesos de manera inteligente; (ii) automatizar sin intervención humana el ajuste de los hiperparámetros de los servicios avanzados que se desarrollan, puesto que los datos que se generan en estas dimensiones claves de la sociedad digital evolucionan con el paso del tiempo, por tanto, estos requieren reentrenarse para mantener la funcionalidad en la toma de decisiones informadas de los actores principales de los sectores ya mencionados; (iii) desplegar los servicios avanzados dentro de los novedosos patrones arquitectónicos de software identificados, dado que permiten que estos puedan ser utilizados de manera versátil y flexible por los tomadores de decisiones relevantes en las dimensiones de la sociedad digital.

7 Referencias

[1] U. Nations, "World Cities Report 2020: The Value of Sustainable Urbanization," 2020. Accessed: Oct. 28, 2021. [Online]. Available: https://unhabitat.org/.

[2] A. M. Toli and N. Murtagh, "The Concept of Sustainability in Smart City Definitions," *Front. Built Environ.*, vol. 6, p. 77, 2020, doi: 10.3389/fbuil.2020.00077.

[3] E. O'Dwyer, I. Pan, S. Acha, and N. Shah, "Smart energy systems for sustainable smart cities: Current developments, trends and future directions," *Appl. Energy*, vol. 237, pp. 581–597, 2019, doi: https://doi.org/10.1016/j.apenergy.2019.01.024.

[4] M. Mohamed, "Challenges and Benefits of Industry 4.0: an overview," *Int. J. Supply Oper. Manag.*, vol. 5, no. 3, pp. 256–265, 2018, doi: 10.22034/2018.3.7.

[5] T. L. Olsen and B. Tomlin, "Industry 4.0: Opportunities and Challenges for Operations Management," *Manuf. Serv. Oper. Manag.*, vol. 22, no. 1, pp. 113–122, Sep. 2019, doi: 10.1287/msom.2019.0796.

[6] M. Lom, O. Pribyl, and M. Svitek, "Industry 4.0 as a part of smart cities," in *2016 Smart Cities Symposium Prague (SCSP)*, 2016, pp. 1–6, doi: 10.1109/SCSP.2016.7501015.

[7] A. Kirimtat, O. Krejcar, A. Kertesz, and M. F. Tasgetiren, "Future Trends and Current State of Smart City Concepts: A Survey," *IEEE Access*, vol. 8, pp. 86448–86467, 2020, doi: 10.1109/ACCESS.2020.2992441.

[8] B. N. Silva, M. Khan, and K. Han, "Towards sustainable smart cities: A review of trends, architectures, components, and open challenges in smart cities," *Sustainable Cities and Society*, vol. 38. pp. 697–713, Apr. 2018, doi: 10.1016/j.scs.2018.01.053.

[9] V. Albino, U. Berardi, and R. M. Dangelico, "Smart Cities: Definitions, Dimensions, Performance, and Initiatives," *J. Urban Technol.*, vol. 22, no. 1, pp. 3–21, Jan. 2015, doi: 10.1080/10630732.2014.942092.

[10] J. Winkowska, D. Szpilko, and S. Pejić, "Smart city concept in the light of the literature review," *Eng. Manag. Prod. Serv.*, vol. 11, no. 2, pp. 70–86, 2019, doi: doi:10.2478/emj-2019-0012.

[11] J. Tian, H. Li, and R. Chen, "The emerging of smart citizen concept under smart city environment," in *The 18th International Conference on Electronic Business, ICEB*, 2018, pp. 739–742.

[12] M. Magalhães, R. P. Duarte, C. Oliveira, and F. C. Pinto, "The Role of the Smart Citizen in Smart Cities," in *Computational Science and Its Applications – ICCSA 2021*, 2021, pp. 295–310.

[13] A. Alexandru, M. Ianculescu, I. A. Marinescu, and T. D. Popescu, "Shaping the Digital Citizen into a Smart Citizen on the Basis of IoT Capabilities," in *2019 22nd International Conference on Control Systems and Computer Science (CSCS)*, 2019, pp. 707–714, doi: 10.1109/CSCS.2019.00126.

[14] M. Ianculescu, O. Bica, A. Balog, and I. Cristescu, "Smart citizen - a participatory co-creator for enhancing online public services," in *2019 11th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, 2019, pp. 1–6, doi: 10.1109/ECAI46879.2019.9041953.

[15] M. Thornbush and O. Golubchikov, "Smart energy cities: The evolution of the city-energy-sustainability nexus," *Environ. Dev.*, vol. 39, p. 100626, 2021, doi: https://doi.org/10.1016/j.envdev.2021.100626.

[16] A. Corsi, F. Florencio de Souza, R. N. Pagani, and J. L. Kovaleski, "Ultimate approach and technologies in smart healthcare: A broad systematic review focused on citizens," *Smart Heal.*, vol. 26, p. 100310, 2022, doi: https://doi.org/10.1016/j.smhl.2022.100310.

[17] P. O. Omolaye, J. M. Mom, and G. A. Igwue, "A holistic review of soft computing techniques," *Appl. Comput. Math.*, vol. 6, no. 2, p. 93, 2017.

[18] A. Sharma, M. Kumar, and S. Agarwal, "A Complete Survey on Software Architectural Styles and Patterns," *Procedia Comput. Sci.*, vol. 70, pp. 16–28, 2015, doi: 10.1016/j.procs.2015.10.019.

[19] I. C. L. Ng and S. Y. L. Wakenshaw, "The Internet-of-Things: Review and research directions," *Int. J. Res. Mark.*, vol. 34, no. 1, pp. 3–21, 2017, doi:

https://doi.org/10.1016/j.ijresmar.2016.11.003.
[20] A. Oussous, F.-Z. Benjelloun, A. Ait Lahcen, and S. Belfkih, "Big Data technologies: A survey," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 30, no. 4, pp. 431–448, 2018, doi: 10.1016/j.jksuci.2017.06.001.
[21] Y. Zhong, L. Sun, and C. Ge, "Key Technologies and Development Status of Smart City," in *Journal of Physics: Conference Series*, 2021, vol. 1754, no. 1, p. 12102, doi: 10.1088/1742-6596/1754/1/012102.
[22] A. Camero and E. Alba, "Smart City and information technology: A review," *Cities*, vol. 93, pp. 84–94, 2019, doi: https://doi.org/10.1016/j.cities.2019.04.014.
[23] I. Dominiković, M. Ćukušić, and M. Jadrić, "The Role of Artificial Intelligence in Smart Cities: Systematic Literature Review," *Lect. Notes Inst. Comput. Sci. Soc. Telecommun. Eng. LNICST*, vol. 378, pp. 64–80, 2021, doi: 10.1007/978-3-030-77417-2_5.
[24] C. Lim, K.-J. Kim, and P. P. Maglio, "Smart cities with big data: Reference models, challenges, and considerations," *Cities*, vol. 82, pp. 86–99, 2018, doi: https://doi.org/10.1016/j.cities.2018.04.011.
[25] H. Lund, P. A. Østergaard, D. Connolly, and B. V. Mathiesen, "Smart energy and smart energy systems," *Energy*, vol. 137, pp. 556–565, 2017, doi: https://doi.org/10.1016/j.energy.2017.05.123.
[26] B. Qolomany *et al.*, "Leveraging Machine Learning and Big Data for Smart Buildings: A Comprehensive Survey," *IEEE Access*, vol. 7, pp. 90316–90356, 2019, doi: 10.1109/ACCESS.2019.2926642.
[27] L. Pérez-Lombard, J. Ortiz, and C. Pout, "A review on buildings energy consumption information," *Energy Build.*, vol. 40, no. 3, pp. 394–398, Feb. 2008, doi: 10.1016/j.enbuild.2007.03.007.
[28] D. S. Vijayan, A. L. Rose, S. Arvindan, J. Revathy, and C. Amuthadevi, "Automation systems in smart buildings: a review," *J. Ambient Intell. Humaniz. Comput.*, 2020, doi: 10.1007/s12652-020-02666-9.
[29] Y. Cao, J. Du, and E. Soleymanzadeh, "Model predictive control of commercial buildings in demand response programs in the presence of thermal storage," *J. Clean. Prod.*, vol. 218, pp. 315–327, 2019, doi: https://doi.org/10.1016/j.jclepro.2019.01.266.
[30] G. Aceto, V. Persico, and A. Pescapé, "Industry 4.0 and Health: Internet of Things, Big Data, and Cloud Computing for Healthcare 4.0," *J. Ind. Inf. Integr.*, vol. 18, p. 100129, 2020, doi: https://doi.org/10.1016/j.jii.2020.100129.
[31] S. Gambhir, S. K. Malik, and Y. Kumar, "Role of Soft Computing Approaches in HealthCare Domain: A Mini Review," *J. Med. Syst.*, vol. 40, no. 12, p. 287, 2016, doi: 10.1007/s10916-016-0651-x.
[32] Z. Rayan, M. Alfonse, and A.-B. M. Salem, "Machine Learning Approaches in Smart Health," *Procedia Comput. Sci.*, vol. 154, pp. 361–368, 2019, doi: https://doi.org/10.1016/j.procs.2019.06.052.
[33] M. F. Slon Roblero *et al.*, "Experience in Spain with the first patients in home hemodialysis treated with low-flow dialysate monitors," *Nefrología*, vol. 42, no. 4, pp. 460–470, 2022, doi: https://doi.org/10.1016/j.nefro.2021.07.001.
[34] J. Gilissen, L. Pivodic, K. T. Unroe, and L. Van den Block, "International COVID-19 Palliative Care Guidance for Nursing Homes Leaves Key Themes Unaddressed," *J. Pain Symptom Manage.*, vol. 60, no. 2, pp. e56–e69, 2020, doi: https://doi.org/10.1016/j.jpainsymman.2020.04.151.
[35] L. A. Zadeh, "Soft computing and fuzzy logic," *IEEE Softw.*, vol. 11, no. 6, pp. 48–56, 1994, doi: 10.1109/52.329401.
[36] D. Ibrahim, "An Overview of Soft Computing," *Procedia Comput. Sci.*, vol. 102, pp. 34–38, 2016, doi: https://doi.org/10.1016/j.procs.2016.09.366.
[37] S. Tamboli and L. Bewoor, "A review of soft computing technique for real-time data forecasting," 2019.
[38] S. Binitha and S. S. Sathya, "A survey of bio inspired optimization algorithms," *Int. J. soft Comput. Eng.*, vol. 2, no. 2, pp. 137–151, 2012.
[39] D. Sharma and P. Chandra, "A comparative analysis of soft computing techniques in software fault prediction model development," *Int. J. Inf. Technol.*, vol. 11, no. 1, pp. 37–46, 2019, doi: 10.1007/s41870-018-0211-3.
[40] F. A. O. Santos, G. S. de Jesus, G. A. Botelho, and H. T. Macedo, "Smart health: Using

fuzzy logic in the monitoring of health-related indicators," in *2016 8th Euro American Conference on Telematics and Information Systems (EATIS)*, 2016, pp. 1–4, doi: 10.1109/EATIS.2016.7520131.
[41] R. Mehta, "Multivariate Fuzzy Logic Based Smart Healthcare Monitoring for Risk Evaluation of Cardiac Patients," in *Medical Informatics and Bioimaging Using Artificial Intelligence : Challenges, Issues, Innovations and Recent Developments*, A. E. Hassanien, R. Bhatnagar, V. Snášel, and M. Yasin Shams, Eds. Cham: Springer International Publishing, 2022, pp. 219–243.
[42] A. Berouine *et al.*, "A Fuzzy Logic-Based Approach for HVAC Systems Control," in *2019 6th International Conference on Control, Decision and Information Technologies (CoDIT)*, 2019, pp. 1510–1515, doi: 10.1109/CoDIT.2019.8820356.
[43] H. Chaouch, C. Çeken, and S. Arı, "Energy management of HVAC systems in smart buildings by using fuzzy logic and M2M communication," *J. Build. Eng.*, vol. 44, p. 102606, 2021, doi: https://doi.org/10.1016/j.jobe.2021.102606.
[44] A. H. Alkeshuosh, M. Z. Moghadam, I. A. Mansoori, and M. Abdar, "Using PSO Algorithm for Producing Best Rules in Diagnosis of Heart Disease," in *2017 International Conference on Computer and Applications (ICCA)*, 2017, pp. 306–311, doi: 10.1109/COMAPP.2017.8079784.
[45] D. Tien Bui, H. Moayedi, D. Anastasios, and L. Kok Foong, "Predicting Heating and Cooling Loads in Energy-Efficient Buildings Using Two Hybrid Intelligent Models," *Applied Sciences*, vol. 9, no. 17. 2019, doi: 10.3390/app9173543.
[46] S. Y. Kung, *Kernel Methods and Machine Learning*. Cambridge: Cambridge University Press, 2014.
[47] A. Géron, *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. O'Reilly Media, 2019.
[48] M. Goyal and M. Pandey, "Towards Prediction of Energy Consumption of HVAC Plants Using Machine Learning," in *Data Science and Analytics*, 2020, pp. 254–265.
[49] P. Kaur, R. Kumar, and M. Kumar, "A healthcare monitoring system using random forest and internet of things (IoT)," *Multimed. Tools Appl.*, vol. 78, no. 14, pp. 19905–19916, 2019, doi: 10.1007/s11042-019-7327-8.
[50] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.
[51] K. Singh and J. Malhotra, "Deep learning based smart health monitoring for automated prediction of epileptic seizures using spectral analysis of scalp EEG," *Phys. Eng. Sci. Med.*, vol. 44, no. 4, pp. 1161–1173, 2021, doi: 10.1007/s13246-021-01052-9.
[52] J. Shafi, M. S. Obaidat, P. V. Krishna, B. Sadoun, M. Pounambal, and J. Gitanjali, "Prediction of heart abnormalities using deep learning model and wearabledevices in smart health homes," *Multimed. Tools Appl.*, vol. 81, no. 1, pp. 543–557, 2022, doi: 10.1007/s11042-021-11346-5.
[53] A. Afram, F. Janabi-Sharifi, A. S. Fung, and K. Raahemifar, "Artificial neural network (ANN) based model predictive control (MPC) and optimization of HVAC systems: A state of the art review and case study of a residential HVAC system," *Energy Build.*, vol. 141, pp. 96–113, 2017, doi: https://doi.org/10.1016/j.enbuild.2017.02.012.
[54] T.-Y. Kim and S.-B. Cho, "Predicting residential energy consumption using CNN-LSTM neural networks," *Energy*, vol. 182, pp. 72–81, 2019, doi: https://doi.org/10.1016/j.energy.2019.05.230.
[55] A. Al-Qamash, I. Soliman, R. Abulibdeh, and M. Saleh, "Cloud, Fog, and Edge Computing: A Software Engineering Perspective," in *2018 International Conference on Computer and Applications (ICCA)*, 2018, pp. 276–284, doi: 10.1109/COMAPP.2018.8460443.
[56] A. Yousefpour *et al.*, "All one needs to know about fog computing and related edge computing paradigms: A complete survey," *J. Syst. Archit.*, vol. 98, pp. 289–330, 2019, doi: https://doi.org/10.1016/j.sysarc.2019.02.009.
[57] P. Mell and T. Grance, "The NIST definition of cloud computing," 2011.
[58] O. Consortium, "The OpenFog Consortium Reference Architecture: Executive Summary." 2017.
[59] A. Bouguettaya *et al.*, "A service computing manifesto: the next 10 years," *Commun. ACM*, vol. 60, no. 4, pp. 64–72, May 2017, doi: 10.1145/2983528.

[60] R. C. A. Hutapea, A. P. Wahyudi, and Suhardi, "Design Quality Measurement for Service Oriented Software on Service Computing System: a Systematic Literature Review," in *2018 International Conference on Information Technology Systems and Innovation (ICITSI)*, 2018, pp. 375–380, doi: 10.1109/ICITSI.2018.8696092.
[61] M. P. Papazoglou and W.-J. van den Heuvel, "Service oriented architectures: approaches, technologies and research issues," *VLDB J.*, vol. 16, no. 3, pp. 389–415, 2007, doi: 10.1007/s00778-007-0044-3.
[62] H. Calderón-Gómez *et al.*, "Evaluating Service-Oriented and Microservice Architecture Patterns to Deploy eHealth Applications in Cloud Computing Environment," *Applied Sciences* , vol. 11, no. 10. 2021, doi: 10.3390/app11104350.
[63] J. Lewis and M. Fowler, "Microservices: a definition of this new architectural term," 2014. https://martinfowler.com/articles/microservices.html (accessed Feb. 16, 2022).
[64] B. M. C. Silva, J. J. P. C. Rodrigues, A. Ramos, K. Saleem, I. de la Torre, and R. L. Rabêlo, "A Mobile Health System to Empower Healthcare Services in Remote Regions," in *2019 IEEE International Conference on E-health Networking, Application & Services (HealthCom)*, 2019, pp. 1–6, doi: 10.1109/HealthCom46333.2019.9009477.
[65] R. T. Hameed, O. A. Mohamad, O. T. Hamid, and N. Ţăpuş, "Patient monitoring system based on e-health sensors and web services," in *2016 8th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, 2016, pp. 1–6, doi: 10.1109/ECAI.2016.7861089.
[66] F. M. Garcia-Moreno, M. Bermudez-Edo, J. L. Garrido, E. Rodríguez-García, J. M. Pérez-Mármol, and M. J. Rodríguez-Fórtiz, "A Microservices e-Health System for Ecological Frailty Assessment Using Wearables," *Sensors* , vol. 20, no. 12. 2020, doi: 10.3390/s20123427.
[67] S. Roca, J. Sancho, J. García, and Á. Alesanco, "Microservice chatbot architecture for chronic patient support," *J. Biomed. Inform.*, vol. 102, p. 103305, 2020, doi: https://doi.org/10.1016/j.jbi.2019.103305.
[68] P. Angulo, C. C. Guzmán, G. Jiménez, and D. Romero, "A service-oriented architecture and its ICT-infrastructure to support eco-efficiency performance monitoring in manufacturing enterprises," *Int. J. Comput. Integr. Manuf.*, vol. 30, no. 1, pp. 202–214, Jan. 2017, doi: 10.1080/0951192X.2016.1145810.
[69] A. Haque, R. Rahman, and S. Rahman, "Microservice-based Architecture of a Software as a Service (SaaS) Building Energy Management Platform," in *2020 6th IEEE International Energy Conference (ENERGYCon)*, 2020, pp. 967–972, doi: 10.1109/ENERGYCon48941.2020.9236617.
[70] K. Bao, I. Mauser, S. Kochanneck, H. Xu, and H. Schmeck, "A Microservice Architecture for the Intranet of Things and Energy in Smart Buildings: Research Paper," 2016, doi: 10.1145/3007203.3007215.
[71] M. A. Jarwar, S. Ali, and I. Chong, "Microservices based Linked Data Quality Model for Buildings Energy Management Services," *CoRR*, vol. abs/1910.0, 2019, [Online]. Available: http://arxiv.org/abs/1910.06115.
[72] T. G. Stavropoulos, E. S. Rigas, E. Kontopoulos, N. Bassiliades, and I. Vlahavas, "A Multi-agent Coordination Framework for Smart Building Energy Management," in *2014 25th International Workshop on Database and Expert Systems Applications*, 2014, pp. 126–130, doi: 10.1109/DEXA.2014.39.
[73] S. Vardoulakis and P. Kinney, "Grand Challenges in Sustainable Cities and Health," *Front. Sustain. Cities*, vol. 1, p. 7, 2019, doi: 10.3389/frsc.2019.00007.
[74] Universidad de Alcalá, "Reglamento de elaboración, autorización y defensa de la Tesis Doctoral. Aplicación del RD 99/2011, de 28 de enero. BOE 10 de febrero de 2011. Aprobado en la Comisión de Estudios Oficiales de Posgrado y la Comisión de Doctorado en Sesión de 18 de enero de ." 2011.
[75] J. Kazanavičius and D. Mažeika, "Migrating Legacy Software to Microservices Architecture," in *2019 Open Conference of Electrical, Electronic and Information Sciences (eStream)*, 2019, pp. 1–5, doi: 10.1109/eStream.2019.8732170.
[76] C. Fan, J. Wang, W. Gang, and S. Li, "Assessment of deep recurrent neural network-based strategies for short-term building energy predictions," *Appl. Energy*, vol. 236, pp. 700–710, 2019, doi: https://doi.org/10.1016/j.apenergy.2018.12.004.
[77] I. Bouzguenda, C. Alalouch, and N. Fava, "Towards smart sustainable cities: A review of

the role digital citizen participation could play in advancing social sustainability," *Sustain. Cities Soc.*, vol. 50, p. 101627, 2019, doi: https://doi.org/10.1016/j.scs.2019.101627.
[78] A. R. Javed *et al.*, "Future smart cities: requirements, emerging technologies, applications, challenges, and future aspects," *Cities*, vol. 129, p. 103794, 2022, doi: https://doi.org/10.1016/j.cities.2022.103794.
[79] N. A. Megahed and R. F. Abdel-Kader, "Smart Cities after COVID-19: Building a conceptual framework through a multidisciplinary perspective," *Sci. African*, vol. 17, p. e01374, 2022, doi: https://doi.org/10.1016/j.sciaf.2022.e01374.

8 Lista de figuras