

MLb-LDLr: LDLaren hartzailearen aldaeren eragina auresateko ikasketa automatikoko eredua

Larrea, Asier¹ ; Galicia, Unai¹ ; Jebari, Shifa¹ ; Alonso, Rocio¹ ; Benito-Vicente, Asier¹ ; Uribe B. Kepa²; Arraste, Sonia³; Gonzalez, Humberto³; Martin, Cesar¹

¹ Biofisika Unitatea (CSIC, UPV/EHU) eta Biokimika eta Biologia Molekularra Saila, UPV/EHU. 48080 Bilbo.
² Center for Cooperative Research in Biomaterials (CIC biomaGUNE), Basque Research and Technology Alliance (BRTA), Paseo de Miramon 182, 20014, Donostia
³ Kimika Organikoa II Saila, (UPV/EHU), 48940 Bilbo

1. Sarrera eta motibazioa

Hiperkolesterolemia Familiarra (HF) gaixotasun autosomiko gainartzaileko ohikoena da mundu mailan eta plasmako LDL-kolesterol (ingelesez Low-density Lipoprotein-Cholesterol edo LDL-C) maila altua du ezaugarri nagusi (1). Plasmako LDL-C pilaketak ez du sintomarik eragiten maila larrietara iritsi arte, gaixotasun kardiobaskularrak hedatzeko aukerak oso altuak diren arte. Arazo horrek **diagnosi goiztiarren** garrantzia azpimarratzen du. Hala ere, HF kasuen % 10 bakarrik diagnostikatzen da (2). LDL-Caren metabolismoan garrantzia duten geneetan gertaturiko mutazioek eragiten dute HFa, eta haien artean LDLaren hartzailea (**LDLR**) da ohikoena (3). Proteina horretan mutazio kaltegarriak gertatuz gero, zelula ezberdinen LDL-Caren barneraketa murrizten eta plasmako LDL-Caren kontzentrazioa handiagotzen da.

Gaur egun 3000 aldaera baino gehiago aurkitu dira LDLRn eta missense motakoak dira gehienak (4). Hala ere, gutxi dira era egokian karakterizatu diren aldaerak. Aldaera bakoitzaren eragina **aurreikusteko gai** diren tresna informatikoen erabilpena asko zabaldu da osasun-zientzietan azken urteotan. **Software** edo eredu horiek gene zein proteinaren hainbat ezaugarri erabiltzen dituzte haien patogenizitatea auresateko, eta era oso azkar eta fidagarrian ematen dituzte emaitzak. Guk proposatutako ereduak karakterizatutako LDLRaren aldaera guztiak erabiltzen ditu ikasketa automatikoko teknikak erabiliz, patogenoen eta onberen arteko ezberdintasunak zeintzuk diren ikasteko eta horren arabera aldaerak sailkatzeko.

2. Ikerketaren helburuak

Ikerketa honen helburua izan da LDLRaren aldaeren eragina asmatzeko gai den eredu bat sortzea. Hiru helburu nagusi izan ditugu:

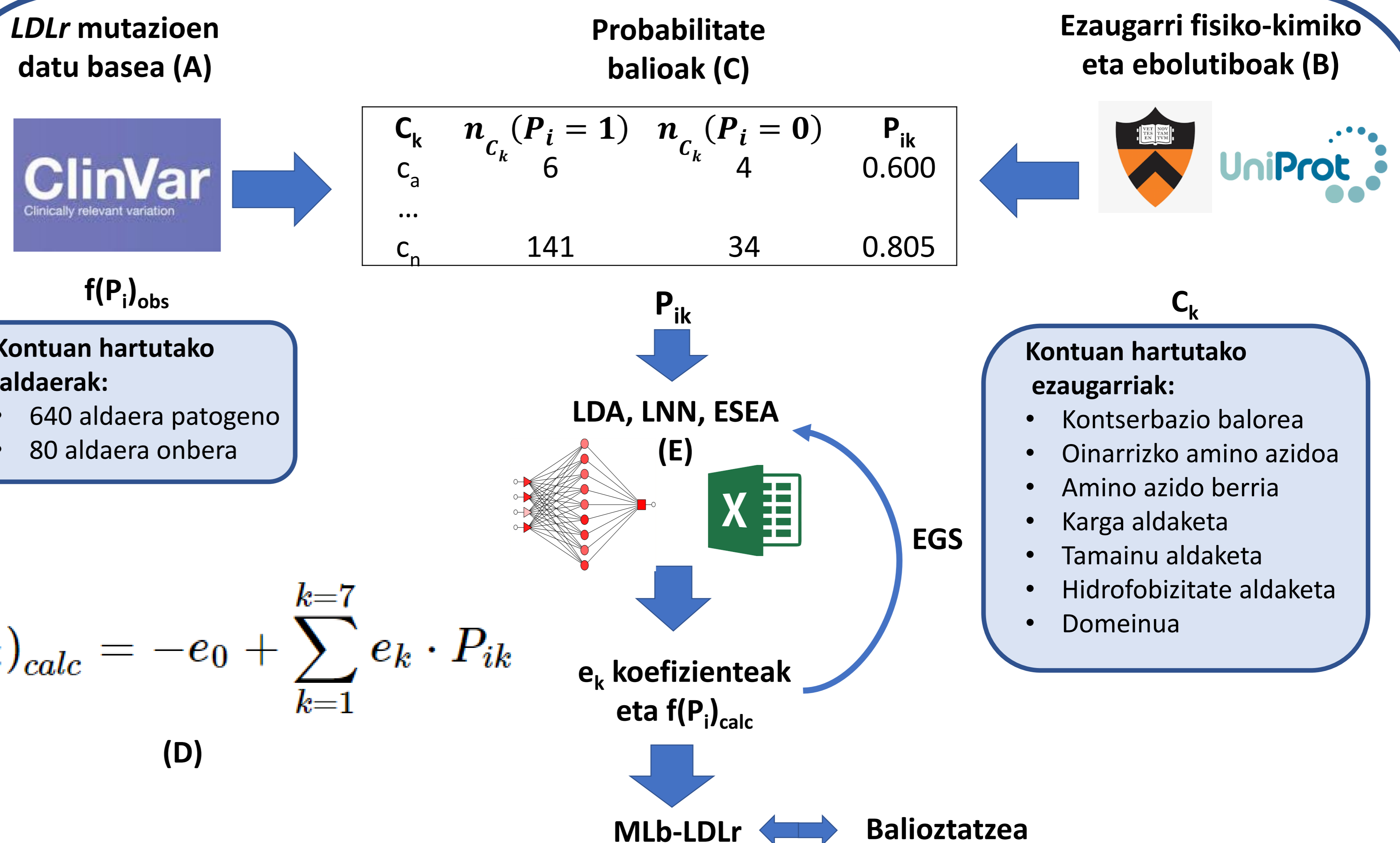
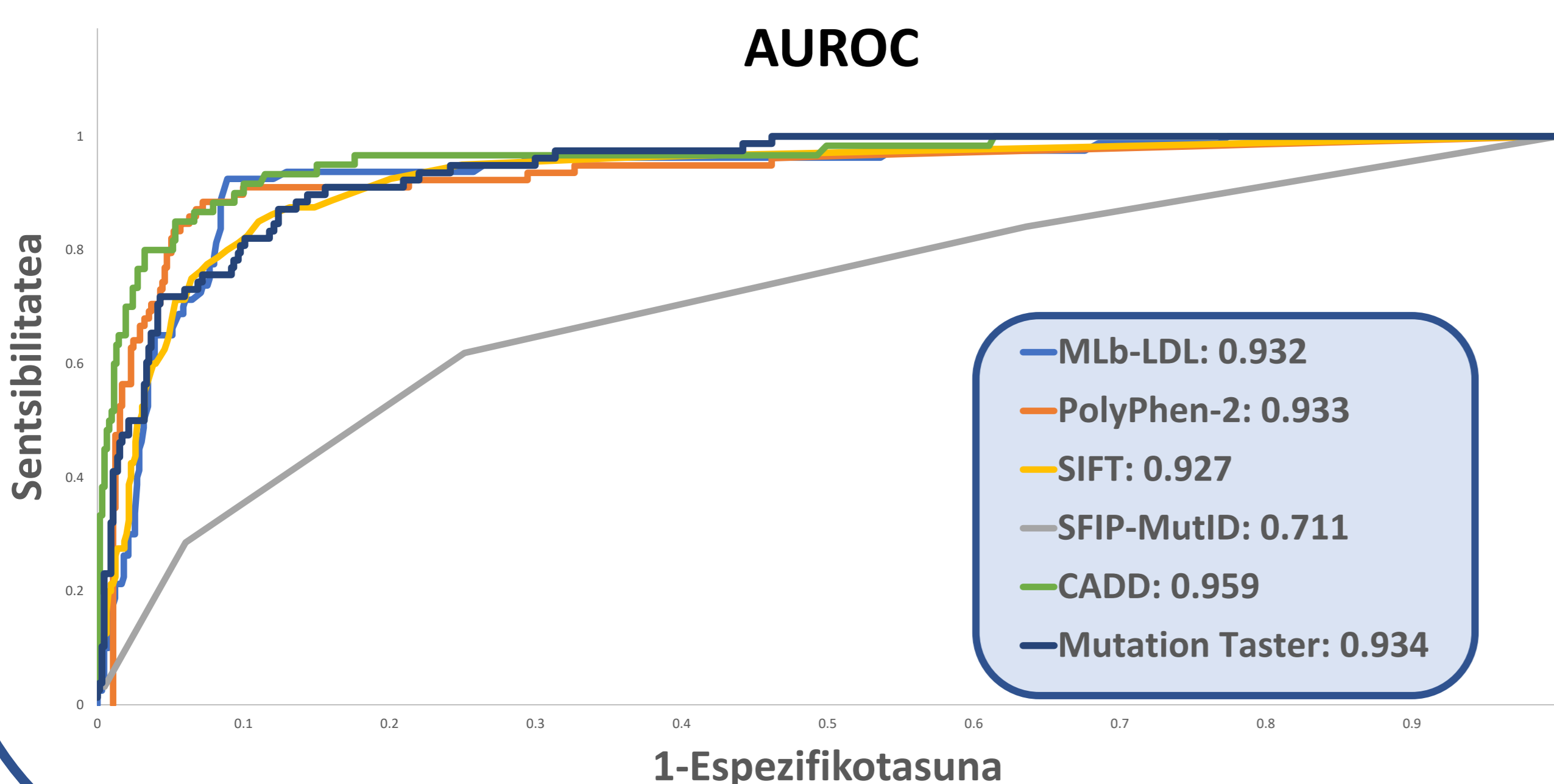
1. Eredu fidagarri bat sortzea
2. Era errazean eguneratzea
3. Beste gene batzuetarako erabilgarria izatea aldaketa gutxirekin

3. Ikerketaren muina

Metodologia

LDLR aldaeren inguruko informazioa ClinVar datu basetik lortu zen (A). Ondoren patogenizitatea zehaztea ahalbidetuko zuten ezaugarriak aukeratu ziren (B). Datu hauekin probailitate balioak lortu ziren: Aldaera baten ezaugarriak kontuan izanik, patogeno edo onbera izateko zer aukera dagoen (C). Behin balio hauek izanda, ereduak sortu zen (D). Hurrengo pausuan ezaugarri bakoitzaren garrantzia edo pisua (e_k) zehaztu zen, ereduaren asmatze tasa hobetzeko. Horretarako hainbat ikasketa automatikoko eredu erabili ziren, nahiz eta amaieran ESEA (Excel solver Evolutionary Algorithm) erabili bakarrik (E). Emaitzarik onenekin MLb-LDLr software-a sortu eta balioztatu zen.

Emaitzak



Zehaztasun espezifikoak

Software	Zehaztasuna
MLb-LDLr	0.918
PolyPhen-2	0.906
SIFT	0.871
SFIP-MutID	0.684
CADD	0.909
Mutation Taster	0.877

Gure ereduaren eraginkortasuna bi teknikekin neurtu zen:

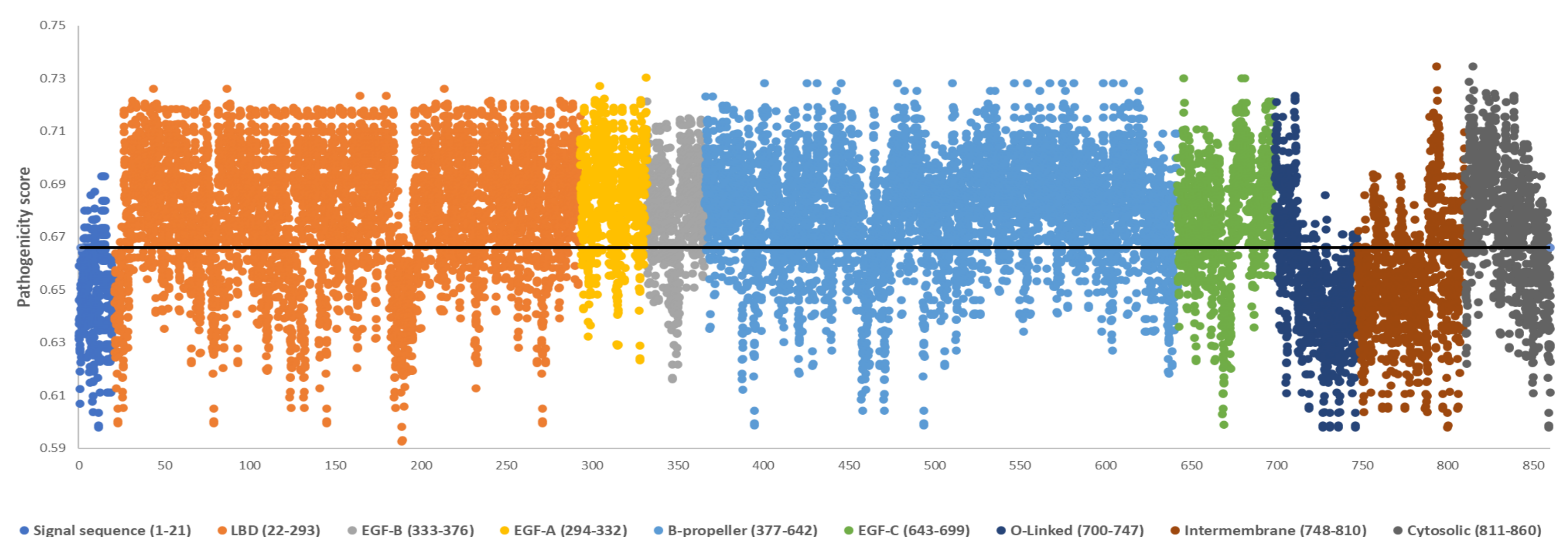
AUROC: Emaitza hau lortzeko software bakoitzaren sentsibilitatea vs 1-espezifiktasuna grafikatu behar da, ondoren kurbaren azpiko azalera lortzeko. Lortutako zenbakiak ereduaren sendotasuna erakusten du.

Zehaztasun espezifikoak: Emaitza hau egoera zehatz batean ereduak duen zehaztasuna erakusten du. Zehaztasun hau ereduaren optimoa izaten da. Software batean aldaera bat bilatzean normalean egoera horretan ematen du emaitza.

4. Ondorioak

Proiektu honek genituen helburu guztiak bete ditu: MLb-LDLr-k fidagarritasun oso handia erakutsi du, merkatuan dauden beste software batzuekin lehiatzeko gai izanik. Eguneratzeari dagokionez, modu oso azkar eta errazean egin daiteke, ESEA eta eskuz eginiko aldatetarako minutu gutxi batzuetan egin ahal baitira. Horrek ahalbidetzen du aldaera berriak agertzean haien datuak ereduaren sartzea, MLb-LDLr-ren fidagarritasuna handituz. Azkenik, ereduak beste proteina batzuen patogenizitatea asmatzeko gai izango litzatekeela uste dugu, aldaketa gutxi batzuekin.

Ereduaren erabilgarritasuna azpimarratzeko, LDLR-an gertatu daitezkeen 16340 aldaera posiblen eragina irudikatu dugu



5. Erreferentziak

- [1] Sjouke et al (2015) *European Heart Journal*, 36(9), 560–565
 [3] Palacios et al (2012) *Atherosclerosis*, 221(1), 137–142

- [2] Benito-Vicente et al (2018) *International Journal of Molecular Sciences*, 19 (11)
 [4] Chora et al (2012) *Nucleic Acids Research*, 40(W1), 452– 457