UAB
Universitat Autònoma de Barcelona

Dipòsit digital de documents de la UAB

This is the **published version** of the bachelor thesis:

Miguélez Montenegro, Aránzazu. *Sistema de Monitoratge Clínic i Predicció Glucèmica de pacients amb DM1.* Treball de Final de Grau (Universitat Autònoma de Barcelona), 2026 (Grau en Enginyeria de Dades)

This version is available at https://ddd.uab.cat/record/337331

under the terms of the CC BY-SA license.

Sistema de Triatge Clínic i Predicció Glucèmica per a Pacients amb DM1

Aránzazu Miguélez

Resum– En el seguiment clínic de la diabetis mellitus tipus 1 (DM1), un dels reptes és passar de la revisió individual de gràfiques a una visió agregada que permeti prioritzar pacients de manera ràpida. En aquest treball es presenta una arquitectura de monitoratge en temps real orientada al triatge clínic de pacients amb DM1. El sistema utilitza Apache Kafka, Spark Structured Streaming, InfluxDB i Grafana per monitoritzar 200 pacients simultanis, classificar automàticament el seu estat glucèmic en set categories de risc i mostrar-ne la prioritat en un dashboard poblacional. A més, s'hi integra una capa analítica de predicció glucèmica a 30 minuts, validada sobre una cohort independent de 20 pacients reals del MetaboNet. Aquesta segona capa es visualitza en un dashboard clínic que compara la glucosa real amb la glucosa estimada, amb l'objectiu de facilitar intervencions preventives abans que apareguin episodis crítics.

Paraules clau– DM1, CGM, Apache Kafka, Spark Structured Streaming, InfluxDB, Grafana, MetaboNet, LightGBM, Clarke Error Grid.

Abstract– In clinical follow-up of type 1 diabetes mellitus (T1DM), one of the main challenges is moving from the individual review of glucose traces to an aggregated view that helps prioritise patients quickly. This work presents a real-time monitoring architecture designed for clinical triage of T1DM patients. The system uses Apache Kafka, Spark Structured Streaming, InfluxDB and Grafana to monitor 200 simultaneous patients, classify their glycaemic status into seven risk categories and display their priority in a population-level dashboard. The architecture also includes a 30-minute glucose prediction layer, validated on an independent cohort of 20 real patients from MetaboNet. This second layer is shown in a clinical dashboard that compares observed and estimated glucose, with the goal of supporting preventive action before critical events occur.

Keywords– T1DM, CGM, Apache Kafka, Spark Structured Streaming, InfluxDB, Grafana, MetaboNet, LightGBM, Clarke Error Grid.

1 Introducció

La diabetis mellitus tipus 1 (DM1) és una malaltia crònica autoimmune en què el sistema immunitari destrueix les cèl·lules beta del pàncrees, responsables de la producció d'insulina [20]. Com que la insulina és necessària perquè la glucosa pugui ser utilitzada pels teixits, la seva manca obliga el pacient a controlar la glucosa de manera continuada i a ajustar el tractament al llarg del dia.

En la DM1, els nivells de glucosa poden variar ràpidament entre episodis d'*hipoglucèmia* i *hiperglucèmia*. Les hipoglucèmies poden provocar símptomes immediats com mareig, confusió o pèrdua de consciència, mentre que les hiperglucèmies mantingudes s'associen a complicacions cardiovasculars, renals, neurològiques i oculars. Per aquest motiu, el control glucèmic no és només una tasca de seguiment, sinó una part central del tractament.

Actualment, la DM1 afecta aproximadament 9,2 milions de persones arreu del món [1]. El seu maneig diari és complex perquè la glucosa depèn de múltiples factors: l'alimentació, l'activitat física, l'estrès, el descans, les dosis d'insulina i altres condicions de salut. Aquesta variabilitat fa que dues persones amb el mateix diagnòstic puguin tenir perfils glucèmics molt diferents.

Els sistemes de monitoratge continu de glucosa, coneguts com a CGM (*Continuous Glucose Monitoring*), han millorat aquest seguiment. Aquests sensors mesuren la glucosa cada pocs minuts i generen, habitualment, una lectura cada 5 minuts, és a dir, fins a 288 mesures diàries per pacient. Aquest volum de dades permet observar tendències i detectar canvis que serien difícils d'identificar només amb controls capil·lars puntuals.

Tot i això, el problema canvia quan el seguiment no és individual, sinó poblacional. Un endocrinòleg pot haver de supervisar desenes de pacients alhora, i revisar manualment totes les corbes de glucosa no és eficient. Les plataformes comercials actuals són útils per consultar l'evolució d'un pacient concret, però ofereixen menys suport quan cal de-

• E-mail de contacte: aranzazu.miguelez@autonoma.cat
• Menció realitzada: Enginyeria de Dades
• Treball tutoritzat per: Antonio Espinosa
• Curs 2025/26

tectar, en una cohort sencera, qui necessita atenció prioritària en aquell moment.

Aquest TFG parteix d'aquesta necessitat i proposa un sistema de monitoratge clínic en temps real. L'arquitectura s'organitza en quatre capes principals: ingestió, processament, emmagatzematge i visualització. A més, incorpora una capa analítica de predicció glucèmica. L'objectiu és passar d'una visualització descriptiva de la glucosa actual a una eina de suport a la decisió: identificar quins pacients requereixen atenció immediata i estimar com pot evolucionar la seva glucosa durant els propers 30 minuts.

La Secció 2 concreta els objectius específics, la Secció 3 revisa l'estat de l'art, la Secció 4 descriu l'arquitectura del sistema, la Secció 5 presenta la metodologia de la capa analítica i la Secció 6 recull els resultats obtinguts.

2 OBJECTIUS

El sistema està pensat per a professionals sanitaris, especialment endocrinòlegs i diabetòlegs, que supervisen pacients amb DM1 en entorns hospitalaris o d'atenció especialitzada. En el projecte es diferencien dues cohorts: una cohort de 200 pacients per validar el pipeline de monitoratge en temps real (Secció 4), i una cohort independent de 20 pacients reals per entrenar i avaluar la capa predictiva (Secció 5). Aquesta separació respon a una limitació pràctica: l'entrenament de models amb dades de diversos anys per pacient té un cost computacional elevat.

L'objectiu general és construir una eina que ajudi a prioritzar pacients i a anticipar situacions de risc. En concret, el sistema ha de permetre: (1) detectar automàticament pacients amb risc glucèmic sense revisar manualment cada historial; (2) predir la glucosa a 30 minuts per facilitar intervencions preventives; i (3) consultar l'evolució retrospectiva de cada pacient com a suport a la presa de decisions clíniques.

Per què cal agregar la informació? L'agregació de dades glucèmiques a escala de cohort aporta tres beneficis principals:

(1) *Triatge automàtic*: el professional pot veure en un únic dashboard quins pacients requereixen atenció prioritària, sense haver d'obrir una gràfica individual per a cada cas.

(2) *Detecció anticipada*: els models predictius estimen la glucosa a 30 minuts vista. Això dona marge per actuar abans que una hipoglucèmia o hiperglucèmia arribi a ser crítica. A més, les prediccions es poden interpretar amb variables com la insulina activa, els carbohidrats recents o la tendència glucèmica.

(3) *Anàlisi retrospectiva*: el sistema permet comparar pacients amb perfils similars i observar patrons de cohort, de manera que les decisions no depenguin només de casos individuals.

La Taula 1 resumeix els objectius específics del projecte.

TAULA 1: OBJECTIUS ESPECÍFICS DEL TFG

OE	Descripció
OE1	Preprocessament de dades reals (MetaboNet)
OE2	Pipeline de *streaming*: Kafka → Spark → InfluxDB
OE3	Classificació clínica automàtica (7 categories)
OE4	Dashboard Grafana i sistema d'alertes
OE5	Predicció glucèmica a 30 minuts

3 ESTAT DE L'ART

3.1 Tecnologia CGM: sensors i estàndards

Els sistemes de monitoratge continu de glucosa (CGM, *Continuous Glucose Monitoring*) han canviat de manera important el seguiment de la DM1. A diferència dels controls capil·lars puntuals, aquests dispositius permeten observar l'evolució de la glucosa de forma contínua al llarg del dia.

El funcionament es basa en un petit filament subcutani que mesura la concentració de glucosa al líquid intersticial. Les lectures s'envien al telèfon mòbil o al receptor del pacient, normalment cada 5 minuts. Això representa fins a 288 mesures diàries per pacient i converteix la DM1 en un escenari amb una gran densitat de dades biomèdiques.

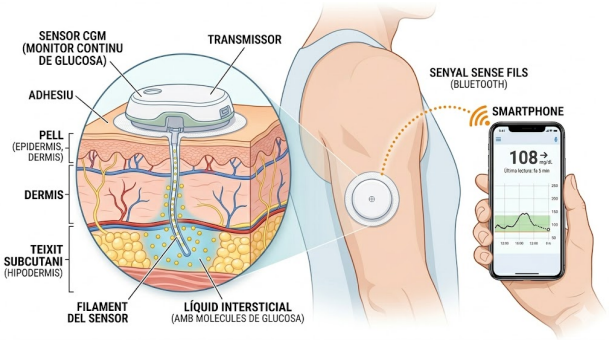


Fig. 1: Secció transversal d'un sensor CGM: el filament mesura la glucosa al líquid intersticial i transmet per Bluetooth a l'app del pacient.

La glucosa intersticial no coincideix exactament amb la glucosa sanguínia. Hi ha un retard fisiològic aproximat de 5–10 minuts respecte de la glucosa capil·lar, sobretot quan la glucosa canvia ràpidament. Aquest retard és conegut i es té en compte en els algoritmes de calibratge i filtratge dels dispositius actuals.

Entre els sensors comercials més utilitzats hi ha el Dexcom G6/G7 [3] i el FreeStyle Libre 3 d'Abbott [4]. Aquests sistemes ofereixen una precisió clínica elevada, no requereixen calibratge manual i es poden integrar amb sistemes automatitzats d'administració d'insulina (*closed-loop systems*).

En l'avaluació clínica del control glucèmic també s'utilitza el *Time in Range* (TIR) [2], que mesura el percentatge de temps que el pacient es manté dins del rang segur de 70–180 mg/dL. De manera complementària, es defineixen el *Time Below Range* (TBR) i el *Time Above Range* (TAR), amb diferents nivells de gravetat: TBR L1 (54–69 mg/dL), TBR L2 ($<$ 54 mg/dL), TAR L1 (181–250 mg/dL) i TAR L2 ($>$ 250 mg/dL). En persones amb DM1, les guies recomanen un TIR $\geq 70\,\%$, ja que s'associa amb un menor risc de complicacions cròniques.

3.2 Eines de monitoratge poblacional

Malgrat la utilitat dels sensors CGM, moltes plataformes comercials continuen centrades en el seguiment individual. Eines com Dexcom Clarity o LibreView permeten consul-

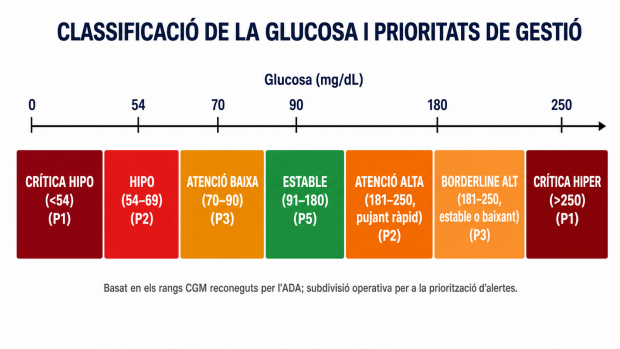


Fig. 2: Esquema dels rangs glucèmics utilitzats en l'avaluació clínica del Time in Range (TIR) en DM1.

tar dades històriques i generar informes, però no estan orientades a prioritzar cohorts completes de pacients en temps real.

Hauptman et al. [6] van estudiar la integració de dades CGM dins la història clínica electrònica hospitalària, amb l'objectiu de facilitar-ne la consulta retrospectiva. Tot i així, aquest enfocament no incorporava processament en temps real ni una ordenació automàtica dels pacients segons risc clínic. De manera semblant, Ferstad et al. [9] van proposar una estratègia de gestió poblacional basada en dades CGM agregades, però el sistema també es plantejava com una anàlisi periòdica i retrospectiva.

Així, encara hi ha marge per desenvolupar eines que combinin en un mateix flux la ingestió contínua de dades multipacient, el processament en *streaming*, la classificació del risc clínic, la predicció glucèmica i una visualització operativa per al professional sanitari.

3.3 Arquitectures de streaming per a salut

En un sistema de seguiment continu de pacients amb DM1, les dades arriben de manera constant des de múltiples sensors. Per això, la infraestructura ha de suportar ingestió contínua, processament amb baixa latència i tolerància a fallades.

Apache Kafka encaixa bé en aquesta necessitat perquè permet gestionar fluxos de dades d'alta freqüència mitjançant un model distribuït i persistent. A més, pot escalar horitzontalment afegint nous *brokers*. Per la seva banda, Apache Spark Structured Streaming permet processar aquests fluxos de manera distribuïda i calcular mètriques sobre finestres temporals amb latència reduïda.

Ed-daoudy et al. [10] van proposar un sistema basat en Kafka i Spark Streaming per predir l'estat de salut a partir de dispositius IoT mèdics. Fernández-Gómez et al. [11] van presentar un marc basat en Spark per al pronòstic de sèries temporals en entorns IoT. Aquests treballs donen suport a l'ús de Kafka i Spark [18] com a base tecnològica per a arquitectures de monitoratge en temps real.

3.4 Predicció glucèmica: revisió i selecció de model

La predicció de glucosa a curt termini és un problema rellevant en diabetologia digital. En aquest treball s'ha fixat un horitzó de 30 minuts perquè proporciona un marge d'actuació raonable abans d'episodis d'hipoglucèmia o hiperglucèmia severa, i és coherent amb el temps d'acció de les insulines ràpides modernes.

La literatura inclou aproximacions basades tant en aprenentatge automàtic clàssic com en *deep learning*. Els models profunds poden ser precisos en determinats escenaris, però sovint necessiten més dades, més recursos computacionals i són més difícils d'interpretar. Oviedo et al. [12] assenyalen que els models clàssics d'aprenentatge automàtic poden oferir un bon equilibri entre precisió, robustesa i interpretabilitat quan es treballa amb conjunts de dades de mida mitjana.

Per aquest motiu s'han considerat models com Random Forest, LightGBM i XGBoost, que permeten obtenir valors SHAP i, per tant, explicar quines variables contribueixen a cada predicció. Aquesta interpretabilitat és especialment important en un context clínic, on el professional necessita entendre per què el model estima una pujada o baixada de glucosa.

Wolff et al. [19] van publicar GluPredKit, una biblioteca Python de codi obert que estandarditza el preprocessament, l'enginyeria de *features* i l'avaluació de models de predicció glucèmica. En aquest TFG, GluPredKit s'ha utilitzat en una fase exploratòria per validar el plantejament inicial del pipeline. Els resultats finals, però, s'han obtingut amb una reimplementació pròpia, per tenir control directe sobre els paràmetres experimentals i sobre la integració amb la resta del sistema.

Finalment, s'ha incorporat una anàlisi d'explicabilitat amb SHAP (*SHapley Additive exPlanations*). Aquesta tècnica permet quantificar la contribució de variables com la glucosa recent, la insulina activa, els carbohidrats ingerits o la tendència temporal en cada predicció individual.

4 Arquitectura del sistema

L'arquitectura del sistema es divideix en quatre capes principals: ingestió, processament, emmagatzematge i visualització. Sobre aquestes capes s'afegeix una capa analítica de predicció glucèmica. La separació per capes facilita el manteniment del sistema i permet substituir una part concreta —per exemple, el model predictiu o la base de dades temporal— sense haver de redissenyar tota la infraestructura.

La infraestructura s'executa amb un únic fitxer `docker-compose.yml`, que orquestra set contenidors Docker. Aquesta decisió es va prendre per facilitar la reproduïbilitat del projecte i evitar dependències manuals en la màquina d'execució. Vegeu l'Annex C per als detalls de configuració.

4.1 Capa 1 — Ingestió de dades

El punt de partida del sistema és el dataset MetaboNet [5], format per dades retrospectives. Com que aquestes dades no es generen en temps real, s'ha implementat un mecanisme de *replay*: un productor Python llegeix les lectures en ordre cronològic i les publica a Kafka [18] simulant l'arribada contínua d'un sensor CGM.

Ús de Kafka en el projecte. Apache Kafka actua com a capa intermèdia entre la generació de lectures i el seu processament. En aquest cas, permet publicar les 288 lec-

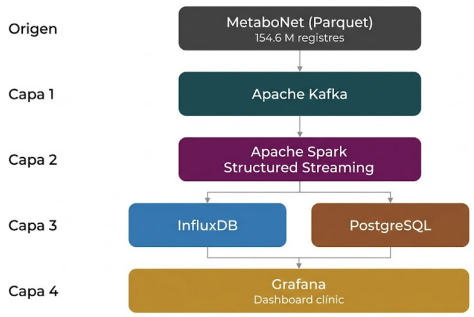


Fig. 3: Arquitectura global del sistema en quatre capes. Les dades flueixen des del dataset MetaboNet fins al dashboard clínic passant per Kafka, Spark, InfluxDB i Grafana. La capa analítica (línia discontínua) és intercanviable i no interfereix amb el pipeline principal.

tures diàries de cadascun dels 200 pacients, és a dir, fins a 57.600 lectures/dia, sense que el consumidor hagi de processar cada missatge exactament en el mateix instant en què es publica.

Les lectures s'organitzen al tòpic `glucose-readings` amb 4 particions. El camp `patient_id` s'utilitza com a clau de partició per mantenir l'ordre cronològic de les lectures dins de cada pacient.

Abans del *replay*, les dades es preprocessen amb DuckDB [21] directament sobre els fitxers Parquet de MetaboNet. Això evita carregar en memòria els 154,6 milions de registres. El preprocessament inclou: (1) selecció de pacients segons criteris de qualitat; (2) normalització temporal dels *timestamps*; (3) divisió de les dades en dos blocs de 6 mesos, un bloc estàtic i un bloc de *replay*; i (4) exportació de les metadades demogràfiques a PostgreSQL.

4.2 Capa 2 — Processament en temps real

Apache Spark Structured Streaming [22] consumeix els missatges de Kafka i els processa en finestres lliscants de 5 minuts, amb un desplaçament d'1 minut. Per a cada finestra es calcula la glucosa mitjana, el pendent de variació glucèmica (mg/dL per minut), la desviació estàndard i els valors mínim i màxim.

Elecció de la finestra temporal. La durada de 5 minuts s'ha escollit perquè coincideix amb la freqüència habitual dels sensors CGM. D'aquesta manera, cada finestra pot incorporar com a mínim una lectura real sense perdre massa resolució temporal, un aspecte important quan es volen detectar canvis glucèmics ràpids.

A partir d'aquestes mètriques, Spark assigna una de **set categories clíniques** a cada pacient. Aquestes categories refinen les cinc zones glucèmiques estàndard del TIR (Figura 2) [2] afegint-hi el criteri de pendent de variació: per exemple, la franja 70–90 mg/dL s'identifica com `ATENCIO_BAIXA` independentment del rang 181–250 mg/dL amb tendència ascendent (`ATENCIO_ALTA`), permetent detectar situacions de risc emergent abans que el pacient surti del TIR. Aquesta subdivisió no modifica la definició clínica del TIR, que continua considerant tot el rang 70–180 mg/dL com a "en rang". Cada categoria té associada una prioritat (P1 = màxima urgència fins a P5 = sense alerta), que després s'utilitza per ordenar els pacients al dashboard. La categoria `UNKNOWN` (P4) s'utilitza quan no hi ha prou dades per fer una classificació fiable; per tant, no representa un risc glucèmic concret. L'Annex C detalla les set categories i els llindars clínics complets.

4.3 Capa 3 — Emmagatzematge dual

Els resultats del processament es guarden en dos sistemes amb funcions diferents:

InfluxDB [23] s'utilitza per a les sèries temporals. En aquesta base de dades s'emmagatzemen tres *measurements*: `patient_status`, amb l'estat clínic calculat per finestra temporal; `cgm_reading`, amb les lectures individuals de glucosa; i `glucose_prediction`, amb les prediccions generades per la capa analítica.

PostgreSQL s'utilitza per a les metadades estàtiques dels pacients, com l'edat, el gènere, el tipus de sensor CGM, el règim d'insulina (AID/SAP/MDI), l'edat al diagnòstic i l'estudi d'origen. Aquesta separació evita barrejar dades temporals d'alta freqüència amb informació clínica que canvia poc o no canvia durant l'experiment.

4.4 Capa 4 — Visualització clínica

Grafana [24] és la capa que consulta el professional mèdic. El sistema connecta Grafana amb InfluxDB per obtenir les dades temporals i amb PostgreSQL per recuperar les metadades dels pacients.

El **dashboard de triatge** mostra els 200 pacients ordenats per prioritat clínica i s'actualitza en temps real. També inclou indicadors com TIR, GMI, CV i comptadors d'episodis crítics per al pacient seleccionat.

El **dashboard de predicció** compara la glucosa real amb la glucosa predita a +30 minuts. També mostra mètriques individuals del model, com RMSE, Clarke A i MCC, consultades des de PostgreSQL. L'Annex B inclou captures dels dashboards i les queries Flux utilitzades.

5 Metodologia de la capa analítica

5.1 Visió general i motivació

La capa de predicció glucèmica té una finalitat concreta: estimar l'evolució de la glucosa amb prou antelació perquè el professional pugui actuar abans que aparegui un episodi de risc. En aquest treball s'ha fixat l'horitzó de predicció en 30 minuts, ja que és un interval curt però clínicament útil.

Des del punt de vista tècnic, el problema es formula com una **regressió de sèries temporals**. A partir de l'historial glucèmic i clínic disponible fins a l'instant t, el model prediu la glucosa als instants $t{+}5, t{+}10, \ldots, t{+}30$ minuts.

Aquesta capa s'ha dissenyat de manera independent respecte al pipeline de *streaming*. Els models s'entrenen *offline* amb dades del dataset MetaboNet i les prediccions resultants s'escriuen a InfluxDB, dins del *measurement* `glucose_prediction`. Des d'aquí, Grafana les pot consultar i representar juntament amb la glucosa real.

5.2 Cohort experimental i selecció

La validació de la capa analítica s'ha fet amb una cohort independent de **20 pacients reals** del dataset MetaboNet (1.406.267 registres bruts, rang temporal 2014–2024;

1.405.889 registres un cop aplicat el *gap* temporal de 90 minuts descrit a la Secció 5.4). Aquesta cohort és diferent de la cohort de 200 pacients utilitzada per al pipeline de *streaming*. La decisió respon a una limitació pràctica: MetaboNet pot contenir fins a 7 anys de dades per pacient, i entrenar sis models sobre una cohort molt gran en un ordinador personal sense GPU no era viable dins del temps disponible. La cohort de 20 pacients manté diversitat clínica i tecnològica, però permet fer una comparació experimental controlada.

La selecció s'ha fet seguint dos criteris:

Criteri de qualitat: cada pacient havia de tenir com a mínim 20.000 lectures CGM vàlides dins d'una finestra de 12 mesos. El filtre s'aplica directament sobre les lectures vàlides de la finestra temporal, per evitar incloure pacients amb moltes files però poques mesures reals. Aquest criteri és independent del que s'aplica a la cohort de 200 pacients del pipeline de *streaming* (Annex C).

Criteri de diversitat tecnològica: s'ha intentat incloure almenys un pacient per tipus de sensor CGM (Dexcom G4, G5, G6, One, Abbott Libre, FreeStyle Libre 2, Medtronic Enlite, Guardian 3 i Guardian 4). Això ajuda a reduir el risc que els models aprenguin artefactes d'un únic sensor. La cohort final inclou 10 pacients amb Dexcom G6, 5 amb Dexcom G4, 2 amb Dexcom One i 1 pacient per a cadascun dels 5 sensors restants. Després del preprocessament i l'aplicació del *gap* temporal, el conjunt queda en 1.405.889 registres.

5.3 Enginyeria de *features*

Cada mostra inclou un conjunt de *features* que descriuen l'estat del pacient a l'instant t: **Historial glucèmic recent:** s'utilitzen els 12 valors de glucosa anteriors a t (lags de 5, 10, . . . , 60 minuts), la mitjana mòbil, la desviació estàndard dels darrers 30 minuts i el pendent instantani. Aquest grup de variables resumeix el nivell actual de glucosa i la direcció del canvi.

Insulina activa (IOB): s'estima amb un filtre de decaïment exponencial sobre les dosis administrades, amb una semivida de 55 minuts, coherent amb el comportament de les insulines ràpides modernes [7]. Aquesta variable és important perquè la insulina activa condiciona directament l'evolució posterior de la glucosa.

Carbohidrats: es calcula la suma de carbohidrats ingerits durant els 60 minuts anteriors. Aquesta variable té cobertura parcial: el 66,8% dels registres de MetaboNet tenen valor de carbohidrats i el 33,2% restant és nul. Es manté dins del model perquè alguns algorismes, com LightGBM, poden gestionar valors nuls de manera nativa.

Variables temporals: l'hora del dia es codifica de forma circular amb $\sin(2\pi h/24)$ i $\cos(2\pi h/24)$. Aquesta codificació evita la discontinuïtat entre les 23h i les 0h i permet capturar patrons circadians de sensibilitat a la insulina, com el *dawn phenomenon* [8].

5.4 Partició temporal i prevenció de *data leakage*

La partició train/test s'ha fet de manera cronològica per a cada pacient: el 80% inicial de les lectures s'utilitza per entrenar i el 20% final per avaluar. Aquesta estratègia és més realista que una divisió aleatòria, perquè en un escenari real el model sempre prediria dades futures a partir de dades passades.

El punt de tall s'estableix segons la data corresponent al percentil 80 dels timestamps amb CGM vàlid, i no simplement segons el nombre de files. Això evita que pacients amb moltes lectures nul·les concentrades al final acabin tenint un conjunt de test poc representatiu.

Per reduir el risc de *data leakage*, s'ha introduït un **gap temporal de 90 minuts** (18 mostres de 5 minuts) entre el final del train i l'inici del test. Aquest marge separa la *feature* més llunyana utilitzada pel model ($t-60$ min) i la target més llunyana ($t+30$ min) de qualsevol dada del train. Amb aquesta partició, el conjunt final queda format per 1.138.678 mostres de *train* (81,0%), 267.211 mostres de *test* (19,0%) i 378 mostres eliminades pel *gap* (0,03%).

5.5 Models avaluats i parametrització

S'han comparat **sis estratègies** sota les mateixes condicions experimentals. Les dues primeres són baselines senzills i les quatre restants són models d'aprenentatge automàtic. Tots els models entrenats s'han implementat amb la configuració `MultiOutputRegressor` de `scikit-learn`: 6 estimadors independents, un per a cada horitzó temporal (5, 10, 15, 20, 25 i 30 minuts).

- **ZOH** — baseline que assumeix que la glucosa futura serà igual a l'última lectura disponible.
- **Naive Linear** — extrapolació lineal de la tendència recent.
- **Ridge** — regressió lineal amb regularització L_2; α seleccionat per *cross-validation*.
- **Random Forest** — 200 arbres, submostreig de 50.000 mostres per pacient.
- **LightGBM** — gradient boosting amb GOSS; 300 estimadors, submostreig de 50.000 mostres. Gestiona nativament els valors nuls de `carbs`.
- **XGBoost** — gradient boosting per nivells; 200 estimadors, submostreig de 30.000 mostres.

L'Annex D recull els hiperparàmetres complets i la justificació de les principals decisions de configuració.

5.6 Enfocament de model per pacient

S'ha optat per entrenar un model **independent per a cada pacient**, en lloc d'un únic model global per a tota la cohort. La motivació és clínica: la resposta glucèmica varia molt entre persones, fins i tot quan comparteixen diagnòstic i tipus de sensor. Un model global tendeix a aprendre el comportament mitjà de la cohort, mentre que un model individual pot adaptar-se millor als patrons propis de cada pacient [12].

Per aquest motiu, les mètriques es presenten com a **mitjana $\pm$ desviació estàndard** sobre els 20 pacients. La desviació estàndard és important perquè mostra si el model funciona de manera consistent o si el bon rendiment mitjà queda concentrat en pocs pacients.

6 RESULTATS

6.1 Pipeline de streaming

El pipeline s'ha validat sobre una cohort de 200 pacients simultanis amb cinc tipus de sensors CGM. La prova ha consistit a publicar una lectura CGM a Kafka i mesurar el temps que triga a aparèixer classificada al dashboard de Grafana. En les 10 proves realitzades, la latència màxima observada ha estat de 58 s, per sota del llindar de 60 s definit a l'objectiu OE2.

Amb càrrega sostinguda de 200 pacients, la latència es manté per sota dels 40 s per batch, amb un *Input Rate* de 3.000–6.000 registres/s i un ús de memòria d'estat proper a 1 GB. La taula de triatge ordena automàticament els pacients per prioritat clínica i s'actualitza en temps real, de manera que es compleixen els objectius OE3 i OE4.

6.2 Capa analítica de predicció

6.2.1 Mètriques de qualitat predictiva

La Taula 2 recull els resultats en el conjunt de test dels 20 pacients, expressats com a mitjana ± desviació estàndard. Com a referència, s'ha considerat un RMSE de 18 mg/dL com a llindar clínicament acceptable per a predicció glucèmica a 30 minuts en diabetologia digital [2].

TAULA 2: COMPARATIVA DE MODELS A 30 MINUTS (20 PACIENTS, MODEL PER PACIENT). RMSE EN MG/DL, MITJANA ± DESVIACIÓ ESTÀNDARD.

Model	RMSE (mg/dL)	R^2	Cl. A	Cl. E
ZOH	26,58±6,69	0,870	75,3%	0,091%
Naive Lin.	29,12±11,96	0,838	73,0%	0,247%
Ridge	18,39±5,43	0,933	84,3%	0,030%
Rnd. Forest	19,17±5,68	0,928	83,5%	0,015%
LightGBM	**18,24±5,36**	**0,936**	**84,9%**	0,022%
XGBoost	18,26±5,39	0,935	85,0%	**0,018%**

Lectura dels resultats. El RMSE expressa l'error quadràtic mitjà en mg/dL i penalitza especialment els errors grans. El R^2 indica la proporció de variabilitat glucèmica explicada pel model. Clarke A correspon al percentatge de prediccions clínicament acceptables segons la Clarke Error Grid [17], mentre que Clarke E identifica els errors de risc màxim, és a dir, prediccions que podrien portar a una decisió terapèutica inadequada.

Resultats principals. Els models entrenats redueixen l'error entre un 28% i un 31% respecte al baseline ZOH (26,58 mg/dL). LightGBM obté el menor RMSE global (18,24 mg/dL) i el millor R^2 (0,936). XGBoost (18,26 mg/dL) i Ridge (18,39 mg/dL) queden molt a prop. Com que la diferència entre aquests tres models és inferior a 0,2 mg/dL, no es pot afirmar una superioritat clara només a partir del RMSE.

La desviació estàndard entre pacients (5,4–5,7 mg/dL) reflecteix la variabilitat clínica de la cohort. Hi ha pacients amb RMSE inferior a 13 mg/dL i perfils glucèmics estables, i altres amb RMSE superior a 25 mg/dL. La Figura 4 mostra que Ridge, LightGBM i XGBoost tenen distribucions semblants. En canvi, el Naive Linear presenta dos *outliers* clars (71 i 43 mg/dL), associats a canvis glucèmics bruscos on l'extrapolació lineal genera valors fisiològicament impossibles.

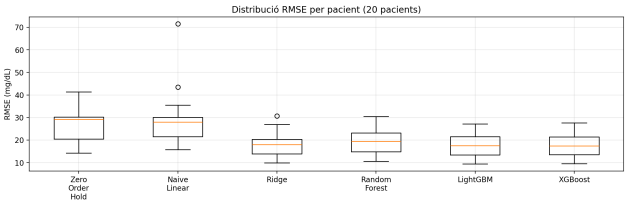


Fig. 4: Distribució del RMSE per pacient i model. Ridge, LightGBM i XGBoost presenten medianes similars (≈18 mg/dL).

Evolució per horitzó temporal. La Figura 5 mostra que, a 5 minuts, les diferències entre models són petites. En un interval tan curt, la glucosa acostuma a canviar poc i el ZOH encara és un baseline competitiu. A mesura que l'horitzó s'allarga, l'avantatge dels models entrenats augmenta. A 30 minuts, la separació entre LightGBM/XGBoost/Ridge (≈26 mg/dL agregat) i ZOH (≈35 mg/dL) arriba a gairebé 9 mg/dL. El Naive Linear es degrada ràpidament perquè és sensible a inversions brusques de tendència.

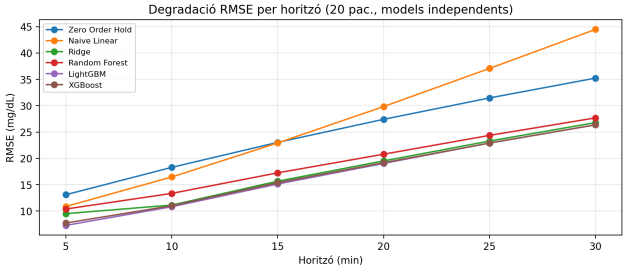


Fig. 5: RMSE per horitzó de predicció (5–30 min). L'avantatge dels models entrenats és mínim a 5 minuts i màxim a 30 minuts, on la separació respecte al ZOH és de ≈9 mg/dL.

6.2.2 Mètriques clíniques complementàries

El RMSE és útil per comparar models, però no indica si els errors es concentren en zones segures o en zones de risc. Per aquest motiu, la Taula 3 incorpora mètriques amb interpretació clínica.

TAULA 3: MÈTRIQUES CLÍNIQUES COMPLEMENTÀRIES A 30 MINUTS (20 PACIENTS). GRMSE: RMSE PONDERAT PER ZONES CRÍTIQUES. MCC: CAPACITAT DE DETECCIÓ D'EPISODIS DE RISC. TG: ANTICIPACIÓ EFECTIVA EN MINUTS SOBRE ZOH.

Model	gRMSE	MCC_{hypo}	MCC_{hyper}	TG (min)
ZOH	36,78	0,287	0,710	0,00
Naive Lin.	44,90	0,381	0,739	−2,82
Ridge	**27,64**	**0,355**	0,785	9,24
Rnd. Forest	29,27	0,220	0,778	7,58
LightGBM	27,84	0,318	0,789	**9,40**
XGBoost	27,78	0,287	**0,791**	9,34

El **gRMSE** dona més pes als errors en zones crítiques: triplica els errors en hipoglucèmia severa (<54 mg/dL) i duplica els d'hiperglucèmia severa (>250 mg/dL). Ridge obté

el millor gRMSE (27,64 mg/dL), cosa que indica que els seus errors són lleugerament menys problemàtics en zones de risc.

El **MCC** (Matthews Correlation Coefficient) avalua la capacitat de detectar episodis de risc i penalitza tant falsos positius com falsos negatius. MCC_{hypo} mesura la detecció d'hipoglucèmies futures (<70 mg/dL), mentre que MCC_{hyper} mesura la detecció d'hiperglucèmies (>180 mg/dL). Ridge obté el millor MCC_{hypo} (0,355), i XGBoost el millor MCC_{hyper} (0,791). El PCC (0,929 per als tres models principals) i el MARE (12,38–12,42%) aporten poca discriminació addicional, perquè les diferències són molt petites.

El **Temporal Gain** (TG) mesura els minuts d'anticipació respecte al ZOH. LightGBM obté el millor valor (9,40 min). El valor negatiu del Naive Linear (−2,82 min) mostra que extrapolar linealment pot ser pitjor que assumir que la glucosa es mantindrà constant.

6.2.3 Clarke Error Grid

La Taula 4 mostra la distribució completa de zones. Les zones A i B no induirien decisions terapèutiques incorrectes; les zones C, D i E representen errors progressivament perillosos.

TAULA 4: CLARKE ERROR GRID (%) A 30 MINUTS. ZONES A+B: CLÍNICAMENT ACCEPTABLES. ZONA E: RISC MÀXIM.

Model	A	B	C	D	E
ZOH	75,3	15,5	4,4	4,6	0,091
Naive Lin.	73,0	15,3	3,2	8,2	0,247
Ridge	84,3	10,4	2,7	2,6	0,030
Rnd.Forest	84,2	10,4	3,4	2,0	**0,015**
LGB	**84,9**	10,0	3,0	2,1	0,022
XGBoost	85,0	9,9	3,1	2,0	0,018

Els quatre models entrenats concentren entre el 94,2% i el 94,9% de les prediccions en les zones A+B. Random Forest és un cas interessant: tot i tenir el pitjor RMSE del grup, obté la menor proporció de prediccions a Clarke E (0,015%). Això mostra que el RMSE no és suficient per valorar un model en un context clínic. Un model pot tenir un error mitjà més alt i, alhora, cometre menys errors greus.

6.2.4 Anàlisi de trajectòries

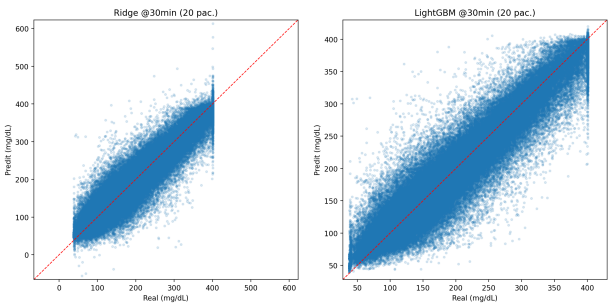


Fig. 6: Scatter predicció vs. real a 30 min per Ridge (esquerra) i LightGBM (dreta). La distribució és simètrica al voltant de la diagonal sense biaixos sistemàtics.

Les mètriques agregades resumeixen el comportament global del model, però no sempre expliquen com es comporta en pacients concrets. Per això també s'han revisat exemples individuals de trajectòries i scatter plots.

Els scatter plots de la Figura 6 indiquen que Ridge i LightGBM no presenten un biaix clar cap a la sobreestimació o la subestimació. La majoria de punts es distribueixen al voltant de la diagonal, tot i que la dispersió augmenta en valors glucèmics extrems.

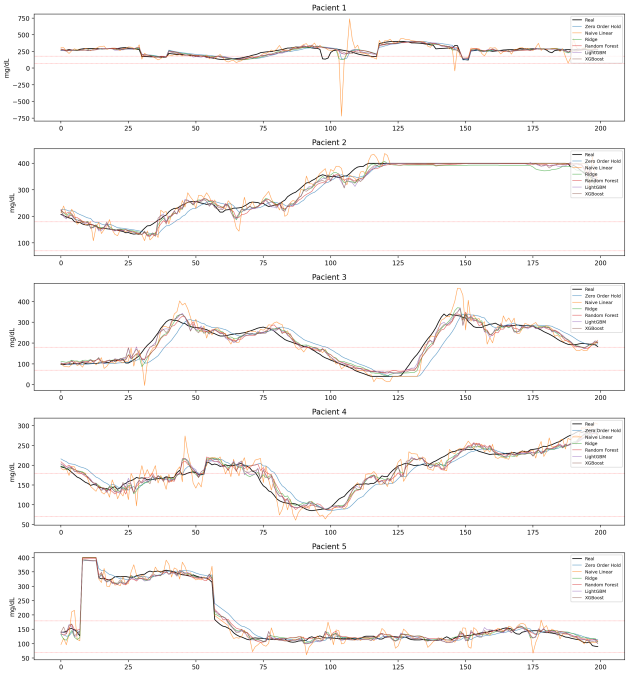


Fig. 7: Trajectòries de glucosa predita per als sis models (200 mostres de test, cinc pacients). La línia negra és la glucosa real. Les línies vermelles indiquen els llindars de 70 i 180 mg/dL.

La Figura 7 permet veure diferències que no apareixen tan clarament en les taules. En el Pacient 1, amb glucosa estabilitzada entre 250 i 300 mg/dL, el Naive Linear falla en un canvi brusc i arriba a predir −750 mg/dL, un valor fisiològicament impossible. En el Pacient 3, amb alta variabilitat i descens cap a 70 mg/dL, LightGBM i XGBoost segueixen millor el canvi. En el Pacient 5, que presenta un descens ràpid de 400 a 100 mg/dL, LightGBM és el model que millor s'adapta a la transició.

6.2.5 Interpretabilitat: anàlisi SHAP

En un entorn clínic, la precisió del model no és l'únic criteri important. També cal poder explicar per què s'obté una predicció determinada. Sense aquesta informació, el model queda com una caixa negra i és més difícil que el professional pugui valorar si la predicció és coherent amb l'estat del pacient.

Per analitzar aquesta qüestió, s'han calculat valors SHAP (*SHapley Additive exPlanations* [13]) sobre el model LightGBM. Els valors SHAP indiquen quina contribució té cada *feature* en la predicció: valors positius empenyen la predicció cap a glucosa més alta, i valors negatius cap a glucosa més baixa.

La Figura 8 mostra un *beeswarm plot* agregat sobre 300 mostres × 20 pacients. Cada punt representa una mostra individual i el color indica el valor de la *feature* (rosat = alt, blau = baix).

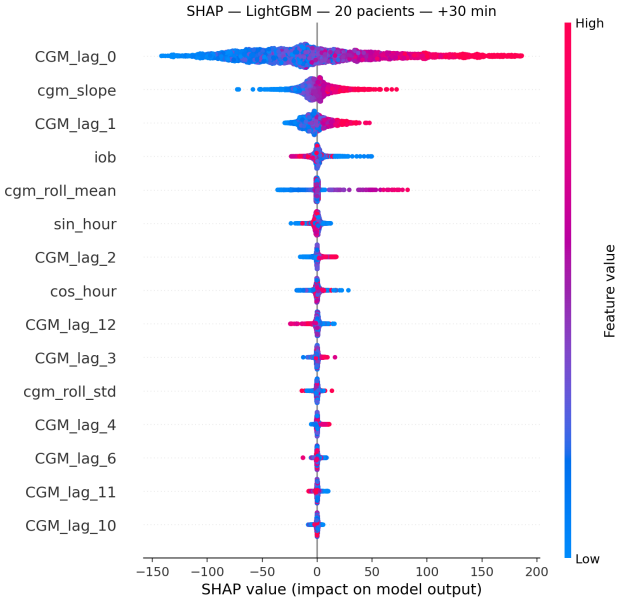


Fig. 8: Beeswarm plot SHAP per a LightGBM (20 pacients, 300 mostres/pacient, horitzó +30 min). Color rosat: valor alt de la *feature*; blau: valor baix.

L'anàlisi mostra quatre patrons principals:

Valor actual de glucosa. `CGM_lag_0` és la *feature* amb més impacte. Quan el valor actual és alt, la contribució SHAP és positiva i empeny la predicció cap amunt; quan és baix, la contribució és negativa.

Tendència recent. `cgm_slope` també té un pes elevat. Una pendent ascendent incrementa la predicció, mentre que una pendent descendent la redueix. Això indica que el model no només té en compte el valor puntual de glucosa, sinó també la seva dinàmica immediata.

Insulina activa. `iob` apareix entre les variables més rellevants. Quan la insulina activa és alta, la contribució SHAP tendeix a ser negativa, cosa coherent amb una baixada posterior de glucosa.

Variables horàries. `sin_hour` i `cos_hour` també tenen pes en el model. Això suggereix que LightGBM aprofita patrons circadians, incloent variacions de sensibilitat a la insulina al llarg del dia.

En conjunt, l'anàlisi SHAP indica que LightGBM basa les seves prediccions en variables fisiològicament coherents. Aquesta interpretabilitat és un argument addicional a favor del seu ús com a model principal del sistema.

6.2.6 Eficiència computacional

A banda de la qualitat predictiva, també s'ha mesurat el cost computacional dels models. Aquesta part és rellevant perquè, en un desplegament real, el sistema hauria de generar prediccions per a molts pacients cada 5 minuts.

TAULA 5: BENCHMARK COMPUTACIONAL PER MODEL (PACIENT DE REFERÈNCIA, MITJANA DE 5 REPETICIONS). T. SINGLE: INFERÈNCIA SOBRE 1 MOSTRA. T. BATCH: 5.000 MOSTRES. T. ENTR.: TEMPS D'ENTRENAMENT PER PACIENT.

Model	Mida (MB)	T. single (ms)	T. batch (ms)	T. entr. (s)
Ridge	<1	3,9	11,0	1–4
XGBoost	4,9	13,6	54,4	6–15
LightGBM	7,4	17,6	159,7	3–12
Random Forest	531,0	205,5	315,2	5–22

Random Forest és el model més costós: ocupa 531 MB per pacient, que equivaldria aproximadament a 106 GB per a 200 pacients. Aquesta mida el fa poc pràctic per a un desplegament amb cohorts grans. En canvi, LightGBM (7,4 MB), XGBoost (4,9 MB) i Ridge (<1 MB) són opcions molt més manejables.

Tots els models compleixen el requisit de producció: una predicció ha d'estar disponible abans del següent cicle del sensor CGM, és a dir, en menys de 5 minuts. Ridge és el més ràpid en inferència individual (3,9 ms), seguit d'XGBoost (13,6 ms) i LightGBM (17,6 ms). Random Forest és molt més lent (205,5 ms), tot i que encara queda per sota del límit operatiu. En inferència batch, XGBoost és més eficient que LightGBM, fet que pot ser útil en reanàlisis retrospectives de cohorts. Pel que fa a entrenament, LightGBM és el model de gradient boosting més ràpid (3–12 s per pacient).

7 DISCUSSIÓ I CONCLUSIONS

7.1 Síntesi dels resultats

El sistema desenvolupat cobreix els objectius OE1–OE5. El pipeline permet ingerir lectures CGM, processar-les en temps real, classificar 200 pacients simultanis i visualitzar el triatge clínic amb una latència inferior a 60 s. A més, la capa analítica compara sis estratègies de predicció sobre 20 pacients reals del MetaboNet. Dins d'aquesta comparació, LightGBM és el model seleccionat per al sistema final (RMSE 18,24 mg/dL, R^2 0,936, Clarke A 84,9% i TG 9,40 min).

7.2 Perfils de cada model

Els resultats mostren que no hi ha un únic model que sigui millor en totes les dimensions. Per això, la selecció del model no s'ha fet només amb el RMSE, sinó considerant també mètriques clíniques i cost computacional. La Taula 6 resumeix aquesta comparació.

TAULA 6: SÍNTESI COMPARATIVA. **NEGRETA**: MILLOR VALOR PER COLUMNA. ↓: MENOR ÉS MILLOR. ↑: MAJOR ÉS MILLOR. T.INF: INFERÈNCIA SINGLE (MS).

Model	RMSE↓	MAE↓	gRMSE↓	MCC_{hypo}↑	TG↑	MB↓	T.inf↓
Ridge	18,39	12,05	**27,64**	**0,355**	9,24	<1	**3,9**
Rnd.For.	19,17	13,43	29,27	0,220	7,58	531	205
LGB	**18,24**	**11,95**	27,84	0,318	**9,40**	7,4	17,6
XGBoost	18,26	11,99	27,78	0,287	9,34	**4,9**	13,6

LightGBM és el model amb menor RMSE (18,24 mg/dL), menor MAE (11,95 mg/dL) i major Temporal Gain (9,40 min). També té una mida assumible en disc (7,4 MB

per pacient), de manera que és viable per a una cohort de 200 pacients.

Ridge no és el millor en RMSE, però presenta avantatges importants: millor gRMSE, millor MCC_{hypo}, mida negligible i latència d'inferència molt baixa. Això el fa especialment interessant si la prioritat clínica és detectar hipoglucèmies futures o si els recursos computacionals són limitats.

XGBoost queda molt a prop de LightGBM en qualitat predictiva i és més eficient en inferència batch. Per aquest motiu, podria ser una bona opció per a reanàlisi retrospectiva de cohorts o per a escenaris on la mida del model sigui una restricció important.

Random Forest queda descartat per a un desplegament amb cohorts grans per la seva mida en disc. Tot i això, obté la menor proporció de prediccions a Clarke E. Aquest resultat és rellevant perquè mostra que un model pot tenir pitjor RMSE i, alhora, menys errors clínicament greus.

7.3 Selecció del model per a producció

La diferència de RMSE entre LightGBM, XGBoost i Ridge és inferior a 0,2 mg/dL. Per tant, no seria rigorós justificar la selecció només amb aquesta mètrica. La decisió s'ha basat en un criteri multicriteri: qualitat predictiva, anticipació, interpretabilitat, gestió de valors nuls i cost operatiu.

S'ha seleccionat **LightGBM** com a model principal perquè combina quatre avantatges en el context específic d'aquest sistema: (1) millor RMSE i MAE globals; (2) major anticipació efectiva respecte al ZOH (9,40 min); (3) gestió nativa dels valors nuls de `carbs`, que representen el 33,2% dels registres i que altres models han d'imputar explícitament; i (4) temps d'entrenament baix dins dels models de gradient boosting, fet que facilita el reentrenament periòdic per pacient.

Cal precisar que Ridge obté millors resultats en dues mètriques de seguretat clínica: gRMSE (27,64 vs 27,84 mg/dL) i MCC_{hypo} (0,355 vs 0,318). Aquests valors indiquen que els errors de Ridge es concentren menys en zones de risc i que detecta millor les hipoglucèmies futures. En un sistema orientat a la detecció d'episodis individuals crítics, Ridge seria probablement l'opció preferent. En canvi, en el context d'aquest TFG —monitoratge poblacional continu de 200 pacients, amb prediccions cada 5 minuts i dades de carbohidrats parcialment absents— el major Temporal Gain i la robustesa davant de valors nuls inclinen la balança cap a LightGBM com a millor opció. Ridge es manté com la recomanació principal per a desplegaments amb recursos computacionals molt limitats o amb focus exclusiu en hipoglucèmies crítiques.

7.4 Limitacions i treball futur

La primera limitació és que la capa analítica treballa amb dades retrospectives. Els models s'entrenen *offline* amb MetaboNet i les prediccions s'avaluen sobre el conjunt de test. La integració completa amb el pipeline de *streaming*, incloent actualització incremental dels models a mesura que arriben noves lectures CGM, queda com a treball futur.

La segona limitació és la mida de la cohort analítica. Els 20 pacients permeten fer una validació experimental amb diversitat de sensors i perfils clínics, però no són suficients per afirmar superioritat estadística entre models tan propers. Un estudi amb una part més gran de MetaboNet, o amb la cohort completa de 1.895 pacients, permetria aplicar tests com Wilcoxon o Diebold-Mariano per quantificar millor les diferències observades.

La tercera limitació és l'absència de variables contextuals. El model actual utilitza historial CGM, insulina activa, carbohidrats i variables temporals, però no incorpora activitat física, son, estrès, hàbits alimentaris ni estat emocional. Aquestes variables podrien explicar part de la variabilitat glucèmica individual. Integrar-les mitjançant wearables, aplicacions de seguiment o qüestionaris estructurats permetria construir models més personalitzats i potencialment anticipar episodis de risc amb més marge temporal.

Com a extensions addicionals, es plantegen quatre línies: (1) avaluar models de *deep learning*, com LSTM o Transformers, especialment per a horitzons superiors a 60 minuts; (2) integrar valors SHAP directament al dashboard; (3) ampliar la cohort analítica per obtenir una validació estadística més robusta; i (4) explorar entrenament en línia (*online learning*) per adaptar els models a noves lectures de cada pacient.

7.5 Conclusions generals

Aquest TFG ha desenvolupat un sistema complet de monitoratge clínic en temps real per a pacients amb DM1. El sistema cobreix el flux des de la ingestió de lectures CGM fins a la visualització del triatge i la predicció glucèmica en un dashboard clínic.

Els resultats mostren que és possible combinar en una mateixa arquitectura tres funcionalitats: triatge poblacional en temps real, predicció glucèmica a 30 minuts i explicabilitat de les prediccions. La part de *streaming* valida el seguiment de 200 pacients amb latència inferior a 60 s, mentre que la capa analítica obté una precisió clínicament acceptable amb LightGBM.

Des del punt de vista metodològic, el treball també mostra la importància d'avaluar els models amb més d'una mètrica. En aquest cas, models amb RMSE gairebé idèntic presenten diferències en gRMSE, MCC, Temporal Gain, mida i temps d'inferència. Aquesta lectura multicriteri és especialment important en entorns clínics, on no tots els errors tenen el mateix impacte.

Finalment, l'arquitectura s'ha dissenyat perquè sigui modular: la capa analítica pot evolucionar sense modificar el pipeline principal. Això facilita futures millores, com l'entrenament en línia, la incorporació de variables contextuals o la substitució del model predictiu per alternatives més avançades.

8 Seguiment de la planificació

El projecte s'ha desenvolupat seguint la planificació prevista, organitzada en tres fases principals: preparació de dades, implementació del pipeline de *streaming* i desenvolupament de la capa analítica. En general, els objectius principals s'han assolit, tot i que s'han fet alguns ajustos per adaptar el projecte al temps disponible i als recursos computacionals.

El principal ajust ha estat separar la validació del pipeline

TAULA 7: ESTAT FINAL DE LES TASQUES

Tasca	Estat	%
Pipeline streaming 200 pacients	Complet	100%
Classificació clínica 7 categories	Complet	100%
Dashboard Grafana + triatge	Complet	100%
Models ML per pacient (6 models)	Complet	100%
Integració predicció al dashboard	Complet	100%

de *streaming*, feta amb 200 pacients, de l'avaluació de la capa predictiva, feta amb una cohort independent de 20 pacients. Aquesta decisió es va prendre pel volum de dades de MetaboNet i pel cost computacional d'entrenar diversos models per pacient. Tot i aquest canvi, l'abast principal del TFG s'ha mantingut i els objectius definits inicialment s'han complert.

AGRAÏMENTS

Vull expressar el meu més sincer agraïment al meu tutor, per la seva orientació, dedicació i suport al llarg del Treball. Les seves aportacions i el seu guiatge han estat fonamentals per al desenvolupament d'aquest projecte.

A la meva família i a la meva parella, pel seu suport incondicional i l'ànim constant durant tot aquest procés. En especial, a la meva mare, perquè la seva experiència convivint amb la diabetis ha despertat el meu interès per aquesta malaltia i ha estat la principal inspiració per escollir aquest tema. La seva fortalesa i capacitat de superació han estat una font de motivació constant.

Finalment, vull donar les gràcies a totes aquelles persones que, d'una manera o altra, han contribuït a fer possible la realització d'aquest treball.

REFERENCES

[1] International Diabetes Federation, "IDF Diabetes Atlas, 10th Edition," 2021.

[2] T. Battelino et al., "Clinical Targets for CGM Data Interpretation," *Diabetes Care*, vol. 42, pp. 1593–1603, 2019.

[3] Dexcom Inc., "Dexcom Clarity," 2024. https://clarity.dexcom.com/

[4] Abbott, "LibreView," 2024. https://www.libreview.com/

[5] M. K. Wolff et al., "MetaboNet: The Largest Publicly Available Consolidated Dataset for T1DM," *arXiv*:2601.11505, 2026.

[6] A. Hauptman et al., "Integration of CGM Data into an EHR System," *J. Diabetes Sci. Technol.*, vol. 19, pp. 426–430, 2023.

[7] A. Haidar et al., "The Artificial Pancreas," *Diabetes Care*, vol. 36, pp. 1468–1474, 2013.

[8] P. J. Campbell et al., "Pathogenesis of the dawn phenomenon in patients with insulin-dependent diabetes mellitus," *New England Journal of Medicine*, vol. 315, pp. 1473–1481, 1986.

[9] J. O. Ferstad et al., "Population-level management of T1DM via CGM," *Pediatric Diabetes*, vol. 22, pp. 982–991, 2021.

[10] A. Ed-daoudy et al., "Applying Apache Spark on Streaming Big Data for Health Status Prediction," *Computers, Materials & Continua*, vol. 70, 2022.

[11] A. M. Fernández-Gómez et al., "A new Apache Spark-based framework for big data streaming forecasting in IoT," *J. Supercomputing*, vol. 79, pp. 11078–11100, 2023.

[12] S. Oviedo et al., "A review of personalized blood glucose prediction strategies for T1DM," *Int. J. Quality in Health Care*, vol. 29, pp. 706–733, 2017.

[13] S. M. Lundberg i S. I. Lee, "A Unified Approach to Interpreting Model Predictions," *NeurIPS*, vol. 30, 2017.

[14] F. Pedregosa et al., "Scikit-learn: ML in Python," *JMLR*, vol. 12, pp. 2825–2830, 2011.

[15] G. Ke et al., "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," *NeurIPS*, vol. 30, 2017.

[16] T. Chen i C. Guestrin, "XGBoost: A Scalable Tree Boosting System," *ACM SIGKDD*, pp. 785–794, 2016.

[17] W. L. Clarke et al., "Evaluating Clinical Accuracy of Systems for Self-Monitoring of Blood Glucose," *Diabetes Care*, vol. 10, pp. 622–628, 1987.

[18] N. Marz i J. Warren, *Big Data: Principles and Best Practices*. Manning, 2015.

[19] M. K. Wolff et al., "GluPredKit," *JOSS*, vol. 9, 2024.

[20] American Diabetes Association, "Standards of Medical Care in Diabetes," *Diabetes Care*, vol. 47, Suppl. 1, 2024.

[21] M. Raasveldt i H. Mühleisen, "DuckDB: an Embeddable Analytical Database," *ACM SIGMOD*, pp. 1981–1984, 2019. https://duckdb.org

[22] Apache Software Foundation, "Structured Streaming Programming Guide," 2024. https://spark.apache.org/docs/latest/structured-streaming-programming-guide.html

[23] InfluxData Inc., "InfluxDB Documentation," 2024. https://docs.influxdata.com/influxdb/

[24] Grafana Labs, "Grafana Documentation," 2024. https://grafana.com/docs/

A ANNEX A — EXPLORACIÓ DEL DATASET METABONET

Aquest annex resumeix l'exploració inicial del dataset MetaboNet, utilitzat com a base de dades principal del projecte. El dataset conté un total de 154,6 milions de registres, 1.895 pacients i 13 estudis clínics diferents. L'anàlisi s'ha dut a terme amb un script Python basat en DuckDB, que permet executar consultes SQL directament sobre el fitxer Parquet sense carregar-lo completament a memòria.

L'objectiu principal d'aquesta exploració ha estat identificar la composició del dataset, la disponibilitat de variables clíniques i demogràfiques, els dispositius CGM emprats i els règims d'insulina disponibles. Aquesta anàlisi ha servit per definir els criteris de selecció de pacients utilitzats posteriorment en el pipeline de preprocessament.

A.1 Distribució per estudi d'origen

La Taula 8 mostra els principals estudis identificats dins del dataset. L'estudi Loop és el més representat, amb més del 61% dels registres totals, seguit de ReplaceBG i IOBP2.

TAULA 8: DISTRIBUCIÓ RESUMIDA DE REGISTRES PER ESTUDI D'ORIGEN

Estudi	Registres	Pacients	%
Loop	94.810.573	845	61,3
ReplaceBG	20.864.819	208	13,5
IOBP2	9.676.885	332	6,3
Flair	6.997.510	113	4,5
DCLP5	6.096.063	100	3,9
DCLP3	6.031.492	112	3,9
PEDAP	5.507.753	65	3,6
Altres	4.666.716	120	3,0
Total	**154.651.111**	**1.895**	**100**

A.2 Sensors CGM i règim d'insulina

Pel que fa als dispositius CGM, el sensor majoritari és el Dexcom G6, utilitzat pel 63,7% dels pacients amb informació disponible. També apareixen altres dispositius com Dexcom G4, Dexcom G5, Medtronic Guardian 3 i FreeStyle Libre 2.

TAULA 9: DISTRIBUCIÓ RESUMIDA PER SENSOR CGM

Sensor	Pacients	%
Dexcom G6	1.059	63,7
Dexcom G4	224	13,5
Dexcom G5	128	7,7
Medtronic Guardian 3	113	6,8
FreeStyle Libre 2	27	1,6
Altres / desconegut	91	5,5
Total	**1.642**	**100**

La majoria de pacients utilitzen sistemes AID, és a dir, sistemes de bucle tancat automàtic. Aquest fet és coherent amb la naturalesa del dataset, orientat a dades de monitoratge continu i tractament intensiu de la diabetis tipus 1.

TAULA 10: DISTRIBUCIÓ PER RÈGIM D'INSULINA

Modalitat	Pacients	Descripció
AID	1.226	Bucle tancat automàtic
SAP	228	Bomba amb sensor, manual
MDI	28	Injeccions manuals

A.3 Cobertura de variables

L'anàlisi de cobertura ha permès classificar les variables segons la seva disponibilitat real. Les variables principals, com l'edat, l'identificador del pacient, el dispositiu CGM, el tipus d'insulina, les lectures de glucosa i les dosis d'insulina, presenten una cobertura elevada. En canvi, variables com la freqüència cardíaca, els passos, les etiquetes d'àpats o les dades ambientals tenen una cobertura residual i no s'han considerat adequades per a l'anàlisi principal.

TAULA 11: COBERTURA RESUMIDA DE LES PRINCIPALS VARIABLES DEL DATASET

Variable	Cobertura	Grup
age, id, source_file	100,0%	Alta
insulin_delivery_modality	99,1%	Alta
cgm_device	94,6%	Alta
weight, height	94,1%	Alta
age_of_diagnosis	94,1%	Alta
ethnicity	93,6%	Alta
insulin	90,8%	Alta
bolus	88,8%	Alta
basal	88,1%	Alta
CGM	84,5%	Alta
carbs	66,8%	Parcial
calories_burned	23,6%	Parcial
heartrate	0,6%	Residual
steps	0,6%	Residual
meal_label	0,1%	Residual
workout_*, air_temp	<0,1%	Residual

A.4 Característiques clíniques i temporals

Les lectures de glucosa CGM presenten una mitjana de 148,2 mg/dL, una mediana de 135,0 mg/dL i un rang entre 1,8 i 500,9 mg/dL. Aquests valors són coherents amb una cohort de pacients amb diabetis tipus 1 en condicions de vida real.

Els estudis amb millor cobertura temporal són ReplaceBG, Loop i Flair, ja que inclouen pacients amb seguiments superiors o iguals a 12 mesos. Aquest criteri temporal ha estat clau per seleccionar els pacients utilitzats en el sistema de monitoratge i predicció.

B ANNEX B: DASHBOARD CLÍNIC: IMPLEMENTACIÓ I CAPTURES

B.1 Construcció del dashboard amb Grafana i InfluxDB

El dashboard s'ha construït amb Grafana [24] connectat a InfluxDB [23] com a *data source* principal i a PostgreSQL per a les metadades demogràfiques. Cada panell utilitza consultes en llenguatge **Flux** adaptades al tipus de visual-

ització: panells *stat* per als indicadors puntuals (TIR, TBR, CV, GMI, episodis crítics), *bar gauge* per a la distribució de categories, *table* per a la taula de triatge i *time series* per a l'evolució temporal de la glucosa.

El TIR es calcula com el percentatge de finestres amb glucosa mitjana en rang 70–180 mg/dL; el GMI amb $\text{GMI} = 3{,}31 + 0{,}02392 \times \overline{\text{CGM}}$; i el CV com $\sigma/\mu \times 100$. Una variable de tipus *query* consulta dinàmicament tots els `patient_id` disponibles i actualitza tots els panells en seleccionar un pacient. Les sèries temporals incorporen *thresholds* que pinten el fons en verd (70–180 mg/dL), groc (180–250 mg/dL) i vermell (>250 o <70 mg/dL), coincidint amb l'estàndard TIR [2].

Els *timestamps* es desplacen de manera que el dataset estàtic cobreix 2025-07-01 → 2026-01-01 i el de replay 2026-01-01 → 2026-07-01, preservant les diferències relatives entre lectures. Els valors clínics no es modifiquen.

B.2 Dashboard de triatge poblacional

La Fig. 9 mostra el dashboard de triatge en funcionament per al pacient `1107_Loop` (home, 12 anys, estudi Loop, 18 mesos de seguiment).



Fig. 9: Dashboard *Monitoratge en Temps Real — TFG UAB* per al pacient `1107_Loop` (home, 12 anys, estudi Loop, 18 mesos de seguiment). Elements visibles: selector de pacient i perfil demogràfic; indicadors TIR (79,4%), TBR L1 (2,18%), TBR L2 (0,135%), TAR L1 (16,4%), TAR L2 (1,86%); CV (33,9) i GMI (6,54); temps amb sensor actiu; distribució de categories; taula de triatge ordenada per prioritat; comptadors d'episodis d'hipoglucèmia (18) i d'hiperglucèmia (43) amb durada mitjana de 28,6 min; evolució temporal de la glucosa a 6 mesos amb zones de color; i pendent de variació per pacient.

Els valors del TIR (79,4%) i del GMI (6,54) indiquen un control glucèmic acceptable, per sobre del llindar clínic recomanat de TIR $\geq$ 70 %, consistent amb un perfil AID. El CV del 33,9 es troba per sota del valor de tall de risc (<36%). La taula de triatge mostra els pacients ordenats per urgència, amb els episodis crítics destacats. Les gràfiques inferiors confirmen que el pipeline processa correctament els 6 mesos de dades estàtiques i que el càlcul del pendent per finestra de 5 minuts és estable al llarg de tot el període.

B.3 Taula de triatge de pacients

La Fig. 10 mostra la taula de triatge en funcionament amb 200 pacients de la cohort, ordenats per prioritat clínica descendent.

La taula evidencia la heterogeneïtat del control glucèmic

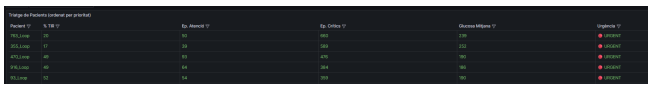


Fig. 10: Taula de triatge ordenada per prioritat. Columnes: identificador del pacient, TIR, nombre d'episodis d'atenció i d'episodis crítics, glucosa mitjana (mg/dL) i nivell d'urgència.

dins la cohort: els pacients `763_Loop` i `355_Loop` presenten un TIR del 20% i 17% respectivament, amb més de 580 episodis crítics, cosa que els situa en prioritat màxima. En contrast, pacients com `540_Loop` (TIR 30%) o `779_Loop` (TIR 50%) mostren un control deficient però amb menor càrrega d'episodis aguts. El criteri d'ordenació combina el TIR amb el nombre d'episodis crítics, de manera que el professional pot identificar immediatament els casos prioritaris sense revisar manualment cada perfil.

B.4 Dashboard de predicció glucèmica

La Fig. 11 mostra el dashboard de predicció en funcionament per al pacient `1087` de la cohort analítica. El dashboard consulta el *measurement* `glucose_prediction` d'InfluxDB per a les prediccions de LightGBM i el *measurement* `cgm_reading` per a la glucosa real, i combina ambdues fonts en una mateixa visualització.



Fig. 11: Dashboard *Predicció Glucèmica a 30 minuts (LightGBM)* per al pacient 1087. Elements visibles: indicadors puntuals RMSE (+30 min), R^2, Clarke A, MCC hipoglucèmia i MCC hiperglucèmia; gràfic de la predicció de glucosa a +30 minuts (línia verda); gràfic de la glucosa real (línia groga) amb zones de color per rang clínic; superposició de les dues sèries per comparació visual directa; i taula de tots els pacients amb mètriques individuals (RMSE, MAE, R^2, Clarke A%, Clarke E%, MCC_{hypo} i TG).

La fila superior de panells *stat* mostra les mètriques individuals del model per al pacient seleccionat, consultades des de PostgreSQL: el RMSE de 25,50 mg/dL (en groc, per sobre del llindar de 18 mg/dL, coherent amb la variabilitat alta d'aquest pacient), l'R^2 de 0,94 (en verd) i el Clarke A del 87,0% confirmen una predicció clínicament acceptable. Els valors MCC_{hypo} (0,793) i $\text{MCC}_{\text{hyper}}$ (0,858) indiquen una bona capacitat de detecció d'episodis de risc per a aquest pacient concret.

Els dos gràfics temporals centrals mostren, per separat, la predicció a +30 minuts i la glucosa real, facilitant la inspecció independent de cadascuna. El gràfic de superposició inferior permet comparar-les directament i identificar visualment l'anticipació efectiva del model: la

predicció (línia verda) precedeix la trajectòria real (línia groga) en aproximadament 30 minuts. El fons vermell indica que el pacient es manté en rang de hiperglucèmia moderada (181–250 mg/dL) durant la major part de la finestra mostrada, amb un descens cap al rang estable al final del període.

La taula inferior recull les mètriques de tots els pacients de la cohort analítica i permet al professional comparar la fiabilitat del model entre individus amb sensors i perfils glucèmics diferents. Això facilita que el clínic valori si la predicció d'un pacient concret és tan fiable com la d'altres de la cohort abans de prendre una decisió terapèutica basada en ella.

C Annex C: Infraestructura Docker Compose

C.1 Pipeline de preprocessament

El preprocessament s'implementa amb DuckDB, que executa consultes SQL directament sobre el fitxer Parquet de 154,6 milions de registres sense carregar-lo a memòria.

C.1.1 Criteris de selecció i fitxers de sortida

S'escullen els $N = 200$ pacients més complets que superen quatre criteris mínims: 12 mesos de seguiment, 20.000 lectures CGM, 2.000 registres d'insulina i 100 esdeveniments de carbohidrats. Aquests criteris són independents dels de la cohort analítica de 20 pacients (Secció 5.2), tot i compartir el llindar de 20.000 lectures CGM, que en ambdós casos representa el mínim per garantir 12 mesos de seguiment continu.

El pipeline s'executa en cinc passos: (1) selecció dels N pacients per DuckDB; (2) exportació de la finestra de 12 mesos de lectures CGM no nul·les; (3) validació i detecció de duplicats per la clau `(id, source_file, date)`; (4) divisió en bloc estàtic (càrrega directa a InfluxDB) i bloc de replay (injectat via Kafka); i (5) normalització de timestamps a `NORM_START = 2026-01-01 UTC` i càrrega de metadades a PostgreSQL. La normalització és necessària perquè MetaboNet abasta més de 10 anys amb rangs temporals heterogenis per pacient; Grafana requereix un rang comú per mostrar tots els pacients simultàniament. Els valors clínics no es modifiquen en cap cas.

Els quatre fitxers de sortida a `data/processed/` són:

TAULA 12: Fitxers de sortida del pipeline

Fitxer	Contingut	Rang
`12m_clean.parquet`	12 mesos, sense nuls CGM	Original
`static_6m.parquet`	Primers 6 mesos normalitzats	2025-07→2026-01
`replay_6m.parquet`	Últims 6 mesos normalitzats	2026-01→2026-07
`patients_200.csv`	Metadades demogràfiques	—

C.2 Producer i consumer: flux en temps real

El fitxer `src/producer.py` reprodueix el bloc de replay sobre el tòpic Kafka `glucose-readings` en mode de màxim throughput (`SPEED=0`), enviant les lectures sense delay per validar la capacitat màxima del pipeline. El paràmetre és configurable: `SPEED=300` simularia temps real (1 s real = 5 min simulats). Cada missatge conté els camps `patient_id`, `time`, `glucose_mgdl`, `data_source`, `ts_epoch` i, opcionalment, `insulin`, `carbs`, `basal` i `bolus`.

El fitxer `src/spark_consumer.py` consumeix els missatges amb Spark Structured Streaming, calcula la pendent de glucosa per finestra de 5 minuts i classifica cada pacient en una de **set categories** clíniques mitjançant dues UDFs (`categorize` i `assign_priority`):

TAULA 13: Categories clíniques i prioritat de triatge

Categoria	Criteri	Prioritat
`CRITICA_HIPO`	Glucosa < 54 mg/dL	P1
`HIPO`	54–69 mg/dL	P2
`ATENCIO_BAIXA`	70–90 mg/dL (proper al límit baix)	P2
`ESTABLE`	91–180 mg/dL (en rang objectiu)	P5
`ATENCIO_ALTA`	181–250 mg/dL + pendent $> 2{,}0$ mg/dL/min	P2
`BORDERLINE_ALTA`	181–250 mg/dL, pendent $\leq 2{,}0$ mg/dL/min	P3
`CRITICA_HIPER`	Glucosa > 250 mg/dL	P1
`UNKNOWN`	Lectura de glucosa absent	P4

Els resultats s'escriuen a InfluxDB al bucket `diabetes_data`. La categoria `ATENCIO_BAIXA` (70–90 mg/dL) és una subdivisió operativa del TIR estàndard que permet detectar tendències descendents abans que el pacient surti del rang segur; no altera la definició clínica del TIR (Secció 3.1).

C.3 Mètriques de rendiment del pipeline Spark

TAULA 14: Resum de mètriques Spark Structured Streaming (200 pacients simultanis)

Mètrica	Pic	Estable
Input Rate (registres/s)	6.000	3.000–4.000
Process Rate (registres/s)	6.000	3.000–4.000
Input Rows acumulats	$\sim$250M	
Batch Duration (ms)	60.000	$\leq$40.000
Operation Duration (ms)	70.000	$\sim$50.000
Global Watermark Gap (s)	300	250–300
State Memory (GB)	2	$\sim$1

El pic inicial de *Batch Duration* ($\sim$60 s) correspon a la càrrega d'estat; un cop estabilitzat, la latència se situa per sota dels 40 s. L'*Input Rate* sostingut de 3.000–6.000 reg/s confirma que el pipeline absorbeix el flux sense acumulació, i l'ús de memòria s'estabilitza a $\sim$1 GB.

C.4 Serveis i configuració

Tota la infraestructura es desplega amb un únic `docker-compose.yml`. El fitxer `src/ingest_static.py` carrega el bloc estàtic a InfluxDB en mode batch en arrencar el sistema, de manera que el dashboard mostra l'historial de 6 mesos immediatament sense esperar el replay.

El tòpic `glucose-readings` es crea automàticament amb 4 particions i factor de replicació 1. La xarxa interna `tfg-net` aïlla els contenidors de la màquina amfitriona i els volums persistents conserven les dades entre reinicis.

TAULA 15: SERVEIS DEL DOCKER-COMPOSE.YML

Servei	Port	Rol
zookeeper	2181	Coordinació Kafka
kafka	9092	Missatgeria streaming
spark-master	8080/7077	Motor de processament
spark-worker	8081	Worker Spark
influxdb	8086	BD sèries temporals
grafana	3000	Dashboard clínic
postgresql	5432	BD metadades

D ANNEX D: DETALLS EXPERIMENTALS DE LA CAPA ANALÍTICA

D.1 Hiperparàmetres i configuració dels models

S'han avaluat sis estratègies sota les mateixes condicions experimentals: mateix conjunt de dades, mateixa partició train/test, mateixa finestra de *features* i mateixa configuració `MultiOutputRegressor`. El 33,2% dels registres no tenen valor de carbohidrats (Secció 5.3): LightGBM gestiona aquests nuls nativament; per a Ridge, Random Forest i XGBoost[1] s'imputa `carbs` $= 0{,}0$, equivalent clínicament a "cap ingesta recent".

Zero Order Hold (ZOH). $\hat{y}_{t+k} = y_t$ per a tot k. No requereix entrenament i és competitiu a horitzons curts (5 min), quan la glucosa canvia poc.

Naive Linear. Extrapola linealment la tendència recent:

$$\hat{y}_{t+k} = y_t + k \cdot (y_t - y_{t-1}) \tag{1}$$

Falla en canvis bruscos de pendent, generant prediccions fisiològicament impossibles. El Temporal Gain negatiu ($-2{,}82$ min) confirma que és pitjor que no predir res.

Ridge. Regressió lineal amb regularització L_2:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 + \alpha \|\boldsymbol{\beta}\|^2 \tag{2}$$

La regularització és especialment útil per les 12 *features* de lag CGM altament correlacionades. α es selecciona per CV de 2 *folds* sobre $\{0{,}01;\ 0{,}1;\ 1{,}0;\ 10{,}0\}$ [14].

TAULA 16: HIPERPARÀMETRES DE RIDGE.

Paràmetre	Valor
`alpha`	CV $\in \{0{,}01;\ 0{,}1;\ 1{,}0;\ 10{,}0\}$
`cv_folds`	2
`scoring`	`neg_mean_squared_error`
`fit_intercept`	`True`

Random Forest. Mitjana de T arbres (*bagging*):

$$\hat{y} = \frac{1}{T} \sum_{t=1}^{T} h_t(\mathbf{x}) \tag{3}$$

Redueix la variància però no corregeix errors entre arbres, a diferència del gradient boosting. Submostreig de 50.000 mostres/pacient [14].

[1] XGBoost admet valors nuls nativament, però per mantenir la comparativa homogènia s'ha aplicat la mateixa imputació als tres models.

TAULA 17: HIPERPARÀMETRES DE RANDOM FOREST.

Paràmetre	Valor
`n_estimators`	200
`max_depth`	20
`min_samples_leaf`	3
`max_features`	$\sqrt{p}$
`random_state`	42
`n_jobs`	-1
`max_train_samples`	50.000

LightGBM. Gradient boosting seqüencial (*leaf-wise*):

$$F_m(\mathbf{x}) = F_{m-1}(\mathbf{x}) + \eta \cdot h_m(\mathbf{x}) \tag{4}$$

Utilitza GOSS (*Gradient-based One-Side Sampling*) i EFB (*Exclusive Feature Bundling*) per reduir el cost computacional. Gestiona nativament els nuls de `carbs` [15].

TAULA 18: HIPERPARÀMETRES DE LIGHTGBM.

Paràmetre	Valor
`n_estimators`	300
`learning_rate`	0,05
`max_depth`	8
`num_leaves`	63
`min_child_samples`	20
`subsample`	0,8
`colsample_bytree`	0,8
`reg_alpha` (L1)	0,1
`reg_lambda` (L2)	1,0
`random_state`	42
`max_train_samples`	50.000

XGBoost. Gradient boosting per nivells (*level-wise*), amb regularització L_1 i L_2 directament a la funció objectiu. Submostreig de 15.000 mostres/pacient [16].

TAULA 19: HIPERPARÀMETRES DE XGBOOST.

Paràmetre	Valor
`n_estimators`	200
`learning_rate`	0,05
`max_depth`	6
`subsample`	0,8
`colsample_bytree`	0,8
`reg_alpha` (L1)	0,1
`reg_lambda` (L2)	1,0
`random_state`	42
`n_jobs`	2
`max_train_samples`	30.000

D.2 Configuració MultiOutputRegressor

Tots els models s'han entrenat amb `MultiOutputRegressor`: 6 estimadors independents, un per a cada horitzó temporal (5, 10, 15, 20, 25 i 30 minuts). Aquesta configuració permet que cada estimador s'especialitzi en el seu horitzó, ja que la relació entre *features* i glucosa futura canvia qualitativament entre 5 minuts (dominada per la inèrcia glucèmica) i 30 minuts (dominada per la insulina activa i la tendència). Per a cada pacient s'entrenen $6 \times 4 = 24$ estimadors, i $24 \times 20 = 480$ en total.