Sergio Nesmachnow · Luis Hernández Callejo
Editors

Smart Cities

5th Ibero-American Congress, ICSC-CITIES 2022
Cuenca, Ecuador, November 28–30, 2022
Revised Selected Papers

Springer

Editors
Sergio Nesmachnow
Universidad de la República
Montevideo, Uruguay

Luis Hernández Callejo
Universidad de Valladolid
Soria, Spain

ISSN 1865-0929 ISSN 1865-0937 (electronic)
Communications in Computer and Information Science
ISBN 978-3-031-28453-3 ISBN 978-3-031-28454-0 (eBook)
https://doi.org/10.1007/978-3-031-28454-0

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This CCIS volume presents selected articles from the 5th edition of the Ibero-American Congress of Smart Cities (ICSC-CITIES 2022), held on November 28–30, 2022 in Cuenca, Ecuador, in a mixed modality, with in-person and online talks and article presentations. This event continues the successful four previous editions of the congress, held in Soria, Spain in 2018 and 2019, Costa Rica in 2020, and Cancún, México, in 2021.

The main goal of ICSC-CITIES 2022 was to provide a forum for researchers, scientists, teachers, decision-makers, postgraduate students and practitioners from different countries in Ibero-America and worldwide to share their current initiatives related to Smart Cities. Articles in this volume address four relevant topics (computational intelligence and urban informatics for smart cities; Internet of Things; optimization, smart production, and smart public services; and smart monitoring and communications) covering several areas of research and applications.

The main program consisted of three round tables, 72 oral presentations and 12 poster presentations from international speakers, highlighting recent developments in areas related to smart cities. Over three hundred distinguished participants from 28 countries gathered presentially or virtually for the congress. The Program Committee of ICSC-CITIES 2022 received 116 manuscripts. 72 submissions were accepted for oral presentation and the best 18 whose contents are within the Computer and Information Science areas were selected to be published in this CCIS volume. All articles have undergone a careful single-blind peer-review process by at least three subject-matter experts before being selected for publication.

We would like to express our deep gratitude to all the contributors to ICSC-CITIES 2022, the congress organizers, and to the authors and reviewers for their endeavors that made the paper-reviewing process efficient and convenient. We also thank the participants of the congress, our institutions, and all readers of this CCIS volume.

December 2022

Sergio Nesmachnow
Luis Hernández Callejo

Organization

General Chairs

Luis Hernández Callejo	Universidad de Valladolid, Spain
Sergio Nesmachnow	Universidad de la República, Uruguay

Track Chairs

Fabián Castillo Peña	Universidad Libre Seccional Cali, Colombia
Ponciano Jorge Escamilla Ambrosio	Instituto Politécnico Nacional, México
Paulo Gondim	Universidade de Brasília, Brazil
Luis Eduardo Tobón	Pontificia Universidad Javeriana, Cali, Colombia
Bernabé Dorronsoro	Universidad de Cádiz, Spain
Ana Ruiz	Universidad San Jorge, Cátedra Mobility Experience, Spain
Humberto Sossa	Instituto Politécnico Nacional, México
Jesús Armando Aguilar Jiménez	Universidad Autónoma de Baja California, México
Victor Alonso Gómez	Universidad de Valladolid, Spain
Sara Gallardo Saavedra	Universidad de Valladolid, Spain
Luis Hernández Callejo	Universidad de Valladolid, Spain
Roberto Villafáfila Robles	Universidad Politécnica de Cataluña, Spain
Irene Lebrusan	Instituto Complutense de Sociología para el Estudio de las Transformaciones Sociales Contemporáneas, Spain
Andrés Navarro	Pontificia Universidad Javeriana, Cali, Colombia
Adriana Correa Guimaraes	Universidad de Valladolid, Spain
Óscar Duque Pérez	Universidad de Valladolid, Spain
Luis Manuel Navas	Universidad de Valladolid, Spain
Jorge Luis Mírez Tarrillo	Universidad Nacional de Ingeniería, Perú
Lorena Parra	Universitat Politècnica de València, Spain
Carlos E. Torres Aguilar	Centro Nacional de Investigación y Desarrollo Tecnológico, México
Daniel Rossit	Universidad Nacional del Sur, Argentina
Pedro Moreno Bernal	Universidad Autónoma del Estado de Morelos, México

Jesús del Carmen Peralta Abarca	Universidad Autónoma del Estado de Morelos, México
Ana Carolina Olivera	CONICET, Argentina
Sergio Nesmachnow	Universidad de la República, Uruguay
Diego Rossit	Universidad Nacional del Sur, Argentina
Mariano Frutos	Universidad Nacional del Sur, Argentina
Teodoro Calonge Cano	Universidad de Valladolid, Spain
Ponciano Jorge Escamilla Ambrosio	Instituto Politécnico Nacional, México
Renzo Massobrio	Universidad de la República, Uruguay and Universidad de Cádiz, Spain
Luis Angelo Steffenel	Université de Reims Champagne-Ardenne, France
Jamal Toutouh	Universidad de Málaga, Spain

Publication Chairs

Luis Hernández Callejo	Universidad de Valladolid, Spain
Sergio Nesmachnow	Universidad de la República, Uruguay
Pedro Moreno Bernal	Universidad Autónoma del Estado de Morelos, México
Diego Rossit	Universidad Nacional del Sur, Argentina

Submission and Conference Management Chair

Santiago Iturriaga	Universidad de la República, Uruguay

Program Committee

Adolfo Ruelas Puente	Universidad Autónoma de Baja California, México
Adrian Toncovich	Universidad Nacional del Sur, Argentina
Adriana Correa-Guimaraes	Universidad de Valladolid, Spain
Agustín Laguarda	Universidad de la República, Uruguay
Alejandro Otero	Universidad de Buenos Aires-CONICET, Argentina
Alejandro Paz Parra	Pontificia Universidad Javeriana, Colombia
Alessandra Bussador	Universidade Federal da Integração Latino-Americana, Brazil

Alexander Vallejo Díaz	Santo Domingo Institute of Technology, Dominican Republic
Alexis Acuna	Universidad Autónoma de Baja California, México
Ana Carolina Olivera	Universidad Nacional de Cuyo, Argentina
Ana Ruiz	San Jorge University, Spain
Andrei Tchernykh	CICESE Research Center, México
Andrés Adolfo Navarro Newball	Pontificia Universidad Javeriana, Colombia
Ángel Zorita Lamadrid	Universidad de Valladolid, Spain
Angela Ferreira	Polytechnic Institute of Bragança, Portugal
Antonio Mauttone	Universidad de la República, Uruguay
Antonio Muñoz	University of Málaga, Spain
Armando Huicochea	Research Center for Engineering and Applied Science, México
Bernabe Dorronsoro	Universidad de Cádiz, Spain
Bernardo Pulido-Gaytan	CICESE Research Center, México
Bouras Abdelkarim	Badji Mokhtar University, Algeria
Carlos Grande	Universidad Centroamericana José Simeón Cañas, El Salvador
Carlos Meza Benavides	Anhalt University of Applied Sciences, Germany
Carlos E. Torres	Tecnológico Nacional de México/CENIDET, México
Christian Cintrano	Universidad de Málaga, Spain
Claudio Paz	Universidad Tecnológica Nacional - FRC, Argentina
Claudio Risso	Universidad de la República, Uruguay
Cleonilson Protasio	Federal University of Paraíba, Portugal
Cristina Sáez Blázquez	University of Salamanca, Spain
Daniel Rossit	Universidad Nacional del Sur, Argentina
Daniel H. Stolfi	University of Luxembourg, Luxembourg
David Peña Morales	Universidad de Cádiz, Spain
Deyslen Mariano	Instituto Tecnológico de Santo Domingo, Dominican Republic
Diego Arcos-Aviles	Universidad de las Fuerzas Armadas ESPE, Ecuador
Diego Alberto Godoy	Universidad Gastón Dachary, Argentina
Diego Gabriel Rossit	Universidad Nacional del Sur-CONICET, Argentina
Edgardo Aníbal Belloni	Universidad Gastón Dachary, Argentina
Edith Gabriela Manchego Huaquipaco	Universidad Nacional de San Agustín de Arequipa, Perú
Eduardo Fernández	Universidad de la República, Uruguay
Emmanuel Millan	UNCuyo, Argentina

Enrique González	Universidad de Salamanca, Spain
Esteban Mocskos	Universidad de Buenos Aires, Argentina
Fabian Castillo Peña	Universidad Libre - Cali, Colombia
Francisco David Moya Chaves	Universidad Distrital Francisco José de Caldas, Colombia
Francisco Valbuena	Universidad de Valladolid, Spain
Franco Robledo	Universidad de la República, Uruguay
Gilberto Martinez	CIC-IPN, México
Gina Paola Maestre Gongora	Universidad Cooperativa de Colombia, Colombia
Gustavo Richmond-Navarro	Instituto Tecnológico de Costa Rica, Costa Rica
Ignacio de Godos	Universidad de Valladolid, Spain
Ignacio Martín Nieto	Universidad de Salamanca, Spain
Irene Lebrusán	Harvard University, USA
Ivania Aguirre	University of Cuenca, Ecuador
Jamal Toutouh	Massachusetts Institute of Technology, USA
Javier Rocher	Universitat Politècnica de València, Spain
Jesús Vegas	Universidad de Valladolid, Spain
Jesús Armando Aguilar	Autonomous University of Baja California, México
Jesús Del Carmen Peralta Abarca	Universidad Autónoma del Estado de Morelos, México
Joao Coelho	Instituto Politécnico de Bragança, Portugal
Jonathan Muraña	UDELAR, Uruguay
Jorge Arturo Del Ángel Ramos	Universidad Veracruzana, México
Jorge Mario Cortés-Mendoza	University of Luxembourg, Luxembourg
Jorge Mírez	Universidad Nacional de Ingeniería, Perú
Jorge Nájera	Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas, Spain
José Alberto Hernández	Universidad Autónoma del Estado de Morelos, México
José Ángel Morell Martínez	Universidad de Málaga, Spain
José Antonio Ferrer	CIEMAT, Spain
José-Ramón Aira	Universidad de Valladolid, Spain
Juan Chavat	Universidad de la República, Uruguay
Juan Espinoza	Universidad de Cuenca, Ecuador
Juan Francisco Cabrera Sánchez	Universidad de Cádiz, Spain
Juan Humberto Sossa Azuela	CIC-IPN, México
Juan José Tarrio	Comisión Nacional de Energía Atómica, Argentina
Juan Manuel Ramírez Alcaraz	Universidad de Colima, México
Juan Mauricio	Universidade Federal da Paraíba, Brazil
Juan Pavón	Universidad Complutense de Madrid, Spain

Juan R. Coca	Universidad de Valladolid, Spain
Leonardo Cardinale	Instituto Tecnológico de Costa Rica, Costa Rica
Lilian Obregon	Universidad de Valladolid, Spain
Lorena Parra	Universitat Politècnica de València, Spain
Luis G. Montané Jiménez	Universidad Veracruzana, México
Luis Garcia	Universidad de Concepción, Chile
Luis Hernandez-Callejo	Universidad de Valladolid, Spain
Luis Manuel Navas Gracia	Universidad de Valladolid, Spain
Luis Marrone	UNLP, Argentina
Luis Tobon	Pontificia Universidad Javeriana Cali, Colombia
Luisena Fernandez	Universidad del Zulia, Venezuela
Manuel Gonzalez	Universidad de Valladolid, Spain
Marcin Seredynski	E-Bus Competence Center, France
Maria Teresa Cepero	Universidad Veracruzana, México
Mariano Frutos	Universidad Nacional del Sur, Argentina
Mario Andrés Paredes Valverde	TecNM/ITS de Teziutlán, México
Melva Inés Gómez Caicedo	Fundación Universitaria Los Libertadores, Colombia
Miguel Davila	Universidad Politécnica Salesiana, Ecuador
Miguel Angel Che Pan	Tecnológico Nacional de México/CENIDET, México
Monica Alonso	Universidad Carlos III de Madrid, Spain
Mónica Montoya Giraldo	CIDET, Colombia
Neiel Israel Leyva Santes	Barcelona Supercomputing Center, Spain
Noelia Soledad Pinto	Universidad Tecnológica Nacional, Argentina
Noelia Uribe-Perez	Tecnalia Research & Innovation, Spain
Oscar Duque-Perez	University of Valladolid, Spain
Oscar Izquierdo	CEDER-CIEMAT, Spain
Pablo Daniel Godoy	Universidad Nacional de Cuyo, Argentina
Pablo Monzon	Universidad de la República, Uruguay
Paul Ayala	Universidad de las Fuerzas Armadas, Ecuador
Paula de Andrés Anaya	Escuela Politécnica Superior de Ávila, Spain
Paula Peña Carro	CIEMAT, Spain
Paulo Gondim	Universidade de Brasília, Brazil
Pedro Curto	Universidad de la República, Uruguay
Pedro Moreno Bernal	Universidad Autónoma del Estado de Morelos, México
Pedro Piñeyro	Universidad de la República, Uruguay
Ponciano Jorge Escamilla-Ambrosio	CIC-IPN, México
Ramiro Martins	Polytechnic Institute of Bragança, Portugal
Raúl Alberto López Meraz	Universidad Veracruzana, México

Renato Andara	Universidad Nacional Experimental Politécnica Antonio José de Sucre, Venezuela
Renzo Massobrio	Universidad de la República, Uruguay
Ricardo Beltrán	Centro de Investigación en Materiales Avanzados, México
Roberto Villafafila	Universitat Politècnica de Catalunya, Spain
Rodrigo Porteiro	UTE, Uruguay
Rogelio Vargas Lopez	Universidad Autónoma de Guadalajara, México
Santiago Iturriaga	Universidad de la República, Uruguay
Sara Gallardo-Saavedra	Universidad de Valladolid, Spain
Sergio Nesmachnow	Universidad de la República, Uruguay
Sesil Koutra	University of Mons, Belgium
Silvia Soutullo	CIEMAT, Spain
Teodoro Calonge	Universidad de Valladolid, Spain
Teresa Batista	Universidade de Évora, Portugal
Tiago Carneiro Pessoa	INRIA, France
Vanessa Guimarães	CEFET/RJ, Brazil
Vicente Leite	Instituto Politécnico de Bragança, Portugal
Vicente Canals	Universidad de las Islas Baleares, Spain
Víctor Alonso Gómez	Universidad de Valladolid, Spain
Víctor Manuel Padrón Nápoles	Universidad Europea de Madrid, Spain

Reviewers

Abelardo Contreras Panibra	Universidad Nacional de Ingeniería, Perú
Adalberto Ospino	Universidad de la Costa, Colombia
Adolfo Heriberto Ruelas Puente	Universidad Autónoma de Baja California, México
Adrián Toncovich	Universidad Nacional del Sur, Argentina
Adrian Criollo	Universidad Politécnica Salesiana, Ecuador
Alberto Redondo	Universidad de Valladolid, Spain
Alejandro Paz Parra	Universidad Santiago de Cali, Colombia
Alejandro César López Bolaños	Universidad Nacional Autónoma de México, México
Alejandro Otero	Universidad de Buenos Aires-CONICET, Argentina
Alessandra Bussador	Centro Universitário Dinâmica das Cataratas, Brazil
Alexander Vallejo Díaz	Instituto Tecnológico de Santo Domingo, Dominican Republic
Alexis Acuña Ramírez	Universidad Autónoma de Baja California, México

Alfonso García Alvaro	Universidad de Valladolid, Spain
Alicia Martínez Rebollar	Tecnológico Nacional de México/Centro Nacional de Investigación y Desarrollo Tecnológico, México
Alvaro Jaramillo Duque	Universidad de Antioquia, Colombia
Amaia Arrinda	University of the Basque Country, Spain
Américo Vicente Teixeira Leite	Instituto Politécnico de Bragança, Portugal
Ana Ruiz Varona	Universidad San Jorge, Spain
Ana Carolina Olivera	CONICET-Universidad Nacional de Cuyo, Argentina
Ana Lilia Sanchéz Brito	Tecnológico Nacional de México/Centro Nacional de Investigación y Desarrollo Tecnológico, México
Andrei Tchernykh	CICESE Research Center, Mexico
Andrés Adolfo Navarro Newball	Pontificia Universidad Javeriana Cali, Colombia
Andrés Llombart	CIRCE - Research Centre, Spain
Andres Montero-Izquierdo	Universidad de Cuenca, Ecuador
Ângela Ferreira	Instituto Politécnico de Bragança, Portugal
Angeles Dennis Figueroa-Negrete	Universidad Autónoma del Estado de Morelos, México
Antonella Cavallin	Universidad Nacional del Sur, Argentina
Antonio Daniel Mauttone Vidales	Universidad de la República, Uruguay
Araceli Ávila Hernández	Centro Nacional de Investigación y Desarrollo Tecnológico, México
Armando Huicochea Rodríguez	Universidad Autónoma del Estado de Morelos, México
Bernabe Dorronsoro	Universidad de Cádiz, Spain
Bernardo Pulido Gaytan	CICESE Research Center, México
Bouras Abdelkarim	Badji Mokhtar University, Algeria
Carlos Enrique Torres Aguilar	Tecnológico Nacional de México/Centro Nacional de Investigación y Desarrollo Tecnológico, México
Carlos Miguel Jiménez-Xamán	Tecnológico Nacional de México/Centro Nacional de Investigación y Desarrollo Tecnológico, México
Carmen Alonso-Garcia	CIEMAT - Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas, Spain
Carolina Solis Maldonado	Universidad Veracruzana, México
César Alejandro Varela Boydo	Centro Nacional de Investigación y Desarrollo Tecnológico, México
Cindy Espinoza Aguirre	Universidad de Cuenca, Ecuador
Claudio J. Paz	Universidad Tecnológica Nacional - FRC, Argentina

Consuelo Gines Palestino	Tecnológico Nacional de México Campus Tepexi, México
Cristina Sáez Blázquez	Universidad de León, Spain
Daniel Rossit	Universidad Nacional del Sur-CONICET, Argentina
Daniel Stolfi	SnT - University of Luxembourg, Luxembourg
Daniel Morinigo-Sotelo	Universidad de Valladolid, Spain
Danielle Rodrigues de Moraes	Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, Brazil
David Peña Morales	Universidad de Cádiz, Spain
David de la Vega	University of the Basque Country, Spain
David Guillermo Pasillas Banda	Universidad de Guadalajara, México
David Peña	Universidad de Cádiz, Spain
David Zabala-Pedraza	Instituto Tecnológico de Santo Domingo, Dominican Republic
Deyslen Mariano Hernández	Instituto Tecnológico de Santo Domingo, Dominican Republic
Diana Sánchez-Partida	Universidad Popular Autónoma del Estado de Puebla, México
Diego Rossit	Universidad Nacional del Sur-CONICET, Argentina
Diego Arcos Avilés	Universidad de las Fuerzas Armadas ESPE, Ecuador
Diego Alberto Godoy	Universidad Gastón Dachary, Argentina
Diego Enrique Morales Polanco	Instituto Tecnológico de Santo Domingo, Dominican Republic
Diego Luis Linares	Pontificia Universidad Javeriana, Colombia
Edgar Vázquez Beltrán	Universidad de Sonora, México
Edgar Vicente Macias Melo	Universidad Juárez Autónoma de Tabasco, México
Edith Gabriela Manchego Huaquipaco	Universidad Nacional de San Agustín de Arequipa, Perú
Egor Shiriaev	North-Caucasus Federal University, Russia
Enrique González González	Universidad de Salamanca, Spain
Esteban Zalamea	Universidad de Cuenca, Ecuador
Estefani Mendoza	Universidad Nacional de Ingeniería, Perú
Fabián Castillo Peña	Universidad Libre de Colombia, Colombia
Fabio Miguel	Universidad Nacional de Río Negro, Argentina
Francisco Chicano	Universidad de Málaga, Spain
Francisco José Sánchez Pacheco	Universidad de Málaga, Spain
Fredy Dulce	Universidad del Valle, Colombia
Gabriel Marro	Universidad San Jorge, Spain

Genesis Sanchez-Sosa	Instituto Tecnológico de Santo Domingo, Dominican Republic
Gilberto Martinez	CIC-IPN, México
Gina Maestre Gongora	Universidad Cooperativa de Colombia, Colombia
Guillermo Pérez	Escuela Politécnica Nacional, Ecuador
Gustavo Richmond-Navarro	Instituto Tecnológico de Costa Rica, Costa Rica
Gustavo Adolfo Gómez Ramírez	Instituto Tecnológico de Costa Rica, Costa Rica
Gustavo Navarro	Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas, Spain
H. Amaris	Universidad Carlos III de Madrid, Spain
Harold Vacca	Universidad Distrital Francisco José de Caldas, Colombia
Hector Felipe Mateo Romero	Universidad de Valladolid, Spain
Heidi Paola Díaz Hernández	Universidad Juárez Autónoma de Tabasco, México
Hugo Estrada	Tecnológico Nacional de México/Centro Nacional de Investigación y Desarrollo Tecnológico, México
Hugo J. Bello	Universidad de Valladolid, Spain
Hugo Sanchez	Anhalt University of Applied Sciences, Germany
Ignaciode Godos Crespo	Universidad de Valladolid, Spain
Irene Lebrusán Murillo	Universidad de Salamanca/Universidad Carlos III de Madrid, Spain
Irene Acosta	Universidad Regional Amazónica, Ecuador
Irene Marisol Revelo Portilla	Pontificia Universidad Católica de Ecuador, Ecuador
Irma Yazmín Hernández Báez	Universidad Politécnica del Estado de Morelos, México
Ismael Minchala	Universidad de Cuenca, Ecuador
Ítalo Bove	Universidad de la República, Uruguay
Itziar Angulo	University of the Basque Country, Spain
Ivan Alonso de Miguel	Universitat de les Illes Balears, Spain
Ivania Carolina Aguirre Pardo	Universidad de Cuenca, Ecuador
Ivonne Yazmín Arce García	Universidad Autónoma del Estado de Morelos, México
J. Guadalupe Velásquez-Aguilar	Universidad Autónoma del Estado de Morelos, México
Jacinto Vidal-Noguera	Universitat de les Illes Balears, Spain
Jaime Paul Ayala	Universidad de las Fuerzas Armadas ESPE, Ecuador
Jamal Toutouh El Alamin	Universidad de Málaga, Spain
Javier Rocher Morant	Universitat Politècnica de València, Spain

Jesús Armando Aguilar Jiménez	Universidad Autónoma de Baja California, México
Jesús del Carmen Peralta Abarca	Universidad Autónoma del Estado de Morelos, México
João Paulo Coelho	Instituto Politécnico de Bragança, Portugal
Jonathan Muraña SIlvera	Universidad de la República, Uruguay
Jorge Nájera Álvarez	CIEMAT, Spain
Jorge del Ángel	Universidad Veracruzana, México
Jorge Gomez-Sanz	Universidad Complutense de Madrid, Spain
Jorge López-Rebollo	Universidad de Salamanca, Spain
Jorge Mírez	Universidad Nacional de Ingeniería, Perú
Jorge-Mario Cortes-Mendoza	University of Luxembourg, Luxembourg
José Alberto Hernández Aguilar	Universidad Autónoma del Estado de Morelos, México
José Ángel Morell Martínez	Universidad de Málaga, Spain
José Antonio Ferrer Tévar	Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas, Spain
Jose Duran	Instituto Tecnológico de Santo Domingo, Dominican Republic
Jose Ignacio Morales Aragonés	Colegio Oficial de Ingenieros de Telecomunicación, Spain
José Ramón Aira Zunzunegui	Universidad Politécnica de Madrid, Spain
Jose-Isidro Hernandez-Vega	Tecnológico Nacional de México, México
Juan Pavón	Universidad Complutense de Madrid, Spain
Juan Carlos Frechoso	Universidad de Valladolid, Spain
Juan Francisco Cabrera Sánchez	Universidad de Cádiz, Spain
Juan Gabriel Ordoñez	Universidad de Nariño, Colombia.
Juan Irving Vasquez Gomez	Instituto Politécnico Nacional, México
Juan Manuel Ramírez Alcaraz	Universidad de Colima, México
Juan Moises Mauricio Villanueva	Universidade Federal da Paraíba, Brazil
Juan Pablo Chavat Pérez	Universidad de la República, Uruguay
Juan Rodríguez	Universidad de la República, Uruguay
Karla María Aguilar Castro	Universidad Juárez Autónoma de Tabasco, México
Laene Oliveira Soares	Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, Brazil
Laura Rosa Guerra Torrealba	Pontificia Universidad Católica de Ecuador, Ecuador
Leonardo Cardinale Villalobos	Tecnológico de Costa Rica, Costa Rica
Lidia Sanz-Molina	Universidad de Valladolid, Spain
Lluc Crespí-Castañer	Universitat de les Illes Balears, Spain
López Cintrano	Universidad de Málaga, Spain

Luis Antonio González Uribe	Universidad Autónoma de Baja California, México
Luis Armando Marrone	Universidad Nacional de La Plata, Argentina
Luis Tobón	Pontificia Universidad Javeriana Cali, Colombia
Luis Gerardo Montané Jiménez	Universidad Veracruzana, México
Luis González Uribe	Universidad Autónoma de Baja California, México
Luis Manuel Navas Gracia	Universidad de Valladolid, Spain
Luis Miguel Quishpe	Universidad Regional Amazónica, Ecuador
Ma. Eloisa Gurruchaga Rodríguez	Instituto Tecnológico de Orizaba, México.
Manuel Ángel González Delgado	Universidad de Valladolid, Spain
Marcelo García Luyo	Universidad Nacional de Ingeniería, Perú
Marcia Eugenio	Universidad de Valladolid, Spain
Marcos Blanco	Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas, Spain
María Sánchez Aparicio	Universidad de Salamanca, Spain
María Teresa Cepero García	Universidad Veracruzana, México
Maria Teresa Folgôa Batista	Universidade de Évora e CIMAC, Portugal
Maria Victoria Reyes Vargas	Universidad Regional Amazónica, Ecuador
Martín Draper	Universidad de la República, Uruguay
Martin Solis Solis	Instituto Tecnológico de Costa Rica, Costa Rica
Matías Galnares	Universidad de la República, Uruguay
Melva Inés Gomez Caicedo	Fundación Universitaria los Libertadores, Colombia
Mercedes Gaitan	Fundación Universitaria Konrad Lorenz, Colombia
Miguel Dávila	Universidad de Valladolid, Spain
Miguel Angel Che-Pan	Tecnológico Nacional de México/Centro Nacional de Investigación y Desarrollo Tecnológico, México
Miguel Angel González	Universidad de Valladolid, Spain
Miguel Angel Muñoz Garcia	Universidad Politécnica de Madrid, Spain
Miguel Aybar	Instituto Tecnológico de Santo Domingo, Dominican Republic
Miguel Mendieta	Universidad Politécnica Salesiana, Ecuador
Miriam Navarrete Procopio	Universidad Autónoma del Estado de Morelos, México
Miriam González Dueñas	Universidad de Guadalajara, México
Moisés Montiel González	Universidad Autónoma del Estado de Morelos, México
Natalia Nuño-Villanueva	Universidad de Salamanca, Spain
Neiel Israel Leyva Santes	Barcelona Supercomputing Center, Spain

Noelia Pinto	Universidad Tecnológica Nacional, Argentina, Argentina
Noelia Uribe-Perez	Tecnalia Research & Innovation, Spain
Olga Quevedo	Universidad de Guayaquil, Ecuador
Óscar Duque Pérez	Universidad de Valladolid, Spain
Oscar Izquierdo Monge	Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas, Spain
Pablo Monzon	Universidad de la República, Uruguay
Pablo Collazzo	University for Continuing Education Krems, Austria
Pablo Daniel Godoy	Universidad Nacional de Cuyo, Argentina
Pablo de Frutos	Universidad de Valladolid, Spain
Paola Duque Sarango	Universidad Politécnica Salesiana, Ecuador
Paolo Piantanida	Politecnico di Torino, Itália
Patricia Ruiz Villalobos	University of Cadiz, Spain
Paul Ayala	Universidad de las Fuerzas Armadas ESPE, Ecuador
Paula Guerra	Kennesaw State University, USA
Paula Peña Carro	Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas, Spain
Paula de Andrés Anaya	Universidad de Salamanca, Spain
Paulo R. L. Gondim	Universidade de Brazilia, Brazil
Pedro Moreno-Bernal	Universidad Autónoma del Estado de Morelos, México
Pedro Antonio Martin-Cervantes	Universidad de Valladolid, Spain
Ramiro José Espinheira Martins	Superior School of Technology and Management-IPB, Portugal
Ramsés David Villalobos Núñez	Universidad Nacional de Ingeniería, Perú
Raúl Alberto López Meraz	Unidad de Ingeniería y Ciencias Químicas/ Universidad Veracruzana, México
Raúl Maján Navalón	Tierra Sin Males, Spain
Renato Alejandro Andara Escalona	Universidad Nacional Experimental Politécnica Antonio José de Sucre, Venezuela
Renzo Massobrio	Delft University of Technology, Netherlands
Ricardo Beltran-Chacon	Centro de Investigación en Materiales Avanzados, Mexico.
Ricardo Peña	Instituto Tecnológico de Santo Domingo, Dominican Republic
Roberto Villafáfila Robles	Universidad Politécnica de Cataluña, Spain
Rodolfo Omar Domínguez García	Universidad de Guadalajara, México
Rogelio Vargas López	Universidad Autónoma de Guadalajara, México
Ronney Boloy	Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, Brazil

Ruby Gines Palestino	Tecnológico Nacional de México, México
Ryszard Edward Rozga Luter	Universidad Autónoma Metropolitana, México
Samanta López Salazar	Tecnológico Nacional de México/Centro Nacional de Investigación y Desarrollo Tecnológico, México
Samuel Borroy	CIRCE Research Centre, Spain
Santiago Damian Quishpe Morales	Pontificia Universidad Católica de Ecuador, Ecuador
Santiago Pindado	Universidad Politécnica de Madrid, Spain
Sara Gallardo Saavedra	Universidad de Valladolid, Spain
Saúl Islas Pereda	Universidad Autónoma de Baja California, México
Silvia Soutullo Castro	CIEMAT, Spain
Sinuhe Gines Palestino	Tecnológico Nacional de México, México
Susana Gómez Redondo	Universidad de Valladolid, Spain
Teodoro Calonge Cano	Universidad de Valladolid, Spain
Thania Guadalupe Lima Tellez	Universidad de Sonora, México
Tiago Carneiro Pessoa	Interuniversity Microelectronics Centre, Belgium
Vanessa de Almeida Guimarães	Centro Federal de Educação Tecnológica Celso Suckow da Fonseca - CEFET/RJ, Brazil
Verónica Pinos-Vélez	Universidad de Cuenca, Ecuador
Vicente José Canals Guinand	Universidad de las Islas Baleares, Spain
Víctor Alonso Gómez	Universidad de Valladolid, Spain
Víctor Hernández Andrés	Universidad de Valladolid, Spain
Víctor Manuel Padrón Nápoles	Universidad Europea de Madrid, Spain
Víctor Martínez	Universitat de les Illes Balears, Spain
Yanitzia Reyes Cárcamo	Tecnológico Nacional de México, México
Yuri Humberto Merizalde Zamora	Universidad de Guayaquil, Ecuador

Contents

Big Data Trends in the Analysis of City Resources

Regina Gubareva[1] and Rui Pedro Lopes[1,2(✉)]

[1] Research Center in Digitalization and Intelligent Robotics (CeDRI), Instituto Politécnico de Bragança, Bragança, Portugal
{regina.gubareva,rlopes}@ipb.pt

[2] Institute of Electronics and Informatics Engineering of Aveiro (IEETA), University of Aveiro, Aveiro, Portugal

Abstract. The operation and management of a municipality generate large amounts of complex data, enclosing information that is not easy to infer or extract. Their analysis is challenging and requires specialized approaches and tools, usually based on statistical techniques or on machine learning and artificial intelligence algorithms. These Big Data is often created by combining many data sources that correspond to different operational groups in the city, such as transport, energy consumption, water consumption, maintenance, and many others. Each group exhibits unique characteristics that are usually not shared by others. This paper provides a detailed systematic literature review on applying different algorithms to urban data processing. The study aims to figure out how this kind of information was collected, stored, pre-processed, and analyzed, to compare various methods, and to select feasible solutions for further research. The review finds that clustering, classification, correlation, anomaly detection, and prediction algorithms are frequently used. Moreover, the interpretation of relevant and available research results is presented.

Keywords: Big data · Smart city · Resources consumption

1 Introduction

The smart city concept is popular and common in scientific literature, characterizing a healthy environment that improves the quality of life and well-being of citizens [10]. Due to the diversity of services, resources, and projects, smart cities manage huge amounts of data, typically within the Big Data concept. One can argue what are the minimum conditions and characteristics for a city to become "smart". However, since nowadays most operations are controlled via comprehensive information and communication technologies, the need to collect, store, integrate, process, and analyze data is prevalent and important in most cities. Over the past ten years, the number of sensors and metering devices has been increasing geometrically. The intention to control and understand everything surrounding us became a significant step in the development of technologies of environmental sensors: smart houses, smart cities, smart devices, IoT,

S. Nesmachnow and L. Hernández Callejo (Eds.): ICSC-CITIES 2022, CCIS 1706, pp. 215–229, 2023.
https://doi.org/10.1007/978-3-031-28454-0_15

and many others. Legacy information is also laying around, in spreadsheets or databases, which can be valuable if correctly accessed and integrated. Citizens and institutions also make use of social networks to convey opinions, criticism, or information about resources, services, or events. The essential questions are how to use this data and how to extract practical and meaningful information from all these measurements. Big data is a set of technologies for processing large amounts of data. It refers not only to the amount of information but also to the data rate, meaning the multiple streams of data that should be processed in real-time. Moreover, large examples of data usually enclose hidden, potentially valuable, patterns. Several unique phenomena associated with high dimensionality, including noise accumulation, spurious correlation, and random endogeneity, make traditional statistical procedures difficult to use. In the Big Data era, large sample sizes allow us to better understand heterogeneity by shedding light on research such as examining the relationship between specific covariates and rare outcomes.

This work is developed within the project "PandIA - Management of Pandemic Social Isolation Based on City and Social Intelligence", which focus on providing detailed information, such as resource consumption trends, estimation of people in each area or household, a heat map of suspected outbreaks, and others to health and municipal authorities and to emergency personal. For that, it uses information from several sources, including pathogen characteristics, infection statistics, municipal information, social networks, and hospital information and statistics. The work described in this paper uses a systematic literature review to understand the nature and purpose of the data generated and collected in the context of a city. It aims to understand what types of data are usually considered, how they what collected, what algorithms are used, and for what purposes. We look for evidence and best practices for using city information in Big Data settings, the impact, and results. The authors do not set out to compare the algorithms with mathematical certainty, we just review various approaches and provide rough estimates of effectiveness. The article should give a base understanding of what to do with smart city data. Broadly speaking, the purpose of this paper is to systematize the basic principles of digital data handling in the formation and development of smart cities. The review summarizes the research being done for the last five years. The literature is categorized according to the algorithms used, the approach to handling data, the nature of data, and the results of the data processing.

The paper is structured in four sections, starting with this introduction. Section 2 describes the methodology followed in this study. The Result and Analysis follow, with the results and associated discussion and it finishes in Sect. 4 with some conclusions.

2 Methodology

The main objective of this literature review is to try to understand the data structure and nature, their sources, the processes of collecting and storing, the algorithms and tasks these are developed to do, and, finally, the purposes or intentions of the results. This literature review follows the approach suggested by Materla, Cudney, and Antony [11] and by Subhash and Cudney [14], including three phases: planning, operation, and dissemination (Fig. 1).

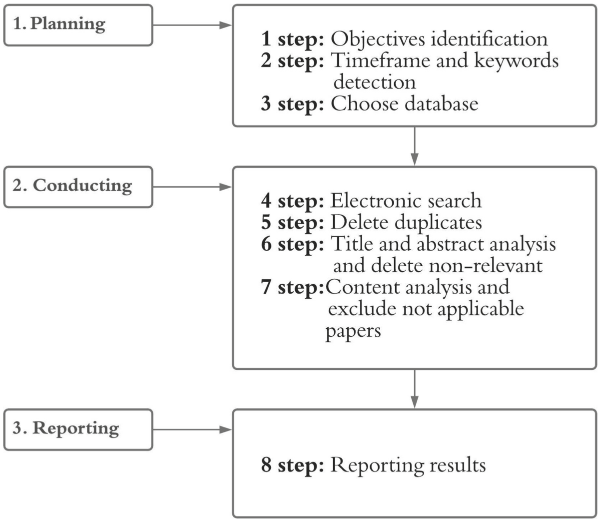

Fig. 1. Phases of the systematic literature review.

To guide the search, a set of research questions was considered:

1. What is the format, structure, and context of the data each paper uses?
2. What algorithms are used for urban data handling?
3. What is the complexity of each algorithm?
4. What are the results achieved after the analysis and what do they mean?

The papers were searched in Scopus and IEEEXplore. These databases were selected because they provide a wide set of areas and the key terms provide an initial focus on the main objective of this work. A total of 230 papers were identified in the first run (Table 1).

Only the papers retrieved from Scopus and IEEEXplore published between January 1st, 2015, and December 31st, 2020, whose text was available in the institutional repositories were considered. Moreover, papers without a peer review process and written in a different language than English were also excluded. After removing the duplicate entries, the total number of papers was 208.

Some guidelines were defined for the title, abstract, and text analysis. After a primary assessment, a detailed analysis of conformity to a chosen theme was made. The title's meaningfulness, associated with the abstract description helped with this. Next, the text was skimmed to assess if all the information needed could be found. So, in summary, the papers were analyzed according to the following steps:

Table 1. Search terms and the number of papers retrieved

Database	Search term	Results
Scopus	TITLE ("big data" AND ("urban data" OR "smart city" OR "geo data" OR "social network" OR "predictive maintenance" OR "algorithms"))	123
IEEEXplore	(("Document Title": "big data") AND ("Document Title": "urban data" OR "Document Title": "smart city" OR "Document Title": "geo data" OR "Document Title": "social network" OR "Document Title": "predictive maintenance" OR "Document Title": "algorithms"))	107
	Total	230

1. Searched articles were limited to the predefined time frame (2016–2021)
2. Papers with non-relevant titles, abstracts, and keywords were excluded
3. Text that did not mention the required subjects was excluded

After analysis of the title, abstract, and text, 196 more papers were excluded for being out of the scope of this work. A total of 12 papers remained for the analysis (Table 2).

Table 2. The results of the search by journals

Phase	Scopus	IEEEXplore	Total
search	123	107	230
del. duplicates	114	94	208
title	42	36	78
abstract	11	8	19
content	7	5	12

3 Analysis and Discussion

The analysis process started with the characterization of the selected references. Content analysis followed, to assess the context and definition of self-study and the purpose of the work described in the paper.

3.1 Characterization

In total, papers from 9 countries were found. Spain has 3 papers and China 2 papers. For United Arab Emirates, Denmark, Malaysua, South Korea, Taiwan, India and Greece, a single paper was found. It should be highlighted that the countries' distribution does not connect with cities' development in these regions.

The distribution by year reveals a peak number of papers in 2018, with 7 papers. The lowest amount of research is observed in 2019 (a single paper), which can reflect the exploration of new approaches in the area. The remaining years (2017 and 2016, two papers were accounted). Despite considerable attention from the scientific community to the issue of the impact of such a resource as digital data on the development of modern socio-economic systems, including cities, this area is only beginning to develop, and the understanding of the use of data as a tool for the development of smart cities remains limited in the scientific literature. In general, it can be noted that research is increasingly focused on the use of digital data as a new socio-economic phenomenon, and attempts are made to conceptualize, classify and evaluate the role of different types of data in socioeconomic processes. In most cases, such studies are related to the use of big data in certain areas of the urban environment, such as transportation, public safety, and environmental protection. At the same time, the literature lacks studies of a general systemic nature on the use of big data for smart cities regardless of the field of application.

3.2 Data Types and Sources

The smart city concept implies integrating multiple information and communication technologies for city infrastructure management: transport, education, health, systems of housing and utilities, safety, etc. Municipal governments collect numerous heterogeneous information, and an "urban data" term can mean various datasets: data from video surveillance cameras, traffic, air quality, energy and water consumption, and images for smart recognition. Therefore for this study, the essential is to recognize and classify different datasets utilized in considered resources.

Trilles et al. describe a methodology of (big) data process produced by sensors in real-time [15]. It assumes that it works with different sensor data sources with different formats and connection interfaces. Wireless sensor networks (WSN) are used for monitoring the physical state of the environment: air pollution, forest fire, landslide, and water quality. Although the system proposed by the authors is designed to process all data types, the WSNs mainly produce numerical data like water level, and the gas concentration in the air, mainly classified as quantitative information. An efficient method to derive spatio-temporal analysis of the data, using correlations was proposed by [3]. The authors use data from Bluetooth sensors installed in light poles. The data was collected from the road sensors in the city of Aarhus in Denmark. The measurements are taken every 5 min and the dataset includes a timestamp, location information, average speed, and a total of automobiles at the time of commit. The data were classified as numerical as there are no text, images, sound, or video information.

Bordogna et al. used in their paper big mobile social data, which included users-generated, geo-referenced and timestamped contents [4]. The content means text data that users posts in modern emerging social systems like Twitter, Facebook, Instagram, and so forth. Hereby, the dataset can be classified as heterogeneous by way of containing the text of social network posts and numerical data of location and time. Wang et al. considered another approach to analysis and evaluated the effectiveness of deep neural networks [16]. The aim of their paper was the monitoring and control of local HIV epidemics. The collection includes statistics on the number of morbidities, mortality, and mortality by region, age, sex, and occupation. The type of data is categorized as text and numerical.

The researchers from Spain, Pérez-Chacón et al., proposed a methodology to extract electric energy consumption patterns in big data time series [12]. The study used the big data time series of electricity consumption of several Pablo de Olavide University buildings, extracted using smart meters over six years. Karyotis et al. presented a novel data clustering framework for big sensory data produced by IoT applications [9]. The dataset was collected from an operational smart-city/building IoT infrastructure provided by the Federated Interoperable Semantic IoT/cloud Testbeds and Applications (FIESTA-IoT) testbed federation. The array is heterogeneous and represents measurements of different types: temperature, humidity, battery level, soil moisture, etc.

Azri et al. presented a technique of three-dimensional data analytics using a dendrogram clustering approach [2]. It is assumed that the algorithm can be applied to large heterogeneous datasets gathered from sensors, social media, and legacy data sources. Alshami et al. tested the performance of two partition algorithms K-Means and Fuzzy c-Mean for clustering big urban datasets [1]. Compared techniques can be applicable to huge heterogeneous datasets in various areas like medicine, business, biology, etc. In the paper, the authors utilized urban data from various data sources, such as the Internet of Things, LIDAR data, local weather stations, and mobile phone sensors.

Chang et al. developed a new iterative algorithm, called the K-sets+ algorithm for clustering data points in a semi-metric space, where the distance measure does not necessarily satisfy the triangular inequality [6]. The algorithm is designed for clustering data points in semi-metric space. To understand what semi-metric space is, it is necessary to briefly consider the concept of metrics in space. The metric is the mapping for some set $d : X \times X \to R$, for which the axioms of non-degeneracy and symmetry have to be satisfied but not necessarily the triangle inequality. If the distance between different points can be zero, the metric is semi-metric. The method was evaluated with two experiments: community detection of signed networks and clustering of real networks. The dataset included 216 servers in different locations, and the latency (measured by the round trip time) between any two servers of these 216 servers is recorded in real-time.

Chae et al. have compared the performance of the deep neural network (DNN), long-short-term memory (LSTM), and the auto-regressive integrated moving average (ARIMA) in predicting three infectious diseases [5]. The study uses four kinds of data to predict infectious diseases, including search query data, social media big data, temperature, and humidity. Data related to malaria, chickenpox, and scarlet fever, for 576 days, were considered. As a result, the data is partly numerical and partly text. The research of Chen et al. focuses on multi-source urban data analysis [7]. The points of interest are geographical, street view, road map, and real-estate data. The record comprises the road network of the city, longitude, latitude, name, and functionality of a structure in the urban environment, and imagery of locations. Obviously, the dataset is ranked as heterogeneous.

Simhachalam and Ganesan presented a multidimensional mining approach in a successive way by finding groups (clusters) of communities with the same multi-dynamic characteristics [13]. The data refers to the statistics of population, migration, tax capacity, dwellings, employment, and commuters.

The majority of the studies assume heterogeneous nature data. There are two research papers with only numerical data and one of the papers investigates image data processing. Text and numerical data are dominant and they are collected from multiple sources (Table 3).

Table 3. Data types and sources.

Paper	Data	Category
[15]	data from different sensors	heterogeneous
[3]	traffic data collected from the road sensors in the city: geographical location, time-stamp, average speed, and total of automobile	numerical
[4]	social networks posts, timestamp, geo-location	heterogeneous
[16]	10-year historical HIV incidence data: the number of morbidity, morbidity, mortality and mortality by region, age, sex, occupation	heterogeneous
[12]	electricity consumption for 6 years for several buildings	numerical
[9]	big sensory data, measurements of different types: temperature, humidity, battery level, soil moisture	heterogeneous
[2]	smart city data	heterogeneous
[1]	data from the Internet of Things, LIDAR data, local weather stations, mobile phones sensors	heterogeneous
[6]	locations and the latency (measured by the round trip time) between any two data points	heterogeneous
[5]	search query data, social media big data, temperature, and humidity	heterogeneous
[7]	geographical data, points of interests data(longitude, latitude, name, and functionality of a structure in the urban environment), street view data, real estate data, mobile phone location data, social network data, micro-blog data, taxi GPS trajectory data, taxi profile data	heterogeneous
[13]	the measurements of the blood tests as the corpuscular volume of test substances and the number of half-pint equivalents of alcoholic beverages drunk per day	numerical

3.3 Algorithms

In general, 12 different approaches to big municipal data processing were considered. The methods can be divided into groups depending on the manner of information handling: clustering, classification, correlation, deep neural network, frameworks, and community detection. Figure 2 illustrates the proportion between different techniques. The most popular approach is clustering, various algorithms of clustering utilized in 5 considered studies.

The approach followed by [15] includes three layers: content layer, services layer, and application layer. The content layer includes sensor network data sources, and the services layer provides database connection, transformations of data, and communications protocols for real-time data handling and processing. The last layer implies client application. The service layer implements the Cumulative SUM (CUSUM) algorithm of anomaly detection. The method considers

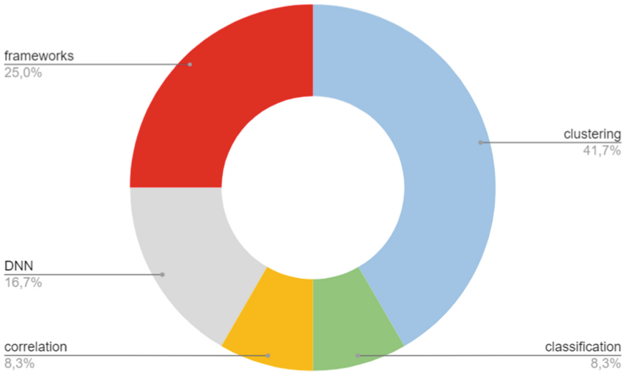

Fig. 2. Methods of urban data analysis

the set of observations following a normal distribution. For each collection of measurements, the cumulative sum is calculated. When the score overcomes the threshold, the algorithm detects anomalies. If the parameter exceeds the threshold, the anomaly will be due to the increase (up-event), and if the sum is greater than the threshold, it will be due to the decrease (down-event). Different data types from multiple sources are processed by a special wrapper and transformed into standard form. Transformed observation is encoded in line according to Open Geospatial Consortium (OGC) standard for Observations and Measurements.

The unique method used in [3] tried to apply correlation methods to urban data analysis. They suggested an efficient method to derive spatio-temporal analysis of the data, using correlations, with Pearson and Entropy-based methods and compares the results of both algorithms. Pearson's coefficient characterizes the presence of linear dependence between two values. The weakness of Pearson correlation is poor accuracy when variables are not distributed normally. Mutual information is the statistical function of two random variables, which describes the quantity of information of one random value in another. The constraint of mutual information is that it has a higher processing complexity than Pearson correlation. The technique continuously calculates the average correlation for sensory road data divided into two sectors until the data runs out. Different types of correlation were tested.

The Long short-term memory (LSTM) neural network models, autoregressive integrated moving average (ARIMA) models, generalized regression neural network (GRNN) models, and exponential smoothing (ES) models to estimate HIV incidence in Guangxi, China, and explore which model is the best and most precise for local HIV incidence prediction were used by [16]. ARIMA is the model used for time series forecasting. LSTM is a recurrent neural network, characterized by the ability to learn long-term dependencies. In this study, several models were built. The model with the lowest mean square error (MSE) was considered the optimal model. GRNN is a feed-forward neural network, which estimates values for continuous dependent variables. The principal advantages of GRNN are fast learning and convergence to the optimal regression surface as the number of samples becomes very large. GRNN is particularly advantageous with sparse data in a real-time environment because the regression surface is instantly defined everywhere, even with just one sample. The method is usually used for functions' approximation, so it can provide very high accuracy, but for huge samples is computationally expensive. ES model is one of the simplest and most widespread practices of series alignment. The method can be presented as a filter that receives the original series members as the input, and the output forms the current values of the exponential average.

The patterns in data related to electricity consumption were searched in [12]. The methodology describes all stages of data processing: data collection, cleaning, transformation, index analysis, clustering, and results. The first stage aims to pre-process the data so that they can be clustered. The second phase consists of obtaining the optimal number of clusters for the dataset by analyzing and interpreting various cluster validation indices. Next, k-means is used for clustering and, finally, retrieves the centroids for each cluster. The processing is done in Apache Spark and the algorithms include big data clustering validity indices (BD-CVIs) and k-means.

The community detection algorithm Girvan-Newman GN) [8] algorithm was modified for big data clustering of IoT sensors by [9]. Their method organizes complex data in blocks, called communities or modules, according to certain roles and functions, organized in a multi-graph. The problem is to find in a given multi-graph a partition of vertices where the objective function is minimized. To achieve this, the graph edges are deleted iterative, depending on the value of the metric. The Edge-Betweenness Centrality (EBC) is the most common metric used, but the computation for this is time-consuming. The authors suggested a new measure approximating EBC, which capitalizes on hyperbolic network embedding and can be considered as the "hyperbolic" analog of EBC. This measure is denoted as Hyperbolic Edge Betweenness Centrality (HEBC), and it is computed by utilizing the hyperbolic node coordinates assigned to the embedded nodes. The novel metric enhances the performance without harming accuracy.

The other technique of data organizing and processing proposed by [2] implies 3D data analytics using a dendrogram (hierarchical) clustering approach. 3D data represents a structure of information that combines, simultaneously, the classification and clustering tasks. The organized data is mapped to a tree structure and retrieved by tree traversal algorithms. Dendrogram clustering is a method of merging objects into bunches. In the study, the bottom-up algorithm of clustering is utilized, which means that each item in a class is assigned to a single cluster. Then combine the clusters until all objects are merged together. An important parameter is a distance between objects in a class. The metric shows a quantitative assessment of the items' similarity ratio according to different criteria. The given research does not provide a selection of the specific parameter, although the choice of metric occurs in the second step of the method. The ability to retrieve information and the efficiency of the structure were measured. In general, the technique demonstrates a good characteristic of information extraction but not the most attractive performance parameters.

Other clustering algorithms, Fuzzy c-Mean (FCM) and K-Means were tested by [1]. The k-Means algorithm is one of the simplest methods but at the same time the most inaccurate. The main idea is that at each iteration, the center of mass is recalculated for each cluster obtained in the previous step, then results are partitioned into clusters again under new centers. The algorithm ends when the cluster is not changed in iteration. The fuzzy c-Mean method allows for obtaining "fuzzy" clustering of large sets of numerical data and makes it possible to correctly identify objects at the boundaries of clusters. However, the execution of this algorithm requires serious computational resources and the initial setting of the number of clusters. In addition, ambiguity may arise with objects remote from the centers of all clusters.

A new approach for clustering data points was designed by [6]. In essence, the method is an extension of the K-set clustering algorithm for semi-metric space. The problem with the K-sets approach is that the triangle distance is not non-negative. Thus the K-sets algorithm may not converge at all and there is no guarantee that the output of the K-sets algorithm is clustering. For solving this difficulty, the definition of triangle distance was adjusted, so that the non-negativity requirement could be lifted. The experimental results confirm the proficiency of the method for the geographic distance matrix and the latency matrix. The deep neural networks for the prediction of infectious diseases were used in [5].

A visual analysis framework for exploring and understanding heterogeneous urban data was presented by [7]. A visually assisted query model is introduced as a foundation for interactive exploration coupled with simple, yet powerful, structural abstractions and reasoning functionalities.

One more clustering method is used by [13]. Fuzzy c-Means (FCM), k-means (KM), and Gustafson-Kessel (GK) clustering algorithms are implemented. According to the paper, the most accurate and effective algorithm is k-means clustering, but the other methods have their own advantages and show higher correctness in certain cases.

3.4 Algorithm Assessment

The algorithms described above have characteristics of performance and scalability that should be understood. Table 4 gives a comprehensive description of the complexity and accuracy of the considered algorithms. For many, it was not easy to evaluate the complexity since the time depends on the characteristics of the machine. Therefore we provide only rough estimates, and all presented assessments are for worst-case values.

The K-sets+ algorithm yields the highest performance from all cluster algorithms. The time complexity is linear $O((Kn + m)I)$, where I is the number of iterations. The other method with linear time is Fuzzy c-mean with $O(nCI)$, where C - number of clusters, I - number of iterations. If we compare the exponent for these two approaches, the apparent fact is that the K-sets+ gives a little

Table 4. Algorithms assessment

Algorithm	Purpose	Complexity	Accuracy	Ref
CUSUM algorithm	anomaly detection	O(n)	—	[15]
Mutual information and Pearson correlation	find correlation between sensory data	$O(n^3)$	+	[3]
LSTM	predict diseases	$O(w)$	85%+	[16]
ARIMA	predict diseases	—	80%	[16]
GRNN	predict diseases	—	76%	[16]
ES	predict diseases	—	74%	[16]
K-means clustering	to extract electric energy consumption patterns	$O(n^2)$	78%	[12]
modification of Girvan-Newman algorithm	community detection	$O(n^2)$	65%–100%	[9]
Dendrogram clustering	produce hierarchical tree structure for data for data retrieval and analytics	$O(n^3)$	—	[2]
K-Means	clustering	$O(n^2)$	87,94%	[1, 12]
Fuzzy c-Mean	clustering	$O(nCI)$	81,91%	[1]
K-sets+	clustering in metric space	$O(Kn+m)$	95%	[6]
DNN	predicting infectious diseases	$O(wnk)$	77%	[5]
VAUD	spatio-temporal data visualisation	—	78,6%+	[7]
K-Means	clustering	70,22%	60,41%–87,94%	[13]
Fuzzy c-Means (FCM)	clustering	68,54%	56,25%–81,91%	[13]
Gustafson-Kessel (GK)	clustering	60,68%	66,19%–95,83%	[13]
Similarity-Matrix-based Clustering	trip clustering	$O(n^3)$	—	[4]

advantage. Considering the accuracy, K-sets+ has 95% as the worst result. The Fuzzy c-Means algorithm gives the complexity on average 81,97%. It is noteworthy that the Girwan-Newman modification provides 100% accuracy for most datasets and only 50% in the case of outliers. It could be used for a dataset with low sparseness if high accuracy is required. The dendrogram clustering method is slower than the others but can produce a hierarchical tree structure for data. K-means clustering is the simplest method but has a quadratic complexity and an accuracy of not more than 88% for different input data.

The deep learning algorithms were compared by the set of parameters: MSE, Root-Mean-Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). From the results given by the authors, it follows that the most accurate algorithm is LSTM, but at the same time, the slowest. The fastest method is ES, but with the worst accuracy. All deep learning algorithms were used for predicting diseases. The accuracy of the ES and GRNN model was relatively poor [16]. The ARIMA model has several requirements: the time series should be stationary with steadily changing differences, and only linear relationships could be captured [16]. The DNN and LSTM models were observed to be sensitive to decreasing trends and increasing trends, respectively [5]. It is worth noticing that the time complexity for the deep neural network is hard to evaluate with O notation. The authors provide real-time results, according to the considered research the fastest model is ES, and the slowest is LSTM.

The CUSUM algorithm is time linear complexity. The solution is straightforward and fast but has limitations that must be taken into account, such as the consideration that all the series must follow a normal distribution and a series of observations cannot have trends [15]. VAUD presents the visualization of heterogeneous urban data. The approach is based on queries to the database, hence the time complexity can not be estimated. The data gathered from mobile phones and stored in one database combines different queries and different results are obtained. The accuracy on average for queries is 76%.

One of the widespread statistical methods applied to big data is correlation. In the listed papers, there is one algorithm that considered the correlation applied to smart city data. The study compared two types of methods: Pearson correlation and Mutual information. The time complexity for both is a cube. But Pearson correlation can discover the linear distribution of data, and mutual information can discover dependencies in more general data distribution cases. However, if an application prioritizes real-time response over accuracy, Pearson correlation will be suitable as it will only give a few false negatives. In other scenarios with different types of data streams (temperature, pollution, etc.), it is better to use mutual information without a priori knowledge of the potential correlations because we do not know the percentage of cases where Pearson correlation will fail to detect the correlations [3].

The assessment of time and accuracy of all proposed algorithms demonstrate that if our purpose is prediction, the best variant for us is deep neural networks like LSTM. For effective clustering, the K-sets+ or Fuzzy c-Means algorithms are the most powerful. If it is necessary to obtain additional analysis, it is possible to find the correlation. Considering the context of municipal data the frameworks are beneficial, as they assume all stages of data processing from storage to visualizing.

3.5 Processing Outcomes and Purpose

The anomaly detection by CUSUM algorithm of the [15], creates the warning message for the client side in the case of rare events. Each event contains a sensor identifier (sender field) and the identifier of the particular observation that has caused the event (identifier field). An event dashboard visualizes this data. The panel shows all sensing nodes of a network on a map using markers. Inside each marker, the amount of events that have been detected for this particular sensing node appears. If this node triggers an event, the marker turns red, if not the marker remains blue.

The analyses based on the correlation and mutual information were used to monitor the traffic of the city. Three sets of experiments have been performed. In the first one, the performance of Pearson correlation and mutual information was compared [3]. The results were visualized on Google Maps. It can be concluded that the Pearson correlation is effective for the linear distribution of data, and mutual information is vital for nonlinear dependencies but requires more time.

The results obtained by [16] are predictions of HIV disease for two years. Each compared algorithm has its metrics. For example, ARIMA includes a moving average process, an auto-regressive moving average process, an auto-regressive moving average process, and an ARIMA process according to the different parts of the regression and whether the original data are stable. To evaluate data accuracy, they compared with original information about HIV cases for 2015 and 2016 years. The same type of outcomes data demonstrates the [5]. They compared the same parameters for LSTM, DNN, and ARIMA to evaluate infectious disease prediction correctness. All cluster algorithms give the same result as a count of clusters and their accuracy.

The electricity consumption data were clustered into 4 and then into 8 groups in [12]. The outcomes are presented as diagrams. The clusters are categorized depending on buildings, seasons of the year, and days of the week.

The modification of the Girvan-Newman method with a novel metric provided by [9], was applied to multidimensional data obtained from an operational smart-city/building IoT infrastructure. The authors presented an accuracy evaluation, modularity, and time comparison of HGN and GN, comparing the execution time of GN and HGN algorithms for graphs with known communities and modularity comparison for 5, 10, 20, 30, and 60-minute sampling. Given that statistics demonstrate the computational efficiency and that algorithm can give accurate outcomes.

The cluster visualization into dendrograms, as tested on the information about 1000000 buildings was presented by [2]. Response time analysis was provided as well, which exhibits that response time for the proposed method is 50–60% faster than non-constellated data.

4 Conclusion

The aim of the article was to figure out what is the trend in the city's infrastructure data processing. The authors were interested in consumption data of electricity, water, heat, data of city traffic, and the methods for creating predicting models, clustering, and classifying. Increasingly, big data are seen as a key resource for the development of the urban environment, which presents opportunities for the optimization of economic processes, the creation of innovations in the social sphere, formation of new management models. The literature review serves as a foundation for future work in resource expenditures data analysis and urban management system creation.

The article presents a detailed analysis of the twelve papers from the last five years. The authors considered the techniques of urban data processing. The input and output data, assessments of algorithms' effectiveness, and methods description are provided. The inspection gives the following results: 1 algorithm of correlation, 2 algorithms of classification, 1 method of anomaly detection, 2 approaches for data visualization, 5 algorithms of predicting, and 6 methods for data clustering. A disproportion between the number of reviewed articles and the number of techniques dues to the fact that more than one method in each research was provided.

The input and output data vary depending on the method and purposes of the research. Predominantly heterogeneous data sources are considered. In one case the images are exploited and in two cases the numerical information is leveraged. The heterogeneous data mean that the information of different bases is used: images, text, and numerical. The video or sound data is not used in reviewed papers. The major part of the investigation offers a clustering model in the capacity of output results. The second place in prevalence is frameworks. The remaining outcomes can be divided between deep neural networks, classification, and correlation models.

The most interesting approach is the leverage of LSTM. Based on surveyed articles LSTM gives the highest accuracy of prediction and is the fastest solution in comparison with similar solutions. The forecasting of social phenomena based on city data is the most desirable result. Although, in the context of the modern situation is a still challenging task, as the whole pipeline of the assembly, processing, and analysis is important. As the given review demonstrates, different data are necessary for various problems, different algorithms give diverse findings. The dilemma of the practical benefits and standardization in smart city data is still open.

Acknowledgments. This work has been supported by FCT - Fundação para a Ciência e Tecnologia within the Project Scope: DSAIPA/AI/0088/2020.

References

1. Alshami, A., Guo, W., Pogrebna, G.: Fuzzy partition technique for clustering Big Urban dataset, pp. 212–216 (2016). https://doi.org/10.1109/SAI.2016.7555984
2. Azri, S., Ujang, U., Abdul Rahman, A.: Dendrogram clustering for 3D data analytics in smart city. **42**(4/W9), 247–253 (2018). https://doi.org/10.5194/isprs-archives-XLII-4-W9-247-2018
3. Bermudez-Edo, M., Barnaghi, P., Moessner, K.: Analysing real world data streams with spatio-temporal correlations: entropy vs. Pearson correlation. Autom. Constr. **88**, 87–100 (2018). https://doi.org/10.1016/j.autcon.2017.12.036
4. Bordogna, G., Cuzzocrea, A., Frigerio, L., Psaila, G.: An effective and efficient similarity-matrix-based algorithm for clustering big mobile social data, pp. 514–521 (2017). https://doi.org/10.1109/ICMLA.2016.188
5. Chae, S., Kwon, S., Lee, D.: Predicting infectious disease using deep learning and big data. Int. J. Environ. Res. Public Health **15**(8), 1596 (2018). https://doi.org/10.3390/ijerph15081596. http://www.mdpi.com/1660-4601/15/8/1596
6. Chang, C.S., Chang, C.T., Lee, D.S., Liou, L.H.: K-sets+: a linear-Time clustering algorithm for data points with a sparse similarity measure, pp. 1–8 (2018). https://doi.org/10.1109/UIC-ATC.2017.8397636
7. Chen, W., Huang, Z., Wu, F., Zhu, M., Guan, H., Maciejewski, R.: VAUD: a visual analysis approach for exploring spatio-temporal urban data. IEEE Trans. Visual. Comput. Graph. **24**(9), 2636–2648 (2018). https://doi.org/10.1109/TVCG.2017.2758362
8. Girvan, M., Newman, M.E.J.: Community structure in social and biological networks. Proc. Natl. Acad. Sci. **99**(12), 7821–7826 (2002). https://doi.org/10.1073/pnas.122653799

9. Karyotis, V., Tsitseklis, K., Sotiropoulos, K., Papavassiliou, S.: Big data clustering via community detection and hyperbolic network embedding in IoT applications. Sensors (Switzerland) **18**(4) (2018). https://doi.org/10.3390/s18041205
10. Kwon, O., Kim, Y., Lee, N., Jung, Y.: When collective knowledge meets crowd knowledge in a smart city: a prediction method combining open data keyword analysis and case-based reasoning. J. Healthcare Eng. **2018** (2018). https://doi.org/10.1155/2018/7391793
11. Materla, T., Cudney, E.A., Antony, J.: The application of Kano model in the healthcare industry: a systematic literature review. Total Qual. Manag. Bus. Excellence 1–22 (2017). https://doi.org/10.1080/14783363.2017.1328980
12. Pérez-Chacón, R., Luna-Romera, J., Troncoso, A., Martínez-Alvarez, F., Riquelme, J.: Big data analytics for discovering electricity consumption patterns in smart cities. Energies **11**(3) (2018). https://doi.org/10.3390/en11030683
13. Simhachalam, B., Ganesan, G.: Performance comparison of fuzzy and non-fuzzy classification methods. Egypt. Inform. J. **17** (2015). https://doi.org/10.1016/j.eij.2015.10.004
14. Subhash, S., Cudney, E.A.: Gamified learning in higher education: a systematic review of the literature. Comput. Hum. Behav. **87**, 192–206 (2018). https://doi.org/10.1016/j.chb.2018.05.028
15. Trilles, S., Belmonte, O., Schade, S., Huerta, J.: A domain-independent methodology to analyze IoT data streams in real-time. A proof of concept implementation for anomaly detection from environmental data. Int. J. Digit. Earth **10**(1), 103–120 (2017). https://doi.org/10.1080/17538947.2016.1209583
16. Wang, G., et al.: Application of a long short-term memory neural network: a burgeoning method of deep learning in forecasting HIV incidence in Guangxi, China. Epidemiol. Infect. **147**, e194 (2019). https://doi.org/10.1017/S095026881900075X