Studies in Big Data 134

Gilberto Rivera
Alejandro Rosete
Bernabé Dorronsoro
Nelson Rangel-Valdez *Editors*

Innovations in Machine and Deep Learning

Case Studies and Applications

Springer

Studies in Big Data

Volume 134

The series “Studies in Big Data” (SBD) publishes new developments and advances in the various areas of Big Data- quickly and with a high quality. The intent is to cover the theory, research, development, and applications of Big Data, as embedded in the fields of engineering, computer science, physics, economics and life sciences. The books of the series refer to the analysis and understanding of large, complex, and/or distributed data sets generated from recent digital sources coming from sensors or other physical instruments as well as simulations, crowd sourcing, social networks or other internet transactions, such as emails or video click streams and other. The series contains monographs, lecture notes and edited volumes in Big Data spanning the areas of computational intelligence including neural networks, evolutionary computation, soft computing, fuzzy systems, as well as artificial intelligence, data mining, modern statistics and Operations research, as well as self-organizing systems. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution, which enable both wide and rapid dissemination of research output.

The books of this series are reviewed in a single blind peer review process.

Indexed by SCOPUS, EI Compendex, SCIMAGO and zbMATH.

All books published in the series are submitted for consideration in Web of Science.

Gilberto Rivera · Alejandro Rosete ·
Bernabé Dorronsoro · Nelson Rangel-Valdez
Editors

Innovations in Machine and Deep Learning

Case Studies and Applications

Springer

Editors
Gilberto Rivera
División Multidisciplinaria de Ciudad Universitaria
Universidad Autónoma de Ciudad Juárez
Chihuahua, Mexico

Bernabé Dorronsoro
School of Engineering
University of Cadiz
Cádiz, Spain

Alejandro Rosete
Universidad Tecnológica de La Habana "José Antonio Echeverría"
La Habana, Cuba

Nelson Rangel-Valdez
Instituto Tecnológico de Ciudad Madero
Tecnológico Nacional de México
Tamaulipas, Mexico

ISSN 2197-6503 ISSN 2197-6511 (electronic)
Studies in Big Data
ISBN 978-3-031-40687-4 ISBN 978-3-031-40688-1 (eBook)
https://doi.org/10.1007/978-3-031-40688-1

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

Preface

Machine Learning (ML) is a branch of Artificial Intelligence (AI) continuously changing and adapting to address emerging challenges. Since the early conceptions of computers, humanity has struggled with developing more capable *thinking* machines. It began with the proposal of the Turing test, which suggests that a computer could be considered "intelligent" if it convinced a human that it is a person. The first models mimicked human thought processes but quickly developed into algorithms that acquire knowledge or skills through experience. Nowadays, this behavior is improved through Deep Learning (DL) techniques, a branch of ML that distinguishes itself by overcoming the finite capacity to absorb knowledge by integrating more experiences into its learning process.

The creation of inference engines scales up innovation in software capabilities. Mainly, ML & DL contribute to the software transition from human-driven to inference-driven. Furthermore, AI R&D plays an essential role in any enterprise innovation department, whether it is to aid decision-makers in making informed decisions as part of a Business Intelligent tool or as part of the developed technology that will be commercialized.

Innovations in Machine Learning and Deep Learning: Case Studies and Applications aims to collect the latest technological applications in the field of ML & DL to innovate tasks related to decision-making, forecasting, information retrieval, interpretable AI, risk management, healthcare, human activity recognition, sustainability, logistics among other topics related to this field. This book consists of twenty-two chapters, organized into three main areas:

First Part: *Analytics-Oriented Applications.* This seven-chapter part advocate to the systematic analysis of data. The studied cases' domains include time-series forecasting, feature selection, pattern classification, reusability frameworks, speech recognition, text classification, and question-answering systems. The original research in this part involves analyzing a rich set of ML & DL techniques such as residual-feedback artificial neural networks, tabu search, holographic neural networks, *k*-nearest neighbors, convolutional neural networks, and text analytics models. Some concrete applications include emotion recognition through speech,

sensationalism detection in news headlines, and Arabic language processing. Lastly, let's point out that this part shows that almost any data is subject to analysis and that ML & DL tools can aid during the process.

Second Part: *Healthcare-Oriented Applications.* These eight chapters highlight tools based on ML & DL techniques oriented to the support of medical care of individuals. This part analyzes applications for attention deficit hyperactivity disorder, classification of mosquitoes on human skin, pneumonia classification from X-ray images, human activity recognition, cholesterol prediction, cancer detection, characterization of the reverberation time of neonatal incubators, and COVID-19 prediction. The issues presented in the previous applications were tackled using strategies such as convolutional neural networks, dissipative particle dynamics simulation, and several classifiers (e.g., random forest and stochastic gradient). Throughout these chapters, the relevance of ML & DL to healthcare is noted in how such techniques can support the prevention of a wide variety of illnesses or conditions or to get a better understanding of them.

Third Part: *Sustainability-Oriented Applications.* This part encompasses seven chapters oriented to developing means to support the balance between society, environment, and economic growth. Humanity must coexist with the environment to be sustainable, i.e., its actions must not compromise its future, and they must support the perdurance of the ecosystems. ML & DL techniques make these tasks more manageable, and this part provides rich, supportive evidence on this topic. The analyzed case studies are CO_2 emissions prediction in freight transportation, planting system detection, ginger disease detection, coconut tree detection, logistics, and anomaly detection in low-cost sensors. In addition, a wide range of artificial neural networks lies within the approaches used to solve the previous applications. The chapters of this part offer a clear point of view of the tremendous impact that ML & DL techniques have in the green industry.

All chapters, rigorously analyzed by the book editors, resulted from a stringent double-blind peer-review process by field experts. The contributions of all the authors enrich the reader's experience and knowledge of ML & DL techniques and applications. The main focus is on the role that AI tools play in the solution of real-life issues. Hence, this book is expected to motivate readers to implement these technologies to become a Smart Business or Industry 4.0 environment. *Innovations in Machine Learning and Deep Learning: Case Studies and Applications* represents a channel to examine our knowledge about how ML & DL influence the solution of daily emerging needs. Lastly, we hope readers find inspiration in this book (or any of its chapters) that motivates research in developing intelligent solutions for real-world problems using ML & DL with related disciplines.

Chihuahua, Mexico — Gilberto Rivera
La Habana, Cuba — Alejandro Rosete
Cádiz, Spain — Bernabé Dorronsoro
Tamaulipas, Mexico — Nelson Rangel-Valdez

Contents

Predirol: Predicting Cholesterol Saturation Levels Using Big Data, Logistic Regression, and Dissipative Particle Dynamics Simulation

Reyna Nohemy Soriano-Machorro, José Luis Sánchez-Cervantes, Lisbeth Rodríguez-Mazahua, and Luis Rolando Guarneros-Nolasco

Abstract Four out of ten Mexican adults have high cholesterol, according to the National Institute of Cardiology. Cholesterol is essential for the production of substances in our body, such as hormones and vitamin D metabolism; it is essential for the absorption of calcium and bile acids. However, excess cholesterol causes hardening and narrowing in the walls of the arteries and can form a clot that causes a heart attack or stroke. Taking into account this problem, in this chapter, we present PREDIROL: Predicting Cholesterol Saturation Levels Using Big Data, Logistic Regression, and Dissipative Particle Dynamics Simulation, which presents an approach with Big Data and mesoscopic simulation techniques with a method of Particle Dynamics DPD (Dissipative Particle Dynamics). Parallel computing using CUDA was implemented to build the DPD model that would represent the cholesterol and blood molecules. However, considering the quantity of cholesterol and blood molecules generated in 3D, which required high computing power, we opted for the 3Dmol.js library based on WebGL for rendering 3D graphics within any web browser. PREDIROL seeks to raise awareness about the care of cholesterol concentration levels since having high levels is detrimental to health, but having low concentration levels, the body does not produce cells in the body. This is a tool for preventive medicine and to improve the lifestyle of users before they develop more serious ailments and even heart attacks or strokes.

R. N. Soriano-Machorro · L. Rodríguez-Mazahua
Tecnológico Nacional de México/Instituto Tecnológico de Orizaba, Orizaba, Veracruz, México
e-mail: M20011509@orizaba.tecnm.mx

L. Rodríguez-Mazahua
e-mail: lisbeth.rm2@orizaba.tecnm.mx

J. L. Sánchez-Cervantes (✉)
CONACYT-Instituto Tecnológico de Orizaba, Orizaba, Veracruz, México
e-mail: jlsanchez@conacyt.mx

L. R. Guarneros-Nolasco
Universidad Tecnológica del Centro de Veracruz, Veracruz, México
e-mail: luis.guarneros@utcv.edu.mx

G. Rivera et al. (eds.), *Innovations in Machine and Deep Learning*,
Studies in Big Data 134, https://doi.org/10.1007/978-3-031-40688-1_12

Keywords Big Data · Mesoscopic simulation · Cholesterol · Logistic regression

1 Introduction

Cholesterol is a serious and fat-like substance, this is a steroid that constitutes the essential component of the cell membrane, and it is a precursor of steroid hormones, which is found in all cells of the human body. Cholesterol is important for good health and is needed to make cell walls, tissues, hormones, vitamin D, and bile acid. This lipid comes from the consumption of foods of animal origin, such as egg yolks, meat, and dairy products made from whole milk. There are two types of cholesterol in the human body, HDL (High-Density Lipoprotein composed of lipids and proteins), commonly known as good cholesterol, and LDL (Low-Density Lipoprotein), known as bad cholesterol, when there is too much cholesterol in the blood with the latter, it can build up in the walls of blood vessels, blocking blood flow to tissues and organs, thereby increasing the risk of heart disease and stroke, and according to the National Institute of Cardiology, four out of ten Mexican adults have high cholesterol [1]. HDL has a diameter between 25 and 10 nm, and phospholipids are its main lipid. They are produced by the liver (30%) and the intestine (70%), and their main function is to extract excess cholesterol from the cells and transport it to the liver for elimination in the form of bile acids and cholesterol in the feces, while LDL are cholesterol-rich particles with a diameter of 20–25 nm that are taken up by the cells of the body and thus provide themselves with the cholesterol they require [2]. In relation to this, cardiovascular diseases encompass a wide range of disorders, including diseases of the heart muscle and the vascular system that supplies the heart, brain, and other vital organs where cholesterol levels take a prominent role. It is the leading cause of death worldwide [3]. Mesoscopic simulation is an alternative, clean, and cheap method to study ways in which there will be no exposure for experiments; it allows to surpass the limitations of molecular simulation because it is directed to the behavior of atoms [4]. Dissipative particle dynamics is a method for mesoscopic simulation that was developed to study the hydrodynamic behavior of complex fluids; it is also useful for the calculation of dynamic properties [5].

High cholesterol can be inherited, although it is often the result of unhealthy lifestyle choices so that it can be prevented and treated. Eating a healthy diet, exercising regularly, and sometimes taking medication help lower high cholesterol.

Big Data offers alternatives to carry out the analysis of large volumes of data, for example: (1) High-Performance Computing (HPC) to run advanced applications quickly, efficiently, and reliably; (2) Use of super-computing that uses coprocessors and accelerators; (3) HPC resources in the cloud that are increasingly accessible, being a service for consumers and, (4) Data analysis is performed by a specific analyst or a group of specialists [6, 7]. In the medical domain, Big Data helps specialists to maintain a communicative, intelligent, and well-oriented relationship to meet specific objectives, eliminating irrelevant information for the treatment of diseases. On the other hand, the mesoscopic simulation, which is located between the atomistic and

macroscopic scale, is the Dissipative Particle Dynamics (DPD), which is capable of capturing processes that are carried out in microseconds, which allows analyzing the streams of molecules and their motions based on predefined capacities and velocity-density functions. Individual molecules can be tracked to identify agglomerations or clusters of them that cause some effect depending on the application domain.

For this reason, the chapter presents PREDIROL, a system for the prediction of cholesterol concentration in the blood, to make the population aware of the care and prevention of cardiovascular diseases with a Big Data approach and mesoscopic simulation techniques with a DPD Particle Dynamics method. Among the expected benefits are: to show the cholesterol saturation levels in the blood by computational simulation and to support people with the prevention of heart attacks or cerebrovascular accidents.

This chapter is organized into four main sections: Sect. 1 covers the Related works that include a comparative analysis of PREDIROL and similar initiatives; Sect. 2 presents the architecture of PREDIROL; Sect. 3 describes the results obtained through a case study; and Sect. 4 presents the general conclusion and future work with some limitations about it.

2 Related Works

This section briefly describes the analysis of a set of literary initiatives related to PREDIROL. The initiatives analyzed are divided or classified into two groups: (1) Model for simulation of fluids and (2) Application of Data Mining for the prevention of cardiovascular diseases.

2.1 Models for the Simulation of Fluids

In this subsection, a summary and analysis of works related to the simulation of fluids is described. In [8], to search for new methods to eliminate excess cholesterol molecules, the DM molecular dynamics simulation method was used, with which the system composed of b-cyclodextrin molecules and cholesterol groups was studied, both independent as in an aquatic environment. The DM simulations were performed with the NAMD software (Not (just) Another Molecular Dynamics program), the integrator used was the standard (Brünger-Brooks-Karplus algorithm) and visualized with the VMD molecular modeling program to set up the simulation. For the simulations, the TIP3P CHARMM model (Chemistry at Harvard Macromolecular Mechanics, sets of force fields widely used for Harvard molecular dynamics) was used, adjusted due to the applied water model does not correctly reproduce the translational diffusion of water. The authors carried out a series of b-cyclodextrin-cholesterol DM simulations in the presence and absence of water. The influence of water is quite significant because the cholesterol molecules are grouped together

with water. The results of the simulations for the removal of excess cholesterol in the body were left to future considerations on the potential use of b-cyclodextrin in the context of combating atherosclerosis disease.

The high-density lipoprotein HDL was studied in [9] due to it being a powerful risk factor for cardiovascular disease. The authors analyzed 47 subjects with different levels of HDL (low and high) and studied their analyses with the R language for statistical computing. To perform the simulations, they used the GROMACS package (v. 4.0). The standard MARTINI lipid force field (model for bio-molecular simulations) was used for the composition of lipid and apolipoprotein components. The result obtained was that the molecular profile of HDL particles was combined with dynamic structural modeling, which revealed marked differences in HDL lipidomic profiles, as well as related clinical and biochemical characteristics between HDL-C subjects. low and high. The approach allowed it possible to demonstrate that changes in lipid composition also induce specific spatial distributions of lipids within HDL particles. In [10], an efficient simulation algorithm based on dissipative particle dynamics (DPD) was proposed; this method was used to study electrohydrodynamic phenomena in electrolytic fluids. The fluid flow is mimicked with DPD particles, while the evolution of the ionic species concentration is described using Brownian pseudo particles. The algorithm was developed from the set of electrokinetic equations for electrolyte fluids, which the authors propose is designed to optimize the computational efficiency of electrolyte simulations at high physiological salt concentrations in fluids. The authors noted that the approach of their algorithm, called Condiff-DPD, is not restricted to electrolyte solutions since the same idea is applied to other mesoscale fluid flow simulations where diffusion of a minor component is important, for example, in microreactors. Terrón-Mejía et al. [11] studied the interfacial and structural properties of fluids confined by different walls at the mesoscopic scale using DPD simulations on the grand canonical ensemble. The entire methodology was implemented in simulation code called SIMES, which is designed to study complex systems at a mesoscopic scale using a graphics processing unit (GPU), for confined fluids DPD was used since in the DPD model, the particles do not represent molecules but rather groups of molecules with smooth boundaries which were hybridized with the Grand Canonical Monte Carlo (MC) methodology to simulate fluids confined by walls, at the mesoscopic scale, and the interfacial properties for confined fluids For the simulation, four different models of confined fluids were analyzed, these confined between smooth walls, symmetrical rough walls, non-symmetrical rough walls, and a combination between a smooth wall and a symmetrical rough wall, in the author's conclusion were described the different behaviors of the confined fluids depending on the walls with which the simulations were carried out and it is defined that to determine the value of the interfacial tension, the geometry of the walls is not fundamental, but the intensity of solid–fluid wall interaction. In [12], a simulation package for red blood cell (RBC) flow with GPU-accelerated chemical transport property based on an adaptation of the transport dissipative particle dynamics (tDPD) method was presented; tDPD was used due to it allows capturing the reaction kinetics at the mesoscopic level. For programming, the languages C/C++ , CUDA, C/ C++ , and MPI (Message

Passing Interface) were used. The simulation package processes all computational workloads in parallel per GPU and incorporates multithreaded programming and non-blocking MPI communications to improve inter-node scalability. The authors mentioned that they used GPU for the processing speed compared to its counterpart, the Central Processing Unit (CPU). The red blood cell model presented by the authors correctly recovers cell membrane viscosity, elasticity, bending stiffness, and chemical transport across the membrane. The lack of fluid flow simulation models of complex RCB structures through complex capillary vessels led Phan-Thien et al. [13] to elaborate a method for the simulation of red blood cell flows in different tubes; the methods used were smoothed dissipative particle dynamics (SDPD) and dissipative particle dynamics (DPD), with pa- parameters that have a specific physical meaning, combined with thermal fluctuations in a mesoscale simulation and the immersed boundary method (IBM), a preferred method for handling fluid–structure interaction problems, the latter has been widely used to handle the fluid-RBC interaction in simulations. The authors' objective was to couple SDPD and IBM to perform RBC simulations on complex flow problems. The authors first developed the SDPD-IBM model and demonstrated with CFD (Computational Fluid Dynamics) simulations the ability of the SDPD-IBM method to simulate red blood cell flows in rectangular, cylindrical, curved, bifurcated, and constrained tubes. Similarly, an overview of DM simulations of nHDL lipidic nanodiscs was presented by Pourmousa and Pastor [14]. For the simulation, there was solvation of all nanodiscs with water molecules and 0,15 M NaCl in 15–16 nm in lateral length cubic boxes using the web interface CHARMM-GUI (Effective Simulation Input Generator and More).

The authors conclude after comparing nHDL simulations that understanding nHDL maturation is challenging for future studies because it not only requires a large number of computational resources but also depends on high-resolution structures of other proteins involved, the apolipoprotein APOA1 interacts with the ATP Binding Cassette A1 (ABCA1) transporter to acquire lipids first. Little is known about the details of this first stage, as single-molecule imaging revealed a dimeric form of the iteration of ABCA1 with APOA1. On the other hand, a GPU-accelerated package for the simulation of flow in nano-to-micropore networks with a mesoscale model of many-body dissipative particle dynamics (mDPD) was described in [15]. Mesoscopic simulations of the flow of Hydrocarbons in shales are a challenge due to the heterogeneous pores in shales with sizes ranging from a few nanometers to a few micrometers. The authors used the method of many-body dissipative particle dynamics (mDPD), which is a mesoscopic model for fluids and solid molecules; the wall is modeled with a non-slip boundary condition that prevents the particles of fluid from penetrating the walls; they also used the Velocity-Verlet algorithm, the programming languages used were CUDA C/C++ with MPI and OpenMP. The authors mentioned that a simulation in CPU takes 15 times more than using parallel programming with GPU; the use of GPU effectively reduces the overhead in communication between rank/nodes. Thanks to non-obstructive angioscopy with the NOGA device, Kojima et al. [16] aimed to study NO-GA-derived aortic ruptured plaque RP and the stereographic distribution and regional increase in wall shear stress (WSS)

using computational fluid dynamics modeling. (CFD), for which the PHOENICS-CFD Works application was used; this model was used to calculate the blood flow velocity distribution and reveal the three-dimensional distribution of WSS within the lumen of the aortic arch. The authors studied 45 patients who underwent 3D-CT three-dimensional computed tomography before coronary angiography. The WSS in the aortic arch was measured by CFD analysis based on the finite element method using uniform inflow and outflow conditions. Aortic PR was detected by NOGA. They concluded that aortic RP detected by NOGA (Transendocardial injection with an electromechanical mapping system) was strongly associated with a higher maximum WSS in the aortic arch derived by CFD using 3D-CT. The maximum value of WSS has an important role in the mechanism not only of aortic atherosclerosis but also of aortic PR.

2.2 Data Mining Application for Prevention of Cardiovascular Diseases

In this part of the section, we continue with the descriptions of initiatives similar to our work. The use of data mining techniques for the prevention of cardiovascular diseases is described briefly below.

Liu et al. [17] investigated the association between the cholesterol efflux capacity and the risk of mortality presented by a patient who has a cardiovascular problem (coronary artery disease). Studying 1737 patients with coronary artery disease, the authors found that cholesterol efflux capacity is not simply explained by circulating HDL cholesterol or apolipoprotein AI levels; it is also independently related to the presence and the extent of atherosclerosis; they conclude by suggesting that cholesterol efflux capacity is a predictor of mortality from all cardiovascular problems in both acute coronary syndrome (ACS) and diabetic ketoacidosis (DKA). Reference [18] performed a meta-analysis of data from patients with no history of cardiovascular disease; they used two methods of increasing complexity to model repeated measurements: cumulative mean values and individual-specific intercepts and slopes for each individual from linear regression models of mixed effects. The objective was to quantify the change in discrimination and risk stratification of individuals according to their expected 5-year CVD risk when information from repeated measures of risk predictors was added to the evaluation of single measures to the levels of risk factors used in the standard risk scores. They used 38 studies with emerging risk factors. The events identified for the risk of cardiovascular diseases were non-fatal myocardial infarction or any cerebrovascular accident, and different measures of HDL, and blood pressure, among others, were used in the models. The authors conclude that if a mean of repeated post-baseline measurement of systolic blood pressure is incorporated, it is possible that total cholesterol and HDL cholesterol in cardiovascular disease risk prediction models result in slight improvements in the discrimination and the reclassification of the risk. In the health industry, large amounts of data are

collected that contain information not observed at first glance, which is useful for decision-making for the early detection of diseases; that is why Singh et al. [19] used advanced data mining techniques to develop an effective cardiac disease prediction system using a neural network. The authors' system used fifteen medical parameters for prediction (age, sex, blood pressure, cholesterol, and obesity). The tool they used was Weka 3.6.11, with a data set of 303 records where 40% served as a training set and 60% as a test set. The authors demonstrated that their model has the best results and helps experts in the medical field to plan a better diagnosis in time for the patient since the system worked realistically and, in the results, it was shown that it predicts cardiac diseases with a 100% accuracy by using neural networks. Heart disease is a type of ailment that is directly related to the human heart and blood circulation; for this reason, the objective of [20] was to collect information on the six parameters that are age, chest pain, electrocardiogram, systolic blood pressure, fasting blood sugar, and serum cholesterol to detect cardiovascular disease. The model they proposed for the diagnosis of cardiovascular diseases (DCD) using the Mamdani fuzzy inference system (DCD-MFIS) showed an accuracy of 87.05%; they also proposed the model for the diagnosis of cardiovascular diseases includes fuzzy logic using Mamdani Fuzzy Inference System (DCD-MFIS), Artificial Neural Network (DCD-ANN) and Deep Extreme Machine Learning (DCD-DEML) approach, these achieved higher precision and accuracy. The DCD extreme deep machine learning they proposed achieves higher accuracy than the previously proposed solutions, which are 92.45%. To verify the performance, they calculated many indicators that determine precise performance. The training of the neural networks is done using the layers of 10−20 neurons that denote the hidden layer. DEML reveals and indicates a hidden layer containing ten neurons, which shows the best result. In [21], an efficient system for the prediction of cardiac diseases was presented; they used databases with the clinical information of cardiac patients and used data mining to process the data set obtained with input values such as the age, sex, chest pain, resting blood pressure, serum cholesterol, fasting blood sugar, among others. The authors used the Weka tool and KEEL (Knowledge Extraction Based on Evolutionary Learning) since it is a set of tools with open-source Java language (GPLv3) to implement development processes for data mining problems and the algorithm C4.5. When the system was tested, it obtained an 86.3% of prediction in the test phase, and it obtained a prediction of 87.3% in the training phase. This tool predicts heart disease early. The main cause of heart disease is high blood pressure and high cholesterol. In [22], the authors use the human eye to detect high levels of cholesterol in the blood because when thick plaques form in the arteries, an occlusion in the retinal vein blood flow is observed. The authors developed a mobile application that uses an algorithm, uses Machine Learning, and data such as height, weight, age, sex, eating habits, and iris photo capture, among others. The result that the application presents to the user, if he has cholesterol in his blood, is a "No" or "YES." The authors use Heroku as a server, and for image processing, they use Python modules. The challenges that the authors found are that the images can vary from person to person, the conditions of the Smartphone camera are different, and the lighting situation, due to the light reflection, affects the diagnosis.

2.3 Comparative Analysis

Tables 1 and 2 show a comparative analysis to observe differences and similarities between them, in addition to the methodologies used by the authors to solve the different problems.

Therefore, after carrying out an analysis of the articles addressed in this chapter, it was found that the works [11, 12, 15] use the GPU architecture, which has advantages such as time, due to it being 15 times faster than using a CPU architecture, that is why the GPU-accelerated simulation was chosen for this work, where it allows parallel processing between one or more GPUs and CPUs. In [10, 11], they use the DPD method, and two variant methods of DPD are also used in [12, 13] and [15] for the simulation of simple and complex fluids; these works allow the justification of the use of the DPD in this work in which different levels of cholesterol in the blood will be simulated because it is the method that best describes complex fluids since there is better control of the transport properties. In [8] and [9], cholesterol was simulated with different simulation methods and software, works on the simulation of red blood cells were found as in [12] and [13]; unlike previous works, it will simulate different levels of cholesterol in the blood.

For the medical field, it was also found that [16] simulates the flow of the aorta for angioscopy, using data from medical studies.

The works [18–20] and [21] used data mining techniques to detect and predict cardiovascular diseases, also in [17–21] used cholesterol as a predictor of risk for heart disease, while [22] proposes an application to detect cholesterol in the blood using machine learning and retinal images, among other data, unlike this work, cholesterol will only be predicted with data provided by the user. PREDIROL uses data mining techniques to predict cholesterol levels and the risk of heart disease.

3 PREDIROL Architecture

This section of the chapter shows the architecture of the PREDIROL application that meets the required needs. Figure 1 shows the architecture to be used for the development of the application.

The layers, modules, and flows that make up the system architecture are described below.

Layers of the PREDIROL Architecture

- Presentation: This layer presents the interface between the user and the system, where the user's information and the predictions of cholesterol levels are displayed.
- 3D Models: This layer contains the 3D model generator component of cholesterol levels.

Table 1 Comparison of work related to simulation

Author	Contribution/Solution	Models	Methodology	Architecture/ Software for simulation
Makieła et al. [8]	The system composed of b-cyclodextrin molecules and cholesterol groups was studied	Simulation of b-cyclodextrin molecules and cholesterol groups	TIP3P CHARMM	Software NAMD
Yetukuri et al. [9]	HDL simulation with studies of 47 subjects with high and low cholesterol	Simulation of Molecular Dynamics on a large scale	The MARTINI standard lipid force field	GROMACS simulation package (v. 4.0)
Medina et al. [10]	Simulation method specially designed for systems with high salt concentrations	DPD algorithm for electrolytes	DPD	Does not use
Terrón-Mejía et al. [11]	Method for the simulation of confined fluids with rough walls	Simulation of confined fluids	DPD and MC	GPU
Blumers et al. [12]	GPU-accelerated RBC (Red Blood Cell Count) simulation package	Simulation of transport with red blood cells	Transport dissipative particle dynamics (tDPD) method	GPU
Phan-Thien et al. [13]	Hybrid numerical method with the immersed boundary method and smoothed dissipative dynamic particle method	Numerical model	Smoothed Dissipative Particle Dynamics (SDPD)	CFD (Computational Fluid Dynamics, fluid dynamics)
Pourmousa and Pastor [14]	Comparison of models for the simulation of HDL nanodiscs	Molecular dynamics simulation	Molecular dynamics with the exchange of replicas and Monte Carlo	CHARMM-GUI
Xia et al. [15]	GPU-accelerated mesoscopic pore flow simulation package based on a many-body dissipative particle dynamics model (mDPD)	Mesoscopic model for solid and fluid molecules	Many-body dissipative particle dynamics model (mDPD)	GPU
Kojima et al. [16]	Study where the maximum value of WSS has an important role in the mechanism not only of aortic atherosclerosis but also of aortic RP	Computational fluid dynamics modeling	Fluid dynamics	PHOENICS-CFD Works application

Table 2 Comparative analysis of articles related to Data mining

Author	Approach/Issue	Contribution/Solution	Methods/tools used
Liu et al. [17]	They investigate the association between cholesterol output capacity and all-cause and cardiovascular mortality in patients with coronary artery disease	Cholesterol output capacity serves as an independent measure to predict cardiovascular mortality	Statistics
Paige et al. [18]	Lack of evaluation of repeated measurements of blood pressure and cholesterol to predict the risk of cardiovascular diseases	Statistical study	Linear Regression
Singh et al. [19]	There is a collection of data that contains useful information to predict the level of risk of heart disease	Algorithm development for an effective heart disease prediction system (EHDPS) using a neural network	Data mining techniques (Neural networks)
Siddiqui et al. [20]	They collect information on six parameters which are age, chest pain, electrocardiogram, systolic blood pressure, fasting blood sugar, and serum cholesterol that Mamdani Fuzzy Expert uses to detect cardiovascular diseases	Artificial neural networks and extreme deep machine learning approaches for the automated diagnosis of cardiovascular diseases	Data mining techniques (Neural networks)
Purushottam et al. [21]	They propose an efficient system for the prediction of cardiac diseases with a set of data provided by the clinical information of cardiac patients	Efficient system for predicting heart disease	Data mining techniques (Decision Tree C4.5) with the Weka and Keel tools
Alhasawi et al. [22]	They propose a mobile application that detects blood cholesterol with iris photography and other data about the user	Mobile application for the detection of cholesterol in the blood	Machine Learning and Heroku server with Python modules to process images

- Predictions: This layer presents a component that calculates the cholesterol prediction.
- Data model: This contains the user profile component.
- Repository: The repository is based on Mongo DB, which will contain all the user profile data.

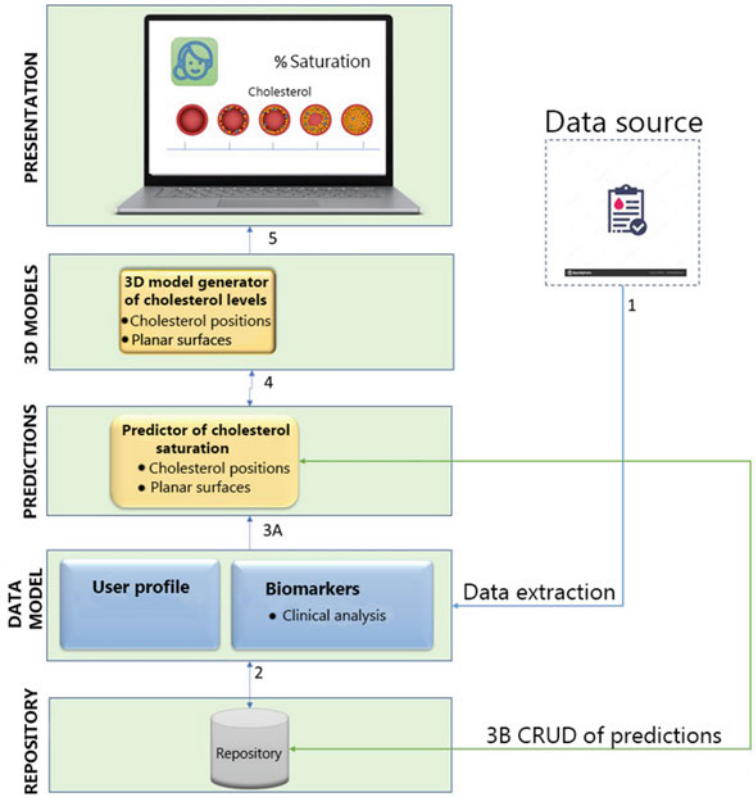


Fig. 1 PREDIROL system architecture

Architecture Components

- Data sources: This module is in charge of storing the analysis of the patients, which contains the levels of cholesterol in the blood.
- Repository: The repository is responsible for saving the patient's data in a database manager.
- Predictions: This module receives the data from the repository of the profile of each user to predict the cholesterol positions that the user currently has and that the user will have in time periods. The module will also predict the thickness of the walls at different periods of time.
- 3D models: In this module, three-dimensional models of cholesterol levels are generated in different periods of time with data from the cholesterol saturation predictor module.
- Presentation: This module shows the user the 3D models of cholesterol saturation in the different periods of time simulated in the 3D cholesterol levels model generator module, as well as the user profile.

Workflow

1. From the data sources (clinical analysis), the information will be extracted and filtered to store the data in the repository.
2. Based on the information stored in the repository, the predictions of cholesterol saturation in the blood will be made at different periods of time and will be stored in the repository.
3. The predictions of cholesterol levels (3A) will be passed to the component that generates the positions of the molecules for the 3D models of the different time periods; these models will be stored in the repository (3B).
4. From the positions of the molecules, the component in charge will generate the three-dimensional models of cholesterol saturation in the time periods calculated by the predictor.
5. The user will be able to visualize through the application the different levels of cholesterol saturation in different periods of time in 3D.

3.1 Big Data Model

Figure 2 shows the data model implemented in MongoDB for the storage of user data and for prediction.

In the data model, five documents are observed. Table 3 shows the model documents with their description.

Each of the documents that make up the data model provides data persistence for the different information used by PREDIROL for its proper functioning. Among the most important information are the records provided by the laboratory of "Sanatorio Escudero in Orizaba, Ver. Mexico" in order to validate PREDIROL. The specialists in the laboratory of the Sanatorio Escudero provided us with 10,355 patient records with three essential fields for the prediction of cholesterol saturation levels, which are "age," "gender," and "cholesterol." It is important to mention that for privacy reasons, the first and last names have been changed. In the following section, we describe how the Cholesterol saturation level prediction module works using the fields "age" and "cholesterol."

3.2 Cholesterol Saturation Level Prediction Module

Logistic regression was used in the prediction module using the variables age versus total cholesterol in mg/dl. Figure 3 shows a graph with the data for model training.

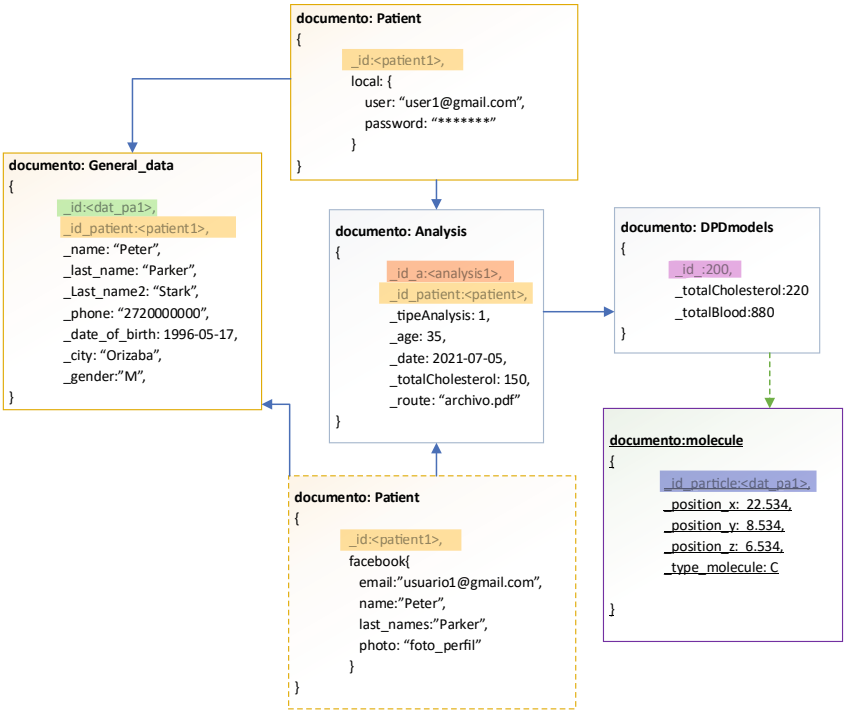


Fig. 2 Big Data model

Table 3 Description of the data model in Fig. 2

Document	Description
Patient	This document stores the patient's data to log in, either using a registered email or through the social network Facebook
General_ data	This stores the patient's data to complement the record, such as full name and age, among others
Analysis	The document stores the amount of cholesterol, as well as the date of the analysis, and the type of analysis that the patient underwent
DPDmodels	In this document, you have stored the data for the construction of the DPD models of the concentration of cholesterol in the blood
Molecule	This document stores the positions of the molecules and the type to which they belong

The model was created in R language. Listing 1 shows a fragment for the creation of the model, which was later trained. A partitioning of 70–30 (Line 3 of Listing 1) was considered for training. That is, 70% of the sample was used for training and 30% for predictions. Additionally, we chose the size 0.9 to obtain a better visualization of the results plot using the plotting package ggplot.

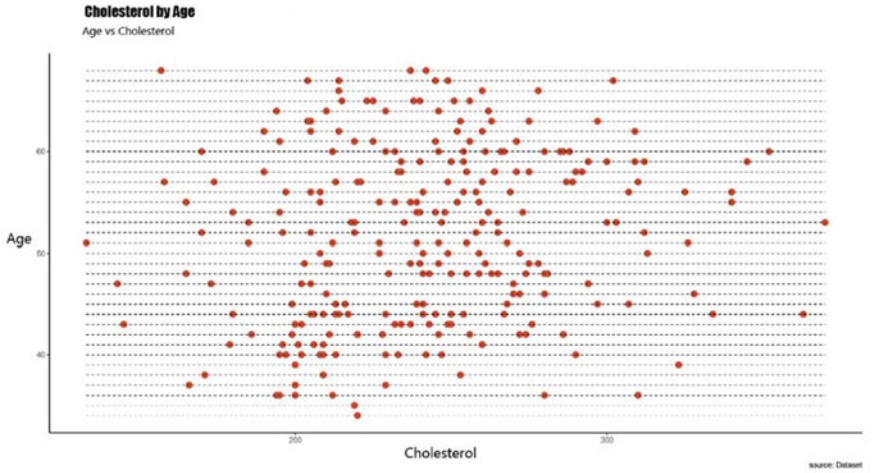


Fig. 3 Cholesterol graph by age

Listing 1: Model created in R language

```
#Loading data
data <- read.csv(file = 'datasetRL.csv', sep = ',')
#Separating the train and test
split = sample.split(data$Cholesterol, SplitRatio = 0.7)
nltrain = subset(data, split == TRUE)
nltest = subset(data, split == FALSE)
#Creating model
set.seed(1234)
LRmodel <- lm(Cholesterol ~ Age, data = nltrain)
summary(LRmodel)
#Predictions for the training set
y_predict <- predict(LRmodel, nltrain)
ggplot() + geom_point(data = nltrain, aes(x = Age, y =
    Cholesterol), size = 0.9) +  geom_line(aes( x =
    nltrain$Age, y = y_predict), color = "red") +
    xlab("Age") + ylab("Choleterol") + ggtitle("Training
    Set Adjustment Curve (nltrain)")
#Predicting cholesterol from a range of age groups
range.ages <- data.frame(Age = seq(20, 80))
predict_value <- predict(LRmodel, range.ages)
glimpse(predict_value)
df_predic=as.data.frame(predict_value)
write.csv(df_predic,'prediction20_80.csv')
```

The prediction module get the age range from 20 to 80 years of a user. (Line 15 of Listing 1). Based on these parameters, the module already trained generate the prediction of cholesterol saturation levels (Lines 14–19 of Listing 1), and it shows the prediction of the cholesterol levels throw a graph, as depicted in Fig. 4.

Finally, a spreadsheet with obtained results is generated as a backup. It is important to mention that this work focuses on the adult population and the prevention of diseases that can be caused by high cholesterol levels; for this reason, the experiments were carried out with age ranges between 20, 60, and 80 years old.

Figure 5 depicts the operating diagram of the prediction of cholesterol concentration in the blood within the system.

This diagram of activities (depicted in Fig. 5) includes five layers: (1) Users that must be registered and authenticated to use the system; (2) Web Application to interact with PREDIROL and visualize the results; (3) Prediction module, that is responsible for generating the predictions of saturation cholesterol levels according to the range of age of the user; (4) Simulation module, which calculates random positions within a cube representing two types of molecules cholesterol and blood and, (5) Repository, that store the information processed in PREDIROL. The following section describes the functionality of the cholesterol level simulation module with dissipative particle dynamics.

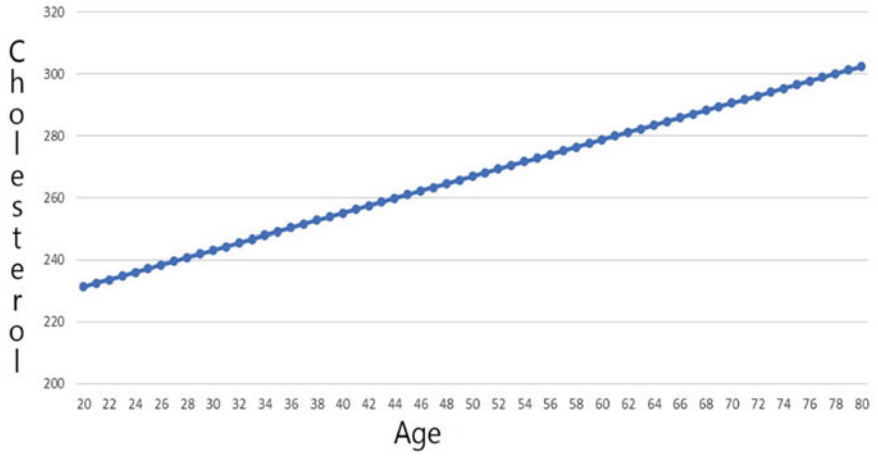


Fig. 4 Graph obtained from the prediction of the age range from 20 to 80 years

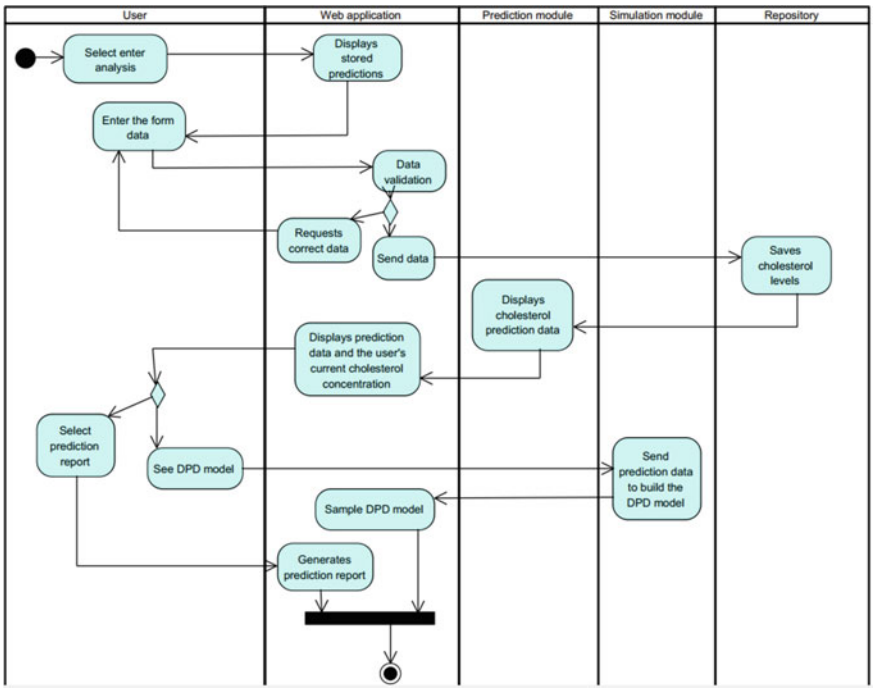


Fig. 5 Diagram of activities of the prediction of PREDIROL

3.3 Cholesterol Levels Simulation Module with Dissipative Particle Dynamics

The simulation module calculates random positions within a cube representing two types of molecules, the amount of blood represented is one deciliter, and each simulated molecule represents one milligram.

In this code list, the size of the box containing the two types of molecules to be simulated is calculated since the size depends on the number of total molecules to be represented. Later, the module calculates the positions of the two types of molecules at random in the form of coordinates (X-axis, Y-axis, Z-axis) inside the box. Figure 6 shows an example of cholesterol at 210 mg/dl, where red represents blood and yellow represents cholesterol.

To develop the cholesterol Levels Simulation module with dissipative particle dynamics, initially, parallel computing using CUDA was implemented to build the DPD model that would represent the cholesterol and blood molecules. However, considering the quantity of cholesterol and blood molecules generated in 3D, the positions (in X axis, the position in Y, and the position in Z), and their movements, which required high computing power, we opted for 3Dmol.js library based on WebGL for

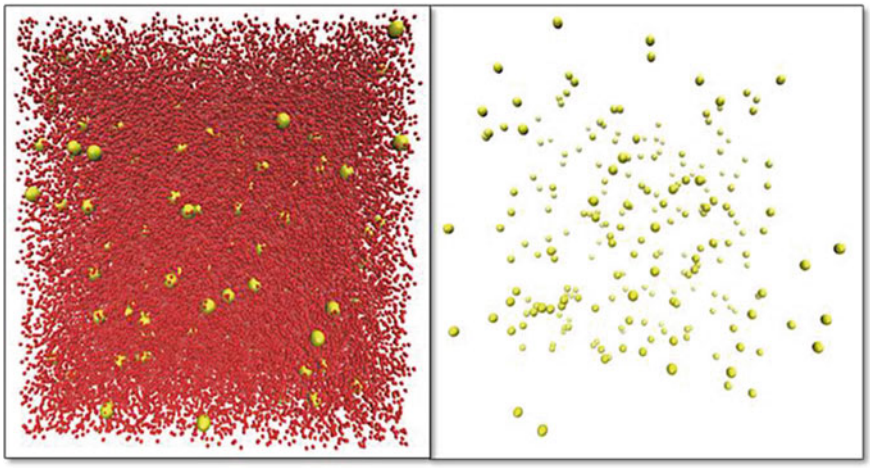


Fig. 6 Concentration of 210 mg/dl of cholesterol in the blood

rendering 3D graphics within any web browser, conserving the DPD model but using technology and more light.

The scale chosen for the representation in final mg/dl is from 1 to 1000; this means that a molecule will represent the molecules that contain 1000 mg. Figure 7 shows an example with the final scale, where 290 mg/dl of cholesterol in the blood is represented, where again red represents blood and yellow represents cholesterol.

For the construction of the DPD cholesterol concentration model, a method was developed to read the data from a document, where the positions of the molecules and their chosen cholesterol concentration are found. The file to be read is found in the

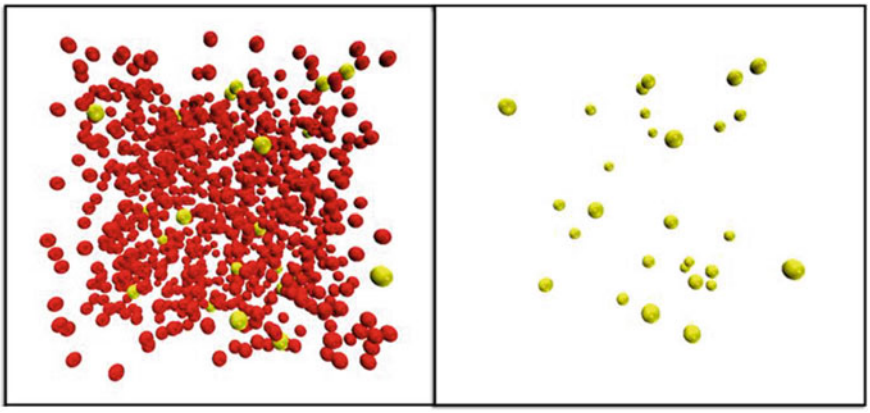


Fig. 7 Blood cholesterol concentration of 290 mg/dl

"moldata" variable; with the "glviewer.addModel" method, the file type is specified; for these simulations, the ".xyz" type is used.

For the construction of the DPD cholesterol concentration model, a method was developed to read the data from a document, where the positions of the molecules and their chosen cholesterol concentration are found. Finally, the properties are added to the model, including the type of representation of the molecule, followed by the size of the radius that this model will have. The molecules type "A" will be represented in red (blood), while type "B" molecules will be represented in yellow (cholesterol); if there is another type of molecule in the file, it will be represented in white. It is important to mention that the DPD model allows adding more molecules for future inclusions, such as triglycerides, oxygen particles, and blood clots, among others.

Section 4 presents a case study that includes the functionality of PREDIROL.

4 Case Study: Prediction of Cholesterol Levels of a Hospital Patients

This section aims to present the results of PREDIROL and the case study. This application will provide a prediction of approximate levels of cholesterol and its relationship with age, in which the increase in cholesterol can be observed if the person does not lead a healthy life; high cholesterol is indicative of diseases associated with the cardiovascular system.

For the development of this application, the data was obtained, and tests were carried out in a hospital. The data of clinical analysis of patients were collected from the month of February of the year 2022; the information collected was from 150 patients during a period of 3 months. The information provided by the Sanatorium Escudero were age, blood cholesterol, triglycerides, and diagnosis of the patients. The age range was from 18 to 88 years.[1]

Table 4 shows some data provided by the sanatorium, omitting data considered sensitive or private; these data were substituted synthetically (random names).

Of this sample, six patients performed the iteration with the system. Figure 8 shows the form where the analysis of Patient 6, Patient 7, Patient 8, Patient 9, Patient 12, and Patient 14 was recorded.

After of capture the data on the form for recording the results of the analysis, the patients chose the option "Get Prediction." Table 5 shows the results obtained from the application for the six patients.

Figures 9A-F depict the results of the predictions of cholesterol levels presented by PREDIROL to the users. Additionally, PREDIROL presents in semaphore format the cholesterol levels normal (green), alert (yellow), and critical (red).

As a complement to the results presented in Figs. 9 and 10 (A and B) shows the Mesoscopic Simulation of cholesterol in the blood through DPD models obtained

[1] Dataset: https://www.dropbox.com/sh/g1jenihy6xz1cy1/AACDRwBQ8J5thsPcJCyCTv_0a?dl=0

Table 4 Example of analysis data provided by the laboratory of sanatorium Escudero

Name	Age	Diagnosis	Cholesterol	Triglycerides
Patient 1	73	DM	201.5	221.1
Patient 2	84	PB. DM	244	174.6
Patient 3	40	DM/IRC	402	660.1
Patient 4	43	PB. DM	105.1	152.5
Patient 5	73	DM	201.5	221.1
Patient 6	24	S/DX	201.5	207.6
Patient 7	54	PB/ DM	344.6	238.1
Patient 8	63	S/DX	286.2	222
Patient 9	31	S/DX	208.2	184.1
Patient 10	31	S/DX	199.5	128.8
Patient 11	36	DMII	211.2	222.9
Patient 12	43	S/DX	269.3	273
Patient 13	22	ANEMIA/HIV	209.2	216.1
Patient 14	62	PROSTATA	235.9	94.1
Patient 15	43	DMII/GEPI	350.7	503.6
Patient 16	28	S/DX	209.2	216.1
Diagnoses				
DM = Diabetes Mellitus 2 IRC = Chronic renal failure S/DX = Undiagnosed PROSTATA = Benign prostatic hyperplasia GEPI = Probably infectious gastroenteritis HIV = Human immunodeficiency virus infection				

from patients 6 and 14, with their current total cholesterol. On the same screen, users can observe other cholesterol levels increasing or vice-versa. We decided to omit the results of the other patients because they are representative results, and presenting more screenshots is unnecessary.

PREDIROL allows the users to generate a report in.pdf format with all information about their current total cholesterol and their age in order to support the medical reports. In addition to the prediction, Fig. 11 (A-B) depicts the section of the prediction graph generated from reports of patients 6 and 14.

Finally, Fig. 12 shows an example of the prediction table according to the age obtained; this table is downloaded in the user report; if they are within the optimal values, it will be green, yellow indicates that a medical assessment is needed, and the red color indicates that cholesterol levels are very high, in the report generated by the user, the table is shown with the prediction in an age range from 20 to 80 years, giving a total of 60 records.

The patients who participated in this case study were of an age range from 20 to 35 years; in this group, it can be observed that their cholesterol levels are within

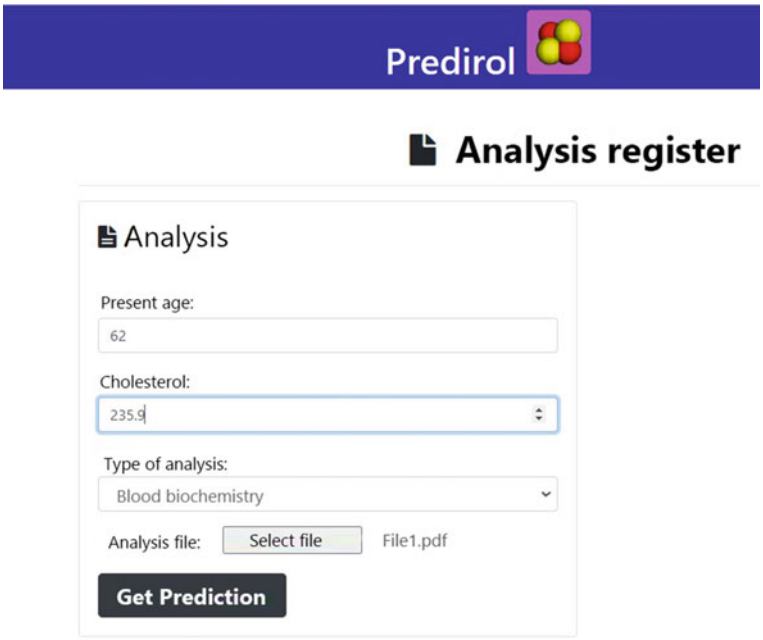


Fig. 8 User interface of the analysis record

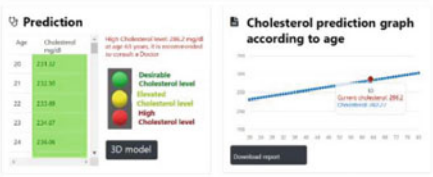
Table 5 Result of the prediction of patients chosen for the case study

Patient	Age	Cholesterol
6	24	201.5
7	54	344.6
8	63	286.2
9	31	208.2
12	43	269.3
14	24	201.5

normal parameters, while the patients from 40 to 65 years have high cholesterol and patients with higher cholesterol values have diagnoses such as "Diabetes mellitus," also associated with very high triglyceride levels, while in cases such as Patient 7 and Patient 12 in which their diagnoses were not reported, however, they reflect elevated cholesterol levels that predispose to developing heart attacks or strokes in the future. In the case of Patient 8 and Patient 12, their age and cholesterol levels coincide with the prediction, for which a diagnosis was not indicated either, but likewise, the warning of cholesterol levels are higher than the normal range.

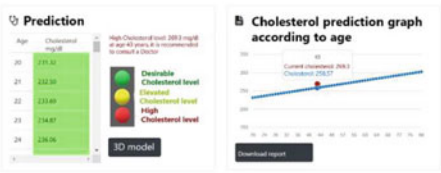
A: Results of Patient 6

B: Results of Patient 7



C: Results of Patient 8

D: Results of Patient 9

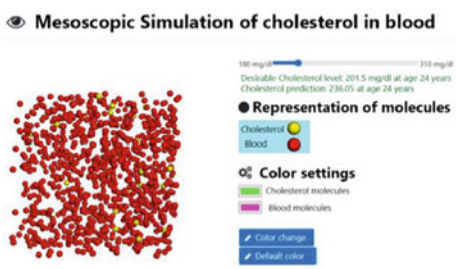


E: Results of Patient 12

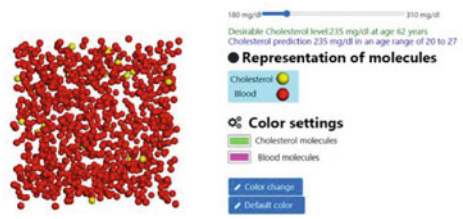
F: Results of Patient 14

Fig. 9 Results of the predictions of cholesterol levels presented from PREDIROL to the users

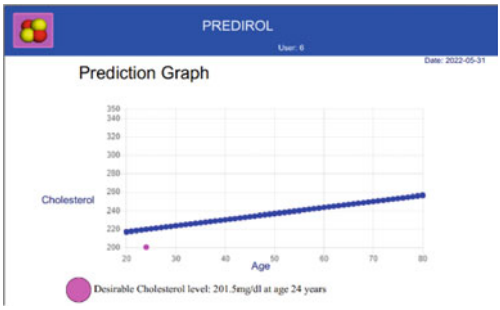
Fig. 10 Simulation of blood and cholesterol



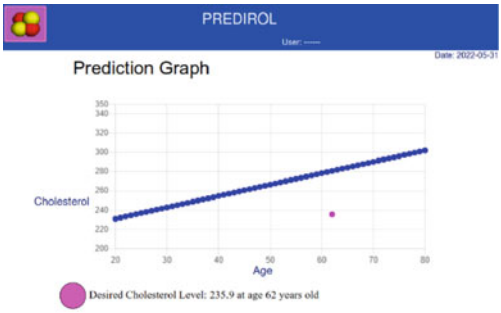
A: Patient 6



B: Patient 14



A: Patient 6



B: Patient 14

Fig. 11 Prediction graph of cholesterol levels

Age	Cholesterol
20	231.32
21	232.50
22	233.69
23	234.87
Age	**Cholesterol**
28	240.80
29	241.98
30	243.17
Age	**Cholesterol**
45	260.94
46	262.13
47	263.31

Fig. 12 Example of user report prediction table

5 Conclusions and Future Work

The application PREDIROL was developed as an initiative to predict possible diseases since an analysis of works related to mesoscopic simulation topics with DPD, Cholesterol, and disease prevention was carried out. For the prediction, logistic regression was used to predict cholesterol levels according to age. The DPD model of the concentration of cholesterol in the blood of the patients was built on a scale of 1 to 1000 mg/dl.; this means that a molecule represents all the molecules that contain 1000 mg in a deciliter of blood. PREDIROL seeks to raise awareness about the care of cholesterol levels since having high levels causes the accumulation of cholesterol and other deposits on the walls of the arteries (atherosclerosis). These deposits (plaques) can reduce blood flow through the arteries and, over time, can lead to heart attacks or strokes. The organism also needs cholesterol for the production of substances such as hormones and the metabolism of vitamin D; it is essential for the absorption of calcium and bile acids. In the data collected from the laboratory of the Sanatorio Escudero, it was observed that patients who had a total cholesterol concentration equal to or less than 80 mg/dl were diagnosed with anemia, being susceptible to other diseases. This application makes patients aware of the care of their total cholesterol levels; this is an indication that the patient is in good health or that he has some condition.

Although PREDIROL offers an alternative for the care of people's cardiovascular health, it has some limitations, such as not correlating cholesterol with pathologies that increase the probability of suffering heart disease, and stroke, among other problems. In this sense, future work considers including triglyceride concentration levels and correlating them with cardiovascular and cerebrovascular diseases, including atherosclerosis, familial combined hyperlipidemia, as well as dysbetalipoproteinemia, hypercholesterolemia, and hypertriglyceridemia, to mention just a few.

Acknowledgements This research chapter was sponsored by Mexico's National Council of Science and Technology (CONACYT) and Mexico's Secretariat of Public Education (SEP) through the PRODEP program. The authors also thank Tecnológico Nacional de México (TecNM) and the members of the REDTCAP network for supporting this work.

References

1. Zárate, A., Manuel-Apolinar, L., Basurto, L., la Chesnaye, E., Saldívar, I.: Colesterol y aterosclerosis. Consideraciones Históricas Y Trat., Arch Cardiol Mex **86**(2), 163–169 (2016). https://doi.org/10.1016/j.acmx.2015.12.002
2. Carvajal, C.: Lipoproteínas: metabolismo y lipoproteínas aterogénicas. Med. Leg. Costa Rica **31**(2), 88–94 (2014)
3. Gaziano, T., Reddy, K.S., Paccaud, F., Horton, S., Chaturvedi, V.: Cardiovascular disease, Disease control priorities in developing countries, 2nd edn, (2006). https://doi.org/10.1056/NEJMra035098

4. Mayoral, E., De La Cruz, E., Longoria, L.C., Nahmad-Achar, E.: Estudio del proceso de adsorción-desorción de contaminantes en medios confinados mediante simulaciones computacionales. Rev. Int. Contam. Ambient. **27**, 253–261 (2011)
5. Satoh, A.: Introduction to practice of molecular simulation: molecular dynamics, Monte Carlo, Brownian dynamics, Lattice Boltzmann and dissipative particle dynamics. Elsevier (2010). https://doi.org/10.1016/C2010-0-67006-4
6. Bolívar, A., García, V., Florencia, R., Alejo, R., Rivera, G., Sánchez-Solís, J. P.: A preliminary study of smote on imbalanced big datasets when dealing with sparse and dense high dimensionality. In: Pattern Recognition: 14th Mexican conference, MCPR 2022, Ciudad Juárez, Mexico, June 22–25, 2022, Proceedings, pp. 46–55. Springer International Publishing, Cham (2022). https://doi.org/10.1007/978-3-031-07750-0_5
7. Pedrycz, W., Martínez, L., Espin-Andrade, R. A., Rivera, G., Gómez, J.M. (eds.).: Preface. In: Computational intelligence for business analytics. Springer, v–vi (2021). https://doi.org/10.1007/978-3-030-73819-8
8. Makieła, D., Janus-Zygmunt, I., Górny, K., Gburski, Z.: Investigation of the influence of β-cyclodextrin on cholesterol lodgement — A molecular dynamics simulation study. J Mol Liq **262**, 451–459 (2018). https://doi.org/10.1016/j.molliq.2018.04.098
9. Yetukuri, L., Söderlund, S., Koivuniemi, A., Seppänen-Laakso, T., Niemelä, P.S., Hyvönen, M., Taskinen, M., Vattulainen, I., Jauhianen, M., Orešič, M.: Composition and lipid spatial distribution of HDL particles in subjects with low and high HDL-cholesterol. J. Lipid. Res. **51**(8), 2341–2351 (2010). https://doi.org/10.1194/jlr.M006494
10. Medina, S., Zhou, J., Wang, Z.-G., Schmid, F.: An efficient dissipative particle dynamics-based algorithm for simulating electrolyte solutions. J. Chem. Phys **142**(2), 24103 (2015). https://doi.org/10.1063/1.4905102
11. Terrón-Mejía, K.A., López-Rendón, R., Goicochea, A.G.: Mesoscopic modeling of structural and thermodynamic properties of fluids confined by rough surfaces. Phys. Chem. Chem. Phys. **17**(39), 26403–26416 (2015). https://doi.org/10.1039/C5CP03823E
12. Blumers, A.L., Tang, Y.-H., Li, Z., Li, X., Karniadakis, G.E.: GPU-accelerated red blood cells simulations with transport dissipative particle dynamics. Comput. Phys. Commun **217**, 171–179 (2017). https://doi.org/10.1016/j.cpc.2017.03.016
13. Ye, T., Phan-Thien, N., Lim, C. T., Peng, L., Shi, H.: Hybrid smoothed dissipative particle dynamics and immersed boundary method for simulation of red blood cells in flows. Phys. Rev. E. **95**(6), 63314(2017). https://doi.org/10.1103/PhysRevE.95.063314
14. Pourmousa, M., Pastor, R.W.: Molecular dynamics simulations of lipid nanodiscs. Biochim. Biophys. Acta (BBA)-Biomembr., **1860**(10), 2094–2107 (2018). https://doi.org/10.1016/j.bbamem.2018.04.015
15. Xia, Y., Blumers, A., Li, Z., Luo, L., Tang, Y.H., Kane, J., Goral, J., Huang, H., Deo, M., Andrew, M.: A GPU-accelerated package for simulation of flow in nanoporous source rocks with many-body dissipative particle dynamics. Comput Phys Commun **247**, 106874 (2020). https://doi.org/10.1016/j.cpc.2019.106874
16. Kojima, K., Hiro, T., Koyama, Y., Ohgaku, A., Fujito, H., Ebuchi, Y., Arai, R., Monden, M., Migita, S., Morikawa, T., Tamaki, T., Murata, N., Akutsu, N., Nishida, T., Kitano, D., Sudo, M., Fukamachi, D., Yoda, S., Takayama, M., Hirayama, A., Okumura, Y.: High wall shear stress is related to atherosclerotic plaque rupture in the aortic arch of patients with cardiovascular disease: a study with computational fluid dynamics model and non-obstructive general angioscopy. J Atheroscler Thromb, 56598 (2020). https://doi.org/10.5551/jat.56598
17. Paige, E., Barrett, J., Pennells, L., Sweeting, M., Willeit, P., Di Angelantonio, E., Wood, A.: Cholesterol efflux capacity is an independent predictor of all-cause and cardiovascular mortality in patients with coronary artery disease: A prospective cohort study. Atheroscler., **249**, 116–124 (2016). https://doi.org/10.1016/j.atherosclerosis.2015.10.111
18. Paige, E., Barrett, J., Pennells, L., Sweeting, M., Willeit, P., Di Angelantonio, E., Wood, A.: Use of repeated blood pressure and cholesterol measurements to improve cardiovascular disease risk prediction: an individual-participant-data meta-analysis. Am. J. Epidemiol **186**(8), 899–907 (2017). https://doi.org/10.1093/aje/kwx149

19. Singh, P., Singh, S., Pandi-Jain, G.S.: Effective heart disease prediction system using data mining techniques. Int. J. Nanomedicine. **13**(T-NANO 2014 Abstracts), 121–124 (2018). https://doi.org/10.2147/IJN.S124998
20. Siddiqui, S.Y., Athar, A., Khan, M.A., Abbas, S., Saeed, Y., Khan, M.F., Hussain, M.: Modelling, simulation and optimization of diagnosis cardiovascular disease using computational intelligence approaches. J Med Imaging Health Inform **10**(5), 1005–1022 (2020). https://doi.org/10.1166/jmihi.2020.2996
21. Purushottam, Saxena, K., Sharma, R.: Efficient heart disease prediction system. Procedia Comput. Sci, **85**, 962–969 (2016). https://doi.org/10.1016/j.procs.2016.05.288
22. Alhasawi, Y., Mullachery, B., Chatterjee, S.: Design of a mobile-app for non-invasively detecting high blood cholesterol using eye images. In: Proceedings of the 51st Hawaii international conference on system sciences, pp. 3227–3235. ScholarSpace (2018)