

# Domingo Martínez, PhD, SNI-I

Desarrollo de Propuestas,  
COA Bioinformática y análisis de datos

---

7 de Octubre de 2024  
Ciudad Universitaria CDMX.



LIIGH-UNAM  
INTERNATIONAL LABORATORY FOR  
HUMAN GENOME RESEARCH

# **Acerca de mí**

---

## **Bioinformática y análisis de datos de salud.**

- LIIGH, 2022 – a la fecha.
  - CINVESTAV, genómica avanzada, 2022-2023.
  - ENESJ - INB, Neurociencias, 2020-2022.
- 



# Contenido

---

1. Servidores
2. Aprendizaje automático
3. Apps web interactiva



# Contenido

---

1. Servidores → **Clúster de alto rendimiento**
  2. Aprendizaje automático → **Red bayesiana**
  3. Apps web interactiva → **Shiny app**
-

---

# Propuesta 1

Estrategias y Metodología para la *Instalación, Mantenimiento y Administración de un Clúster* (HPC), dirigido al **computo de alto rendimiento** en **bioinformática, análisis de datos y biología de sistemas.**

---

# HPC

---



**High-Performance Computing**, sistema que usa muchos **recursos computacionales** (procesadores, memoria, almacenamiento), para resolver **problemas complejos** y realizar **cálculos intensivos** rápidos y eficientes (Veerla et al., 2023).

---

# HPC

---

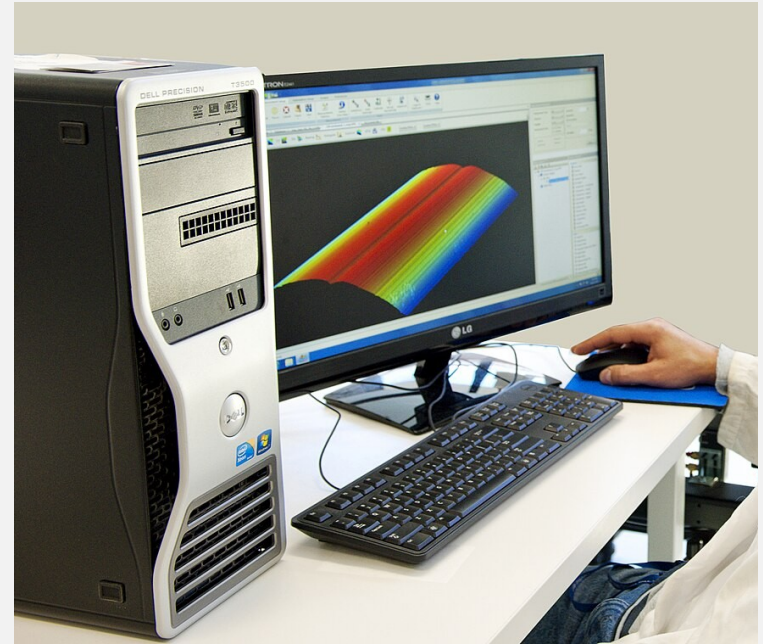
Clúster de **computadoras interconectadas** (cliente/nodo) que **trabajan en paralelo** (Parth et al., 2014) y son *administradas* por un **servidor central** (servidor/nodo principal) (Alsmadi et al., 2016).

---

# Encuentre las diferencias...



VS



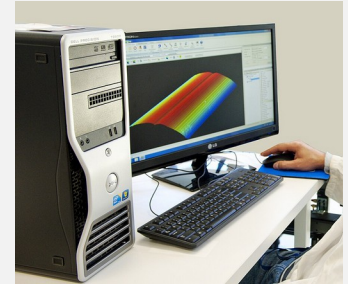
Fuente: BIFI Institute,  
[https://commons.wikimedia.org/wiki/File:BIFI\\_HPC\\_INFRASTRUCTURE.jpg](https://commons.wikimedia.org/wiki/File:BIFI_HPC_INFRASTRUCTURE.jpg)

Fuente: Flickr user ipasadelade (artist: Jennie Groom),  
[https://commons.wikimedia.org/wiki/File:Scientist\\_working\\_with\\_personal\\_computer\\_2013\\_\(Dell\\_Precision\\_T3500\\_Workstation\).jpeg](https://commons.wikimedia.org/wiki/File:Scientist_working_with_personal_computer_2013_(Dell_Precision_T3500_Workstation).jpeg)



Hoy, el manejo de *big-data*, análisis bioinformático y modelado de sistemas biológicos, se realizan **colaborativamente** entre diferentes grupos de investigación (Beaman et al., 2012).

**¿Por qué no crear un HPC con sus computadoras?**



# Objetivo

---

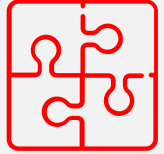


**Implementar un HPC**, mediante la **interconexión** de *PCs con las que actualmente cuenta la RAI*, **personas, grupos o laboratorios** que participan en investigaciones multidisciplinarias relacionadas con **bioinformática, análisis de datos y biología de sistemas**.

---

# Estrategias

---



**Necesidades**

**Inventario de recursos**

**Diseño de la infraestructura**

**Manuales de operación y mantenimiento**

---

## Montaje

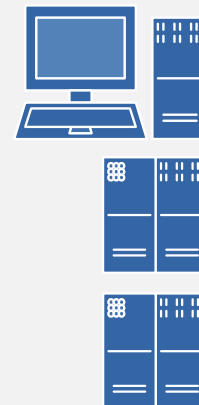
Arquitectura  
Cronograma  
Elección nodo principal  
Ubuntu en server  
LDAP en server  
Ubuntu desktop en nodos  
Integrar LDAP con nodos para autenticación  
Slurm en todos los nodos  
NFS clúster  
NFS en nodos  
Software en server  
Sistema de módulos  
Pruebas  
Registro de usuarios  
Lanzamiento

## Mantenimiento

Ganglia para monitoreo  
Cronograma de actualizaciones  
Ajustar parámetros para mejorar rendimiento  
Documentar la guía de instalación y configuración.  
Crear y publicar la wiki del clúster en github.

## Gestión

LDAP para gestionar cuentas de usuarios.  
Políticas de uso de recursos mediante SLURM.  
Desarrollo de scripts y herramientas de automatización de tareas.  
Reportes periódicos sobre uso y rendimiento.  
Revisiones periódicas de la infraestructura.  
Planeación de expansiones.



---

## Propuesta 2

Pipeline de **Aprendizaje Automático Bayesiano**  
para *analizar y predecir* el riesgo de **hipertensión**  
en **personas mexicanas con Lupus**  
Heritematoso Sistémico, con datos del  
Registro Mexicano de Lupus.

---

# Lupus heritematoso sistémico (LES)

---

Enfermedad autoinmune que puede causar **daño a cualquier órgano**, se caracteriza por un **conjunto** ampliamente **heterogéneo** de **manifestaciones clínicas** (Goulielmos et al., 2018).

---



Fuente: Doktorinternet,  
[https://commons.wikimedia.org/  
wiki/File:Lupusfoto.jpg](https://commons.wikimedia.org/wiki/File:Lupusfoto.jpg)

# Datos epidemiológicos **LES**

---

**Prevalencia mundial:** 3 a 7713.5 / 10000 personas (Goulielmos et al., 2018).

**Incidencia:** 1.5 a 11 / 10000 personas (Goulielmos et al., 2018).

**Prevalencia en México:** se desconoce,  
pero varía geográficamente (Barber et al., 2023).

**Proporción:** 90 % mujeres (Hernández-Ledesma et al., 2023).

**Hipertensión:** induce daño renal y afección CV

(Bruce et al., 2003; Mok et al., 2013; Nikpour et al., 2011)



# Síndrome metabólico (SM)

---

- **Factores de riesgo** de **enfermedades CV**.
- Aumentan **riesgo de diabetes** tipo 2.
- Enfermedades CV: **primera causa de muerte**.



***La hipertensión es un factor importante en el SM y es común en personas con LES.***



# FACTORES

## Modificables

- **Hipertensión**
- Colesterol
- Triglicéridos
- IMC
- Tabaquismo
- Actividad física
- Alcoholismo
- Dieta

## No modificables

- Edad
- Sexo
- Historia familiar



# La hipertensión es modificable

Es primordial identificar **alertas tempranas** para **predecirla, prevenirla y tratarla**, especialmente en **población mexicana de lupus**, la cual figura entre las poblaciones con *más daño acumulado* y con **más mortalidad**.

# Pregunta de investigación

---



¿Cuáles son las **variables críticas** para **predecir** la **hipertensión** en **personas mexicanas con Lupus?**



# Relevancia

---

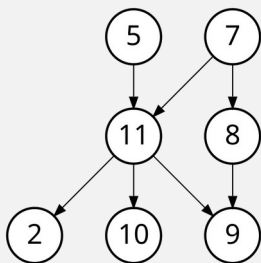


- Diagnóstico difícil y tardío
  - Datos ruidosos y censurados
  - **Modelo Redes Bayesiana**
  - App para alertas tempranas
  - Mejor calidad de vida
  - Reducción de decesos
  - Reducción gastos hospitalarios
-

# Red Bayesiana

---

Modelo gráfico que **codifica** la **relación probabilística** entre variables de interés.



## Ventajas:

- Nas y censurados
  - Semántica causal y probabilística
  - Conocimiento experto
  - Inferencias
-

# Objetivo

---



**Implementar un HPC**, mediante la **interconexión** de *PCs con las que actualmente cuenta la RAI*, **personas, grupos o laboratorios** que participan en investigaciones multidisciplinarias relacionadas con **bioinformática, análisis de datos y biología de sistemas**.

---

# Implementación

## Pre-procesamiento

Selección de variables  
Curado del conjunto de datos  
Imputación o remoción de valores perdidos  
SMOTE para balanceo entre casos y controles

## Estimación

Estimación de la estructura inicial.  
Método Delphi.  
Reducción de dimensiones.  
Partición.  
Tablas de distribución posteriores.

## Predicción

Validación cruzada.  
Predicción del riesgo de padecer hipertensión.  
Evaluación del modelo.  
Inferencias de interés.

**Modelo  
Redes  
Bayesianas**

---

## Propuesta 3

Desarrollo de un **sistema web interactivo** (Shiny App) para **consulta, visualización y predicción de lupus** en México..

---



# Necesidades grupos sociales

## Pacientes

- Características
- Comorbilidades
- Calidad de vida
- Afrontamiento
- Manifestaciones
- Tratamiento

## Médicos

- Prevalencia
- Incidencia
- F. riesgo
- F. protección
- Alertas
- Daños
- Consulta App

## Investigadores

- Conjunto de datos
- Modelos
- Exploración

## Público

- Demográficos
- Gráficos
- Consulta desde celular

# Objetivo

---

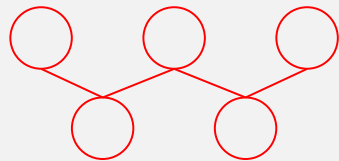


Desarrollar una **plataforma web** de acceso abierto,  
basada en una **Shiny app**, para  
**consultar, visualizar, y predecir** de forma  
**interactiva**, distintas cuestiones del lupus en México.

---

# Programación reactiva

---

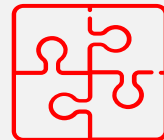


Útil **para desarrollar apps** o sistemas web interactivos. **Manipula** la *propagación de flujos de datos* y el **manejo de eventos asíncronos**.

Algunos **cambios en el programa** (e.g., selección de variables) **se actualizan** en otras partes del script **sin** tener que **reescribir nuestro código**.

# Metodología

---



Desarrollar una **Shinny app** (Chang et al., 2024) en el lenguaje de **programación R** (R Core Team, 2024), embebida en un **dashboard** de un script en formato **Quarto R Markdown** (Teague et al., 2024), para luego ser **montada en un servidor estable** y seguro en el Laboratorio de Visualización Científica Avanzada (LAVIS) de la UNAM.

---

## Demográficos



## ¿Tienes lupus?



Averigua cómo  
afrontan la  
enfermedad otras  
personas de tu edad

## Epidemiología Factores de riesgo



## Elige un modelo de análisis.



# Pipeline

## Pre-procesamiento

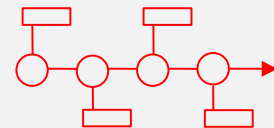
- Integración del conjunto de datos que cubra las necesidades de cada grupo.
- Diseño de la App.

## Programación secuencial

- Desarrollo de gráficos y modelos estáticos.
- Mapas.
- Gráficos de barras.
- Densidad.

## Programación reactiva

- Agregar reactividad a los gráficos.
- Agregar reactividad a los modelos.
- Usabilidad.
- Pilotaje.
- Lanzamiento



---

# Gracias por su atención

Domingo Martinez, PhD.

CONAHCYT Researcher at the

International Laboratory for Human Genome Research

Professor at Escuela Nacional de Estudios Superiores,

National Autonomous University of Mexico, Juriquilla Mex.

Member of the National System of Researchers SNI-I

Boulevard Juriquilla 3001 Col. Jurica la Mesa, C.P. 76230

+52 (442) 1 92 62 31 to 46

+52 (442) 374 22 20

Ldmv82@gmail.com

domingo.martinez@comunidad.unam.mx

---

# Referencias



© 2024 Canonical Ltd. (2024a). *Ubuntu 24.04.1 LTS* (Versión 24.04.1 LTS) [Software]. Ubuntu and Canonical are registered trademarks of Canonical Ltd. <https://ubuntu.com/download/desktop>

© 2024 Canonical Ltd. (2024b). *Ubuntu server OpenLDAP. Install and configure LDAP*. <https://ubuntu.com/server/docs/install-and-configure-ldap>

© 2024 Canonical Ltd. (2024c). *Ubuntu server OpenSSH Server*. <https://ubuntu.com/server/docs/openssh-server>

© The kernel development community. (2022). *Network File System (Version 6.11.0-rc8)* [Software]. The Linux Kernel. <https://docs.kernel.org/admin-guide/nfs/index.html>

Alsmadi, I., Khamiseh, S., & Xu, D. (2016). Network Parallelization in HPC Clusters. *2016 International Conference on Computational Intelligence (CSCI)*, 584-589. <https://doi.org/10.1109/CSCI.2016.0116>

Barber, M. R. W., Falasinnu, T., Ramsey-Goldman, R., & Clarke, A. E. (2023). The global epidemiology of SLE: Narrowing the knowledge gaps. *Rheumatology*, 62(Supplement\_1), i4-i9. <https://doi.org/10.1093/rheumatology/keac610>

Barquera, S., Pedroza-Tobías, A., Medina, C., Hernández-Barrera, L., Bibbins-Domingo, K., Lozano, R., & Moran, A. E. (2015). Global Overview of the Epidemiology of Atherosclerotic Cardiovascular Disease. *Archives of Medical Research*, 46(5), 328-338. <https://doi.org/10.1016/j.arcmed.2015.06.006>

Baumer, B., & Udwin, D. (2015). R Markdown. *WIREs Computational Statistics*, 7(3), 167-177. <https://doi.org/10.1002/wics.1348>

Beaman, R. S., Traub, G. H., Dell, C. A., Santiago, N., Koh, J., & Cellinese, N. (2012). TOLKIN - Tree of Life Knowledge and Information Network: Filling a Gap for Collaborative Research in Biological Systematics. *PLoS ONE*, 7(6), e39352. <https://doi.org/10.1371/journal.pone.0039352>

Becker, J., Arenas, C., Dean, A., Mconigliario, Vuksan, V., Li, B., Hawson, Rawks, D., Satterly, N., Buchbinder, J., Pocock, D., Massie, M., Brousse, N., & Nicholes, B. (2024). *Ganglia* [Software]. <https://github.com/ganglia>

Bruce, I. N., Urowitz, M. B., Gladman, D. D., Ibanez, D., & Steiner, G. (2003). Risk factors for coronary heart disease in women with systemic lupus erythematosus: The Toronto Risk Factor Study. *Arthritis & Rheumatism*, 48(11), 3159-3167. <https://doi.org/10.1002/art.11296>

Chang, W., Joe, C., Allaire, J., Siervet, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., Borges, B., Posit Software, P., & jQuery Foundation. (2024). *shiny: Web Application Framework for R* (Version 1.9.1) [Software]. <https://cran.r-project.org/web/packages/shiny/index.html>

Drenkard, C., & Lim, S. S. (2019). Update on lupus epidemiology: Advancing health disparities research through the study of minority populations. *Current Opinion in Rheumatology*, 31(6), 689-696. <https://doi.org/10.1097/BOR.0000000000000646>

Eckel, R. H., Grundy, S. M., & Zimmet, P. Z. (2005). The metabolic syndrome. *The Lancet*, 365(9468), 1415-1428. [https://doi.org/10.1016/S0140-6736\(05\)66378-7](https://doi.org/10.1016/S0140-6736(05)66378-7)

Furlani, J., Osel, P., Owen, R. K., Lakata, M., Mein, K., & Delaruelle, X. (2024). *Environment Modules* (Version Revision 672e287e) [Software].

Gaziano, T. A., Bittton, A., Anand, S., Abrahams-Gessel, S., & Murphy, A. (2010). Growing Epidemic of Coronary Heart Disease in Low- and Middle-Income Countries. *Current Problems in Cardiology*, 35(2), 72-115. <https://doi.org/10.1016/j.cpcardiol.2009.10.002>

Goulelmos, G. N., Zervou, M. I., Vazgiourakis, V. M., Ghodke-Puranik, Y., Garyfallos, A., & Niewold, T. B. (2018). The genetics and molecular pathogenesis of systemic lupus erythematosus (SLE) in populations of different ancestry. *Gene*, 668, 59-72. <https://doi.org/10.1016/j.gene.2018.05.041>

Hare, E., & Kaplan, A. (2017). Designing Modular Software: A Case Study in Introductory Statistics. *Journal of Computational and Graphical Statistics*, 26(3), 493-500. <https://doi.org/10.1080/10618600.2016.1276839>

Hernández-Ledesma, A. L., Martínez, D., Elizabeth, F.-B., Román-López, T. V., Karen, N.-R., Sandra Valentina, V. D. V., Donajil, O.-Z., Lizbet, T.-N., Angélica, P.-A., Estefanía, T.-V., Gabriel, F.-V., Maria, G.-A., Florencia, R., Sarael, A., Rentería, M. E., Ruiz-Contreras, A. E., Deshiré, A.-R., & Alejandra, M.-R. (2023). *Lupus*

RGMX: Social and Clinical Characteristics and their Contribution to Quality of Life in a Mexican Cohort with SLE. <https://doi.org/10.1101/2023.02.23.23286331>

Knoop, J., Zdun, U., Gesellschaft für Informatik, & Köllen Druck + Verlag (Eds.). (2016). *Software Engineering 2016: 23-26. Februar 2016, Wien, Österreich*. Tagung Software Engineering, Bonn, Gesellschaft für Informatik.

Lewis, M. J., & Jawad, A. S. (2016). The effect of ethnicity and genetic ancestry on the epidemiology, clinical features and outcome of systemic lupus erythematosus. *Rheumatology*, kew399. <https://doi.org/10.1093/rheumatology/kew399>

LUPUSRGMX. (2024). *LUPUSRGMX, Registro Mexicano de Lupus*. <https://lupusrgmx.liigh.unam.mx/quienes-somos.html>

Martínez, D. (2024a). *Lupus Shiny App* [Software]. [https://github.com/LuisDomingo/Lupus\\_ShinyApp](https://github.com/LuisDomingo/Lupus_ShinyApp)

Martínez, D. (2024b). *Notas Configuración Cluster* [Software]. [https://github.com/LuisDomingo/Notas\\_configuracion\\_Cluster](https://github.com/LuisDomingo/Notas_configuracion_Cluster)

Martínez, D. (2024c). *Red Bayesiana Hipertensión Lupus* [Software]. [https://github.com/LuisDomingo/Red\\_Bayesiana\\_Hipertension\\_Lupus](https://github.com/LuisDomingo/Red_Bayesiana_Hipertension_Lupus)

Mendoza-Pinto, C., Etcheberry-Morales, I., García-Carrasco, M., Munguia-Realpozo, P., Méndez-Martínez, S., Osorio-Peña, A. D., & Montiel-Jarquín, A. J. (2022). Twenty-year trends in all-cause mortality of patients with systemic lupus erythematosus in Mexico: Results from a nationwide health registry. *Lupus*, 31(3), 382-391. <https://doi.org/10.1177/09612032211078228>

Mensah, G. A., Roth, G. A., & Fuster, V. (2018). The Global Burden of Cardiovascular Diseases and Risk Factors. *Journal of the American College of Cardiology*, 74(20), 2529-2532. <https://doi.org/10.1016/j.jacc.2019.10.009>

Mok, C. C., Kwok, R. C., & Yip, P. S. F. (2013). Effect of Renal Disease on the Standardized Mortality Ratio and Life Expectancy of Patients With Systemic Lupus Erythematosus. *Arthritis & Rheumatism*, 65(8), 2154-2160. <https://doi.org/10.1002/art.38006>

Nikpour, M., Urowitz, M. B., Ibanez, D., Harvey, P. J., & Gladman, D. D. (2011). Importance of cumulative exposure to elevated cholesterol and blood pressure in development of atherosclerotic coronary artery disease in systemic lupus erythematosus: A prospective proof-of-concept cohort study. *Arthritis Research & Therapy*, 13(5), R156. <https://doi.org/10.1186/ar3473>

Ordoñez, J. M., Rios-Issa, D., Santos-Lozano, A., Lucia, A., Torres, A., Kosogodanov, A., & Camacho, J. M. (2023). A Bayesian network model for predicting cardiovascular risk. *Computer Methods and Programs in Biomedicine*, 231, 107405. <https://doi.org/10.1016/j.cmpb.2023.107405>

Parth, D., Pojia, D., Sharayu, M., & Tanaji, B. (2014). High Performance Computing Clusters. *International Journal of Computer Applications*, 105(3), 24-29. <https://doi.org/10.5120/18359-9494>

Payne, R. A. (2012). Cardiovascular risk. *British Journal of Clinical Pharmacology*, 74(3), 396-410. <https://doi.org/10.1111/j.1365-2125.2012.04219.x>

Pelaez-Ballestas, I., Sanin, L. H., Moreno-Montoya, J., Alvarez-Nemegyei, J., Burgos-Vargas, R., Garza-Elizondo, M., Rodríguez-Amado, J., Goycochea-Robles, M.-V., Madariaga, M., Zamudio, J., Santana, N., Cardiel, M. H., & Grupo de Estudio Epidemiológico de Enfermedades Musculo Articulares (GEEMA). (2011). Epidemiology of the Rheumatic Diseases in Mexico: A Study of 5 Regions Based on the COPCORD Methodology. *The Journal of Rheumatology Supplement*, 86(0), 3-8. <https://doi.org/10.3899/jrheum.100951>

Pradipta, G. A., Wardoyo, R., Muehlhoff, A., & Sanjaya, I. N. H. (2021). Radius-SMOTE: A New Oversampling Technique of Minority Samples Based on Radius Distance for Learning From Imbalanced Data. *IEEE Access*, 9, 74763-74777. <https://doi.org/10.1109/ACCESS.2021.3080316>

R Core Team. (2024). *The R Project for Statistical Computing* (Version 4.3.3) [Software]. <https://www.r-project.org/>

Reyes-Perez, P., Hernández-Ledesma, A. L., Román-López, T. V., García-Vilchis, B., Ramírez-González, D., Lázaro-Figueroa, A., Martínez, D., Flores-Ocampo, V., Espinosa-Méndez, I. M., Rentería, M. E., Ruiz-Contreras, A. E., Alcauter, S., & Medina-Rivera, A. (2023). Building national patient registries in Mexico: Insights from the MexOMICS Consortium. *En bioRxiv*. <https://doi.org/10.1101/2023.09.11.23295377>

Roth, G. A., Mensah, G. A., Johnson, C. O., Adolorato, G., Ammirati, E., Baddour, L. M., Barengo, N. C., Beaton, A. Z., Benjamin, E. J., Benzigler, C. P., Bonny, A., Brauer, M., Brodmann, M., Cahill, T. J., Carapetis, J., Catapano, A. L., Chugh, S. S., Cooper, L. T., Coresh, J., ... Fuster, V. (2020). Global Burden of Cardiovascular Diseases and Risk Factors, 1990-2019. *Journal of the American College of Cardiology*, 76(25), 2982-3021. <https://doi.org/10.1016/j.jacc.2020.11.010>

Salvaneschi, G., Margara, A., & Tamburelli, G. (2015). Reactive Programming: A Walkthrough. *2015 IEEE/ACM 37th IEEE International Conference on Software Engineering*, 953-954. <https://doi.org/10.1109/ICSE.2015.303>

Scutari, M., & Silander, T. (2004). *bnlearn: Bayesian Network Structure Learning, Parameter Learning and Inference* (Version 5.0.1) [R]. <https://cran.r-project.org/web/packages/bnlearn/index.html>

Slurm Developers Group. (2024). *Slurm workload manager* (Version 24.05) [https://slurm.schedmd.com/documentation.html].

Suo, X., Huang, X., Zhong, L., Luo, Q., Ding, L., & Xue, F. (2024). Development and Validation of a Bayesian Network-Based Model for Predicting Coronary Heart Disease Risk From Electronic Health Records. *Journal of the American Heart Association*, 13(1), e029400. <https://doi.org/10.1161/JAHA.123.029400>

Wagie, C., Allaire, J., Schloerke, C., Dervieux, C., Krumbiegel, J., Canouil, M., Woodhull, G., Lannone, R., Manning, A., Xie, Y., & Wickham, C. (2024). *Quarto* (Version 1.6.9) [Software]. <https://github.com/quarto-dev/quarto-cli>

Verla, H., Chidukuri, V., Mohammed, Y. J., Palakonda, V., & Chintala, R. R. (2023). HPC Driven Innovations in Network Management for Greater Efficiency and Productivity. *2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA)*, 1047-1053. <https://doi.org/10.1109/ICIRCA57980.2023.1022775>

Wickham, H., François, R., Henry, L., Müller, K., Vaughan, D., & Posit, S. (2023). *dplyr: A Grammar of Data Manipulation* (Version 1.1.4) [R]. <https://cran.r-project.org/web/packages/dplyr/index.html>

CRÉDITOS: Esta plantilla para presentaciones es una creación de Slidesgo, e incluye iconos de Flaticon, infografías e imágenes de Freepik

Por favor, conserva esta diapositiva para atribuirnos