

---

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

Armengol Romero, Marçal. *Models híbrids CNN-seqüencials per a predicció financera : comparació entre CNN-Transformer i CNN-LSTM amb anàlisi de característiques i interpretabilitat..* Treball de Final de Grau (Universitat Autònoma de Barcelona), 2026 (Enginyeria Informàtica)

---

This version is available at <https://ddd.uab.cat/record/334401>

under the terms of the  license.

# Models híbrids CNN-seqüencials per a predicció financera

Marçal Armengol Romero

**Resum**— Aquest article presenta un estudi comparatiu de dues arquitectures d'aprenentatge profund híbrides, CNN-LSTM i CNN-Transformer, aplicades a la predicció financera intradia amb intervals de cinc minuts en el mercat nord-americà. L'ús de capes convolucionals com a extractores de patrons locals, es combinen amb estructures seqüencials i mecanismes d'atenció per capturar dependències temporals complexes. Els hiperparàmetres s'han obtingut amb optimització bayesiana i poda anticipada dels intents amb menys potencial. El rendiment s'ha avaluat amb mètriques estadístiques clàssiques i variants orientades a la seguretat, com l'Accuracy\*, calculada sobre el subconjunt de prediccions amb més confiança del model. S'ha analitzat la lògica interna mitjançant Gradients Integrats. Els resultats demostren la superioritat estadística de l'arquitectura CNN-LSTM, així com una preferència de les xarxes per extreure característiques directament del preu brut en comptes d'indicadors tradicionals com l'RSI. Finalment, una simulació de backtesting aplicant les comissions d'un entorn realista valida l'obtenció de beneficis del model millor optimitzat, assolint un rendiment positiu de l'1,3% en condicions de mercat simulades.

**Paraules clau**—Aprenentatge profund, Arquitectures híbrides, Intel·ligència artificial explicable, Predicció financera.

**Abstract**—This paper presents a comparative study of two hybrid deep learning architectures, CNN-LSTM and CNN-Transformer, applied to intraday financial forecasting using 5-minute intervals in the US market. The use of convolutional layers as local pattern extractors is combined with sequential structures and attention mechanisms to capture complex temporal dependencies. Hyperparameters have been obtained using Bayesian optimization and early pruning of the least promising trials. Performance has been evaluated using classical and security-focused statistical metrics, such as Accuracy\*, computed on the subset of the model's most confident predictions. The internal logic of the models has been analyzed through Integrated Gradients. The results demonstrate the statistical superiority of the CNN-LSTM architecture, as well as a preference of the networks for extracting features directly from raw price data instead of traditional indicators like the RSI. Finally, a backtesting simulation applying commissions from a realistic environment validates the profitability of the best-optimized model, achieving a positive return of 1.3% under simulated market conditions.

**Index Terms**—Deep Learning, Explainable artificial intelligence, Financial Time Series Forecasting, Hybrid Architectures.



## 1 INTRODUCCIÓ

La predicció de tendències en el sector financer és un dels camps més complexos i dinàmics de l'enginyeria econòmica i la intel·ligència artificial. El modelatge del preu dels actius borsaris s'enfronta a desafiaments intrínsecs derivats de la naturalesa no lineal de les sèries temporals, l'elevada presència de soroll, la influència d'esdeveniments macroeconòmics inesperats i la interacció de múltiples factors externs que alteren la dinàmica de mercat. En aquest entorn altament volàtil, els enfocaments estadístics tradicionals sovint resulten insuficients per capturar patrons subtils i dependències a llarg termini de caràcter no lineal.

L'evolució de l'aprenentatge profund ha obert noves vies gràcies al desenvolupament d'arquitectures capaces d'extraure abstraccions complexes directament de les dades primàries. En aquest sentit, les Xarxes Neuronals

Convolucionals (CNN, de l'anglès Convolutional Neural Networks) [1], dissenyades originalment pel processament d'imatges, han demostrat una gran eficàcia en sèries temporals gràcies a la seva capacitat per detectar patrons locals i canvis estructurals a curt termini mitjançant filtres convolucionals. Tanmateix, per a una predicció financera robusta, cal combinar aquesta capacitat amb models seqüencials que comprenguin l'evolució del mercat al llarg del temps.

Aquest treball sorgeix de la motivació d'avaluar i comparar dues de les principals línies de recerca en aquest àmbit: la memòria recurrent, representada per les xarxes *Long Short-Term Memory (LSTM)* [2], i els mecanismes d'atenció global, utilitzant l'arquitectura del *Transformer* [3]. Les dues estratègies es plantegen de forma híbrida, usant un extractor local de paràmetres CNN com a part comuna de les dues arquitectures. Una motivació afegida del treball, és comprovar si els resultats obtinguts es podrien traduir en un benefici econòmic mitjançant un entorn de simulació de pràctica (*backtesting*) amb comissions de mercat reals.

Per dur a terme la comparació de les arquitectures,

- 
- E-mail de contacte: [m.armengol.2004@gmail.com](mailto:m.armengol.2004@gmail.com)
  - Menció realitzada: Computació
  - Treball tutoritzat per: Albert Romero Sanchez (Ciències de la Computació)
  - Curs 2025/26

aquest treball pren com a referència dos articles. D'una banda, l'arquitectura del CNN-Transformer i l'esquema de certes mètriques de comparació es basen directament en el treball de J.P. Morgan AI Research desenvolupat per Zeng et al. (2023) [4]. Aquest estudi destaca per utilitzar un bloc convolucional acoblat a un codificador Transformer basat en mecanismes d'atenció global. D'altra banda, per a la formulació de l'arquitectura recurrent CNN-LSTM, ens basem en el model proposat per Lu et al. (2020) [5]. Aquest aprofita l'ús de la memòria recurrent per processar de manera contínua la seqüència de característiques extretes prèviament per la capa convolucional.

Aquest article es planteja amb la intenció de contrastar ambdues metodologies en condicions equivalents, però introduint variacions significatives respecte als marcs conceptuals originals per motius metodològics i computacionals. Principalment, ens diferenciarem de l'enfocament de J.P. Morgan en la freqüència i l'horitzó de predicció. Mentre que en el paper de referència s'usen veles d'1 minut i s'executa una predicció de caràcter immediat, en la nostra investigació s'usen intervals de 5 minuts i un horitzó de predicció fix d'una hora. Aquest ajust metodològic és vital per atenuar el soroll d'alta freqüència i es justifica per les limitacions computacionals en el volum de dades gestionable en l'entorn de treball.

Així doncs, l'objectiu general d'aquest Treball de Final de Grau és dissenyar, implementar, optimitzar i comparar de manera rigorosa el rendiment de dos models híbrids de predicció financera (CNN-LSTM i CNN-Transformer) en un entorn operatiu intradia de cinc minuts prenent com a fonament i contrastant els resultats amb els papers de referència.

Per assolir aquesta fita, es plantegen els següents objectius específics:

- Processament i qualitat de dades. Es busca desenvolupar un flux de preprocessament cronològic estricte sobre sèries temporals d'alta freqüència de l'índex Dow Jones (any 2025), garantint la integritat de les dades i evitant qualsevol tipus de filtratge de dades.
- Desenvolupament d'arquitectures híbrides. Implementar ambdues estructures neuronals adaptant-les a les restriccions del projecte per avaluar l'eficàcia de la memòria recurrent davant els mecanismes d'atenció en condicions de mercat equivalents.
- Optimització Avançada. Executar una cerca automatitzada de configuracions òptimes per ponderar adequadament factors crítics com el nivell de regularització o la longitud de la seqüència d'entrada.
- Avaluació multicriteri i interpretabilitat. Analitzar exhaustivament la capacitat predictiva dels models més enllà de l'encert global mitjançant mètriques d'alta confiança i comprendre la lògica interna dels models mitjançant mètodes d'intel·ligència artificial explicable (gradients integrats).
- Validació pràctica financera. Desenvolupar una estratègia de backtesting que repliqui de manera

fidedigna les condicions operatives, incloent-hi comissions reals per mesurar-ne la rendibilitat neta final.

Pel que fa a la metodologia, aquesta s'ha articulada entorn d'un ecosistema tecnològic basat en Python i desenvolupat de forma modular mitjançant una sèrie de jupyter notebooks clarament documentats que garanteixen la neteja i la traçabilitat del codi. El procés s'inicia amb la ingesta automatitzada de les dades de cotització de l'índex Dow Jones a través de la API de Massive, les quals es manipulen mitjançant les llibreries Pandas i NumPy per, posteriorment, dissenyar i entrenar les xarxes híbrides sobre el framework d'aprenentatge profund PyTorch. Atesa l'alta demanda computacional de les capes LSTM i Transformer, la computació s'ha executat a la versió gratuïta de Google Colab, aprofitant l'acceleració per GPU NVIDIA T4 per reduir els temps de processament sense incórrer en costos d'infraestructura. Finalment, la cerca dels millors paràmetres s'ha automatitzat mitjançant el framework Optuna per optimitzar la configuració de les xarxes, mentre que la validació financera dels senyals resultants s'ha simulat sota condicions de mercat reals mitjançant la llibreria de backtesting Backtrader.

La resta d'aquest article s'estructura com es descriu a continuació. La Secció 2 descriu detalladament la naturalesa del conjunt de dades extret de l'API de Massive i les tècniques de preprocessament aplicades. La Secció 3 detalla el disseny estructural de les dues arquitectures híbrides implementades i aborda l'estratègia d'optimització d'hiperparàmetres mitjançant el framework Optuna. La Secció 4 explica l'estratègia per conèixer la importància de cada feature. La Secció 5 presenta els resultats experimentals obtinguts, dividits en rendiment estadístic i simulació financera de backtesting. Finalment, la Secció 6 conté les conclusions principals del treball i, finalment, en la Secció 7 s'exposen les línies de continuació futures.

## 2 PREPROCESSAMENT DE DADES

### 2.1 Adquisició i qualitat de dades

Les dades històriques d'alta freqüència s'han obtingut mitjançant l'API de la plataforma Massive [6]. El conjunt de dades seleccionat comprèn els registres històrics de les 10 empreses amb major capitalització i rellevància dins de l'índex Dow Jones durant l'any 2025. Per a cada actiu, s'han extret les variables clàssiques de mercat formatades en veles o barres de 5 minuts (OHLCV): preu d'obertura (*Open*), preu màxim (*High*), preu mínim (*Low*), preu de tancament (*Close*) i volum de transaccions (*Volume*).

Com a etapa inicial del preprocessament, es va dur a terme la normalització del fus horari de tot el conjunt de dades, fixant-lo de manera uniforme a l'hora local de Nova York (EST/EDT). Aquesta acció és indispensable per alinear estrictament la seqüència temporal amb l'activitat real del mercat on cotitzen els actius de l'estudi. En realitzar la descàrrega primària, es va comprovar que el volum inicial

de dades obtingudes divergia notablement entre les diferents empreses (veure Taula 1). Aquesta diferència en el nombre de dades, és degut a l'activitat en mercats secundaris o plataformes de negociació alternatives fora de les hores oficials. Determinades empreses d'alt perfil tecnològic i liquiditat global, com ara Amazon, registren un flux constant d'operacions en aquests períodes, estenent el seu historial diari en comparació amb actius de caràcter més tradicional.

TAULA 1  
Comparativa del volum de dades obtingudes

| Empresa                 | Files totals | Dins Mercat | Fora Mercat | % Oficial |
|-------------------------|--------------|-------------|-------------|-----------|
| Apple Inc.              | 46,140       | 19,644      | 26,496      | 42.6%     |
| Amazon.com Inc.         | 46,431       | 19,645      | 26,786      | 42.3%     |
| Johnson & Johnson       | 24,573       | 19,642      | 4,931       | 79.9%     |
| JPMorgan Chase & Co.    | 29,012       | 19,643      | 9,369       | 67.7%     |
| Visa Inc.               | 24,598       | 19,642      | 4,956       | 79.9%     |
| Walmart Inc.            | 34,880       | 19,642      | 15,238      | 56.3%     |
| Microsoft Corporation   | 41,818       | 19,644      | 22,174      | 47.0%     |
| The Coca-Cola Company   | 33,448       | 19,643      | 13,805      | 58.7%     |
| McDonald's Corporation  | 23,232       | 19,642      | 3,590       | 84.5%     |
| The Walt Disney Company | 29,722       | 19,643      | 10,079      | 66.1%     |

Per tal d'homogeneïtzar els intervals temporals i garantir una alineació cronològica exacta entre tots els actius, s'ha optat per restringir el conjunt de dades exclusivament a l'horari de la sessió ordinària de la borsa de Nova York (NYSE), establert de 09:30 a 16:00 (EST/EDT). Amb aquesta decisió es descarta el soroll i la baixa claredat dels mercats fora d'hores, estandarditzant la matriu d'entrada a un nombre fix de veles per jornada operada.

Prosseguint amb el control de qualitat de les dades, es va dissenyar un protocol per verificar la integritat estructural i física de la informació subministrada. Aquest procés va avaluar la presència de valors nuls, registres invàlids o buits temporals, així com la detecció d'incongruències lògiques en la formació de les espelmes, com ara anomalies on el preu màxim fos aritmèticament inferior al preu mínim. Així mateix, es van calcular les jornades d'obertura oficial sense dades i el recompte d'intervals de 5 minuts absents (veure Taula 2).

TAULA 2  
Resultats de les comprovacions fetes

| Comprovació                                | Quantitat |
|--------------------------------------------|-----------|
| Files amb NaN                              | 0         |
| Files amb ticker/datetime buit o invàlid   | 0         |
| OHLC: $high < max(open, close, low)$       | 0         |
| OHLC: $low > min(open, close, high)$       | 0         |
| Dies amb el mercat obert però sense valors | 10        |
| Files amb valors negatius                  | 0         |
| Intervals de 5 minuts faltants             | 20        |

Com es pot observar en la taula, les dades subministrades per *Massive* manifesten una gran robustesa estructural

i no presenten cap valor invàlid ni incongruències lògiques en les variables OHLCV. Les úniques llacunes d'informació detectades es restringeixen a l'absència de 10 dies complets de cotització, corresponents de manera uniforme al dia 9 de gener de 2025 per a les 10 empreses analitzades, així com a un conjunt marginal de 20 intervals de cinc minuts repartits de forma asíncrona entre les diferents companyies.

Pel que fa al tractament d'aquests valors absents, derivats de períodes puntuals sense activitat comercial o de talls menors en el flux del proveïdor de dades, s'ha aplicat de manera sistemàtica la tècnica d'ompliment cap endavant. Aquest mètode consisteix a assignar de forma passiva l'últim valor conegut a les observacions posteriors nul·les, garantint la continuïtat de la sèrie temporal sense introduir informació sintètica ni generar biaixos de visió de futur. Cal destacar que la incidència d'aquesta mancança de dades ha estat residual, afectant menys del 0,001% del volum total de dades, fet que consolida un conjunt d'entrenament altament net i estadísticament representatiu.

2.2 Estructura i partició

L'experimentació s'ha configurat sota un enfocament d'horitzó temporal fix per modelar el comportament del mercat a mitjà termini dins de la mateixa jornada operativa. La variable objectiu es defineix com una tasca de classificació binària amb un desplaçament operatiu fixat exactament en una hora vista (12 intervals de 5 minuts cap al futur). Addicionalment, per enriquir l'espai de característiques, es calcula l'Índex de Força Relativa (RSI) [7] segons la formulació clàssica de Wilder, i es va afegir una característica que permet indicar el pas del temps entre vela i vela a la xarxa neuronal. Aquest atribut permet que el model identifiqui possibles discontinuïtats temporals, com les produïdes pel tancament del mercat o pels caps de setmana, evitant que interpreti totes les observacions consecutives com si estiguessin separades pel mateix interval temporal.

A cada instant de temps (corresponent a una finestra de cotització de 5 minuts), el model rep un vector de característiques multidimensional que condensa la informació de mercat disponible de forma normalitzada (veure Taula 3).

TAULA 3  
Estructura del vector d'entrada

| Característica | Símbol       | Descripció                                                 |
|----------------|--------------|------------------------------------------------------------|
| Open           | $O_t$        | Preu d'obertura de l'actiu en l'interval $t$ .             |
| High           | $H_t$        | Preu màxim assolit durant l'interval $t$ .                 |
| Low            | $L_t$        | Preu mínim assolit durant l'interval $t$ .                 |
| Close          | $C_t$        | Preu de tancament de l'actiu en l'interval $t$ .           |
| Volume         | $V_t$        | Volum total de transaccions negociades en l'interval $t$ . |
| RSI            | $RSI_t$      | Relative Strength Index calculat per a l'instant $t$ .     |
| Delta Minutes  | $\Delta m_t$ | Diferencial de minuts                                      |

A diferència d'altres dades, les sèries temporals financeres presenten una forta dependència de l'ordre seqüencial. Per aquest motiu, s'ha descartat qualsevol tècnica de validació creuada aleatòria, aplicant en el seu lloc una

divisió cronològica estricta i lineal de l'històric de l'any 2025. S'ha usat un 60% de la sèrie inicial pel conjunt d'entrenament, un 15% pel conjunt de validació, i el 25% restant pel conjunt de test.

### 2.3 Normalització i prevenció de filtratge de dades

Per facilitar la convergència dels algorismes basats en gradients, les variables de tipus continu s'han de normalitzar uniformement. S'ha triat la tècnica d'escalat de mínims i màxims (*Min-Max Scaling*), que confina els valors en un rang numèric estrictament delimitat entre 0 i 1.

A més, un dels errors més comuns en el disseny de sistemes d'intel·ligència artificial aplicats a les finances és el filtratge de dades o *data leakage*, el qual infla de manera artificial les mètriques d'encert durant la fase de test extreient informació del futur.

Per prevenir el filtratge de dades, els valors mínims i màxims utilitzats en el procés de normalització s'obtenen exclusivament del conjunt d'entrenament. A més, la generació de matrius de seqüències s'ha programat per assegurar que cap finestra d'entrada del conjunt de test o validació contingui elements que se solapin amb conjunts previs.

## 3 ARQUITECTURES DELS MODELS

### 3.1 Disseny de les arquitectures

Ambdues arquitectures comparteixen una primera etapa basada en una xarxa neuronal convolucional (CNN), encarregada d'extreure característiques rellevants de la seqüència d'entrada. A partir d'aquest punt, els models difereixen en el mecanisme emprat per capturar les dependències temporals: el model CNN-LSTM utilitza una xarxa LSTM, mentre que el model CNN-Transformer es basa en mecanismes d'autoatenció. Ambdues arquitectures s'han implementat en Python mitjançant la llibreria PyTorch [8].

A partir d'aquest vector de característiques comú, els models es bifurquen en dues filosofies de processament de llarg termini:

1. Model CNN-LSTM [5]: Integra capes de memòria a curt i llarg termini (*Long Short-Term Memory*). L'arquitectura recurrent processa el vector de patrons convolucionals de manera contínua i seqüencial. Mitjançant les seves portes de control (obertura, oblit i sortida), la LSTM aconsegueix regular el flux d'informació a través del temps, sent ideal per capturar la dependència temporal i la direccionalitat contínua de les sèries financeres al llarg del canal de dades.
2. Model CNN-Transformer [4]: Substitueix la recurrència per un codificador. Aquest model prescindeix del processament seqüencial pas a pas; en el seu lloc, afegeix una codificació de posició a les característiques de la CNN i aplica mecanismes d'autoatenció multi-cap d'abast global. Això permet a la xarxa avaluar simultàniament la influència de tots els intervals temporals de la finestra, identificant correlacions distants en el temps.

### 3.2 Estratègia d'optimització d'hiperparàmetres

L'espai de cerca associat a les arquitectures

d'aprenentatge profund és altament complex i no lineal. Ateses les fortes limitacions computacionals degut a que s'han usat entorns limitats (la versió gratuïta de Google Colab), es van descartar les tècniques de cerca per força bruta com el *Grid Search*. En el seu lloc, s'ha implementat el framework Optuna [9]. Optuna té un enfocament bayesià que aprèn de l'obtingut en cada intent construint un model probabilístic de les mètriques obtingudes per dirigir de forma intel·ligent la cerca cap a les regions més prometedores de l'espai paramètric. Es va buscar optimitzar la mètrica F1-Score en el conjunt de validació.

Per fer viable aquest procés i mitigar el cost en temps de computació, s'ha configurat una estratègia de podat anticipat. Mitjançant la poda de cerca d'Optuna, el sistema monitoritza la corba d'aprenentatge de cada experiment època a època; si el rendiment d'un intent se situa per sota de la mitjana de l'històric en els mateixos punts de control, l'entrenament es cancel·la immediatament. Aquest mecanisme ha permès avaluar un total de 40 intents complets per arquitectura, limitant la durada màxima a 20 èpoques per intent si no es patia un podat previ. Utilitzant aquesta estratègia, es van poder obtenir els hiperparàmetres que maximitzaven el rendiment per cada arquitectura (veure Taula 4).

TAULA 4

Hiperparàmetres maximitzadors per arquitectura

| Paràmetre          | CNN-LSTM  | CNN-Transformer |
|--------------------|-----------|-----------------|
| Longitud Seqüència | 12        | 8               |
| Batch Size         | 64        | 32              |
| Learning Rate      | 9.783e-05 | 0.0003099       |
| Dropout            | 0.1118    | 0.4901          |
| d_model            | 128       | 64              |
| Num. Heads         | —         | 8               |
| Hidden size        | 256       | —               |
| Num Layers         | 3         | 1               |
| LSTM dropout       | 0.3826    | —               |
| Weight Decay       | 8.609e-06 | 8.244e-05       |
| optimizer          | AdamW     | Adam            |

L'anàlisi de les dades obtingudes en la taula 4 revela un comportament altament diferenciat entre les dues arquitectures d'aprenentatge profund. En primer lloc, el model CNN-LSTM tendeix a demanar finestres temporals d'entrada més llargues (12 intervals, equivalents a 1 hora de mercat) per poder estructurar correctament la seva memòria interna i estabilitzar les cadenes de recurrència.

En contraposició, el model CNN-Transformer convergeix millor utilitzant una longitud de seqüència més curta (8 intervals, equivalents a 40 minuts d'històric). Aquesta diferència s'explica perquè el mecanisme d'atenció, com que no depen d'un processament seqüencial pas a pas, acusa més fàcilment el soroll si la finestra s'estén en excés en intervals d'alta freqüència. Així mateix, s'observa que el Transformer requereix mecanismes de control del sobreajustament (overfitting) molt més severs, forçant un dropout elevat (0.49), una taxa d'aprenentatge

més conservadora més, fets que constaten la tendència d'aquesta arquitectura a memoritzar el soroll de fons en sèries de caràcter financer. També pot ser degut al fet que s'està usant un nombre baix de dades pel que sol ser usual per aquests tipus de models que necessiten una quantitat molt elevada de dades.

#### 4 INTERPRETABILITAT

Més enllà de l'avaluació de la capacitat predictiva d'un model mitjançant mètriques d'encert estadístic, es vol comprendre al detall quines característiques de l'entrada han aportat més a la predicció feta pels models i com es diferencien. Comprendre quins factors motiven una predicció permet mitigar l'efecte de "caixa negra" i validar si les xarxes exploten patrons econòmics lògics o si, per contra, pateixen un sobreajustament basat en artefactes del soroll de fons. Amb aquest propòsit, en aquest treball s'ha implementat el mètode d'atribució axiomàtic de Gradients Integrats [10] mitjançant el framework d'intel·ligència artificial explicable Captum [11].

##### 4.1 Representació

Per indagar en la lògica interna dels models, es van generar representacions gràfiques en forma de mapes de calor a partir d'un lot d'avaluació extret del conjunt de test. En la Fig. 1 i la Fig. 2, es poden veure les matrius d'atribució resultants per a una mostra del conjunt de test de les arquitectures CNN-LSTM i CNN-Transformer respectivament.

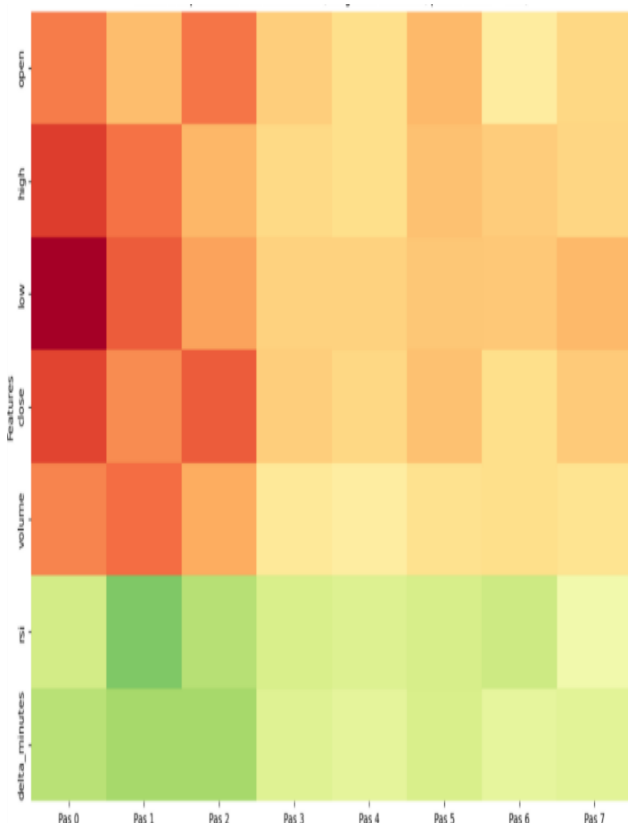


Fig. 1. Mapa de calor d'anàlisi d'importància de l'arquitectura CNN-LSTM

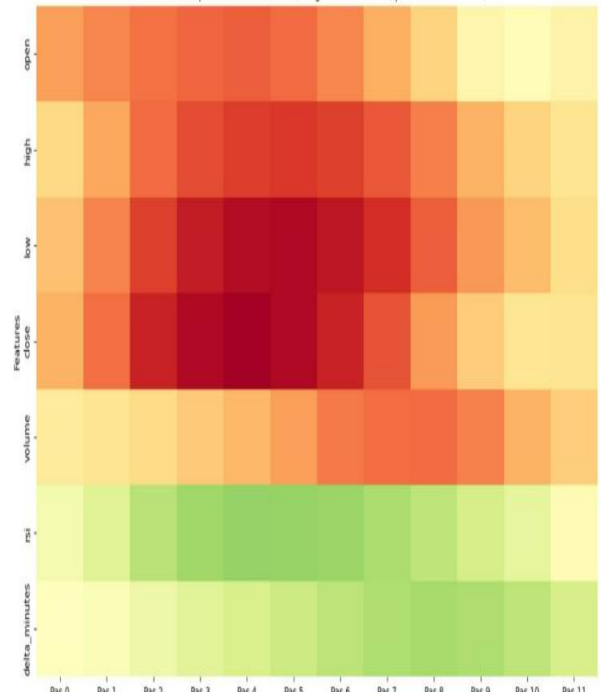


Fig 2 Mapa de calor d'anàlisi d'importància de l'arquitectura CNN-Transformer

L'eix vertical d'aquestes il·lustracions recull les variables d'entrada analitzades (els preus crus d'obertura, màxim, mínim, tancament, volum, així com l'RSI i la variable de control *delta minutes*), mentre que l'eix horitzontal mostra la progressió temporal. És un matís d'extrema rellevància conceptual assenyalar que l'eix temporal no està subdividit en les espelmes o barres temporals brutes de 5 minuts ingressades al sistema; en el seu lloc, l'eix horitzontal es desglossa en tokens convolucionals. Aquests tokens o vectors de característiques són el producte vectorial resultant de l'extracció de característiques duta a terme pel mòdul frontal CNN. Per tant, cada pas de temps en el mapa de calor representa una combinació comprimida i filtrada de la dinàmica local de preus abans de ser processada per l'etapa seqüencial (LSTM) o d'atenció (Transformer).

L'anàlisi visual d'aquests mapes permet desxifrar la direccionalitat de la influència a partir de la intensitat i la naturalesa cromàtica dels colors. En aquestes representacions, els tons més vius denoten un pes d'atribució o responsabilitat superior en el veredict final del model: el color verd identifica una contribució decisiva cap a una predicció de caràcter alcista (pujada del preu a una hora vista), mentre que el color vermell assenyalava aquells factors que la xarxa interpreta com a indicadors sòlids de caràcter baixista.

##### 4.2 Anàlisi de l'atribució

L'escrutini detallat dels mapes de calor revela divergències estructurals profundes en la manera com ambdues arquitectures processen i ponderen la informació al llarg del canal temporal i de l'espai de variables.

En primer lloc, en el CNN-Transformer, es pot veure com s'exhibeix un patró temporal molt marcat i asimètric. Tendeix a focalitzar la seva màxima intensitat d'atribució

en els tokens més llunyans en el temps, és a dir, en aquells situats a l'inici de la finestra temporal de la seqüència. Aquest fet constata la naturalesa del mecanisme d'atenció el qual, gràcies als encastaments de posició, prioritza ancoratges històrics llunyans per contextualitzar el moviment futur de l'actiu. Des de l'òptica de les variables, les característiques amb major magnitud d'atribució són els valors de preu mínim (*Low*) i de tancament (*Close*). Això denota que l'arquitectura aprèn a buscar patrons de caràcter estructural i geomètric en el preu brut per sobre de les oscil·lacions d'alta freqüència. Per contra, l'indicador oscil·lador RSI i la variable delta\_minutes (que computa el temps transcorregut entre l'obertura de la sessió i l'instant actual) presenten coloracions molt més tènues i properes a la neutralitat, evidenciant una rellevància gairebé nul·la en el procés de decisió.

Pel que fa al CNN-LSTM, es veu un contrast homogeni amb el Transformer, el model recurrent presenta una distribució de la importància temporal molt més equilibrada, uniforme i distribuïda al llarg de tota la sèrie. La xarxa LSTM processa la informació com un flux seqüencial continu on cada transició aporta informació incremental de valor. Aquesta topologia obliga a la xarxa a mantenir un nivell d'atenció sostingut sobre la totalitat del flux d'entrada, sense dependre de punts d'ancoratge aïllats o llunyans.

Un dels resultats més significatius obtinguts en aquest anàlisi, i que es manifesta com un tret general compartit per ambdues arquitectures, és la devaluació sistemàtica dels indicadors tècnics precalculats com l'RSI. En tots els experiments, les variables basades en oscil·ladors matemàtics tradicionals van rebre de forma unànime les cotes d'atribució més baixes de l'estudi. Aquesta troballa revesteix un gran interès en l'àmbit de la intel·ligència artificial aplicada a les finances, ja que suggereix empíricament que les xarxes neuronals profundes híbrides posseeixen prou capacitat d'abstracció per extreure les seves pròpies característiques i indicadors optimitzats a partir del preu cru (OHLC), preferint aquests patrons matemàtics interns abans que dependre d'indicadors sintètics clàssics dissenyats per a l'anàlisi tècnica manual.

## 5 RESULTATS

### 5.1 Mètriques usades

Per mesurar l'eficàcia de les classificacions binàries realitzades per les arquitectures híbrides, s'ha implementat un ventall complet de mètriques basades en la matriu de confusió (Veritables Positius, Falsos Positius, Veritables Negatius i Falsos Negatius). Aquestes mètriques s'avaluen de forma agregada i es descriuen a continuació.

1. Exactitud (*Accuracy*): Mesura la proporció global de prediccions correctes (tant de pujades com de baixades) respecte al total de mostres avalades.
2. Precisió (*Precision*): Avalua la fiabilitat del model en emetre senyals de compra, computant el percentatge d'encerts sobre el total de prediccions positives formulades.
3. Sensibilitat (*Recall*): Determina la capacitat de la xarxa per detectar els moviments alcistes reals que succeeixen en el mercat.

4. Valor F1 (*F1-score*): S'utilitza la mitjana harmònica (*Macro F1*) entre la precisió i la sensibilitat. Aquesta mètrica és l'eix central de la funció objectiu d'Optuna, ja que penalitza els models que mostren un biaix extrem cap a una única classe.
5. Àrea Sota la Corba ROC (ROC AUC): Avalua la capacitat de discriminació de la xarxa a diferents llindars de probabilitat, mesurant la robustesa estadística del classificador independentment del punt de tall operatiu.

Més enllà d'aquests indicadors clàssics, aquest article incorpora una mètrica de seguretat d'alta rellevància conceptual extreta directament del paper de referència de J.P. Morgan AI Research. La que ells anomenen *Accuracy\**. Les xarxes neuronals profundes emeten una distribució de probabilitat continuada per a cada predicció (entre 0 i 1). En lloc d'avaluar la totalitat de la mostra, l'*Accuracy\** restringeix el càlcul de l'exactitud exclusivament sobre el 25% de les mostres que presenten els nivells de certesa o confiança més elevats per part del model (les probabilitats més properes als extrems 0 i 1). Des de l'òptica de l'enginyeria financera, aquesta mètrica és un indicador crític de viabilitat: permet simular un entorn on un algorisme d'arbitratge només executa operacions en el mercat real quan disposa d'una seguretat extrema en el patró detectat, ignorant els intervals de dubte o alta volatilitat.

### 5.2 Disseny del marc de simulació financera

Un model amb excel·lents mètriques estadístiques en fase de laboratori pot fracassar de manera sistemàtica en operar en un entorn dinàmic real. Per esmenar aquesta bretxa entre la teoria i la pràctica, s'ha considerat oportú dissenyar una estratègia de simulació financera (*backtesting*) operatiu. Aquesta simulació s'executarà utilitzant exclusivament l'arquitectura híbrida que presenti el millor comportament global d'acord amb les mètriques estadístiques descrites en l'apartat anterior.

La infraestructura de la simulació s'ha programat mitjançant el framework Backtrader [12], replicant de manera fidedigna la lògica d'execució d'ordres sobre les barres de 5 minuts del conjunt de test de l'índex Dow Jones. Dins d'aquest disseny operatiu, la incorporació de les comissions de transacció esdevé el factor més determinant i crític de l'estudi. Moltes investigacions teòriques assumeixen mercats ideals lliures de fricció, un fet que infla artificialment els resultats en operatives intradia d'alta freqüència on el degoteig de compres i vendes és constant.

Per garantir un entorn realista i lliure de biaixos optimistes, s'han auditat les estructures tarifàries reals de la plataforma Interactive Brokers. A partir d'aquesta referència institucional, s'ha establert un esquema de comissions fix per operació dins de l'algorisme. Aquesta parametrització permetrà mesurar l'impacte net del cost de transacció sobre la corba d'equitat (*equity curve*) i avaluar de forma transparent si el guany teòric derivat dels patrons neuronals és capaç d'obtenir beneficis.

### 5.3 Resultats

Es van comparar els indicadors de rendiment obtinguts per a les dues arquitectures híbrides sobre el conjunt de test, és a dir, els últims mesos de la borsa de l'any 2025 (veure Fig. 3).

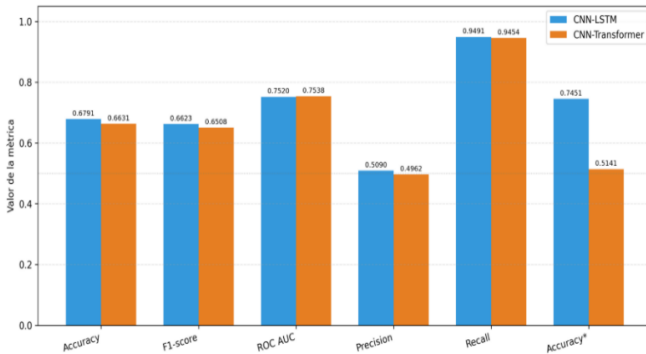


Fig. 3. Comparació de les mètriques obtingudes en el conjunt de test entre les arquitectures CNN-LSTM i la CNN-Transformer.

En primera instància, l'escrutini de l'Àrea Sota la Corba ROC (ROC AUC) revela que ambdues estructures neuronals assolixen cotes fortament equivalents i notablement elevades, situant-se entorn de la franja del  $\sim 0,75$ . Des de l'òptica de la teoria de la informació, aquestes magnituds confirmen de manera robusta que les dues arquitectures híbrides disposen d'una capacitat intrínseca de separació de classes molt superior a l'atzar, aconseguint descodificar patrons no lineals consistents dins de la sèrie d'alta freqüència de l'índex Dow Jones.

Tanmateix, quan es desglossa aquest comportament global mitjançant l'anàlisi creuada de la precisió i la sensibilitat, aflora un fenomen asimètric d'especialització extrema en ambdós sistemes. Tant el model CNN-LSTM com el CNN-Transformer han convergit cap a una detecció gairebé absoluta dels vectors pertanyents a la classe de baixada de mercat, registrant cotes de sensibilitat que sobrepassen el 94%. Aquest comportament es manifesta a costa d'una degradació severa en la precisió de la classe alcista, la qual oscil·la en valors propers al 50%. Aquest pronunciat desequilibri indica que els models actuen amb una sensibilitat asimètrica, generant un volum massiu de falsos positius en la classe de pujada. En termes pràctics i de mercat, l'entrenament ha induït un biaix operatiu: els models són altament eficaços identificant vectors de correcció o caigudes del preu, però exhibeixen dificultats estructurals per discriminar amb certesa les finestres d'ascens net.

Aquesta asimetria s'accentua de forma crítica en sotmetre els models a l'avaluació de la mètrica Accuracy\* (la taxa d'encert confinat exclusivament al 25% de les mostres amb un valor de probabilitat més proper als extrems). En aquesta instància, l'arquitectura CNN-LSTM assolix un valor de 0,7451. Tot i que a nivell nominal representa una xifra excel·lent, suggerint que el model és capaç d'encertar el 74,5% de les seves operacions quan assolix llindars de màxima certesa, quan s'avaluen de manera interna quins són els valors reals que s'estàn predint, es revela que aquest indicador es troba altament condicionat pel biaix direccional abans esmentat. La immensa majoria de les decisions encastades dins d'aquest percentil de màxima confiança corresponen de manera monòtona a ordres de caràcter

baixista.

En clar contrast, el model CNN-Transformer registra un col·lapse en aquesta mètrica de seguretat, obtenint un 0,5141. Aquest resultat evidencia la incapacitat de l'arquitectura basada en atenció global per filtrar amb èxit les operacions de confiança alta en sèries temporals financeres sorolloses, reduint la seva probabilitat d'encert real pràcticament al nivell de la línia base de l'atzar.

Com a síntesi d'aquest bloc teòric, es conclou que el model CNN-LSTM manifesta una clara superioritat en termes de mètriques estadístiques i capacitat de filtratge condicional respecte a l'alternativa Transformer. Tot i les limitacions derivades de l'asimetria del mercat, l'acoblament de capes convolucionals amb la memòria recurrent ha demostrat una aptitud superior per estructurar les dependències temporals del Dow Jones de l'any 2025.

Havent demostrat la seva superioritat en el pla estadístic i en l'Accuracy\*, el model CNN-LSTM va ser l'arquitectura seleccionada per governar l'estratègia operativa en la simulació de mercat. El *backtesting* es va configurar amb un capital inicial de 100.000 \$ per a cada actiu, representant un fons total simulat d'1.000.000 \$ distribuït uniformement entre les 10 empreses seleccionats de l'índex Dow Jones durant el període de test de l'any 2025.

L'estratègia de negociació implementada segueix una lògica d'execució directa i rigorosa: el sistema genera una ordre de compra d'accions d'una empresa en l'instant en què el mòdul neuronal prediu una tendència alcista per a l'horitzó d'una hora vista. Posteriorment, un cop transcorregut aquest interval temporal, les posicions es liquiden mitjançant una ordre de venda. Per garantir la consistència del risc, la política d'operacions prohibeix estrictament el solapament de compres; és a dir, no es permet obrir una nova posició sobre un actiu si aquest ja disposa d'una ordre de compra executada i activa en el mercat.

Els resultats nets de la simulació operativa, un cop deduïdes les estructures de comissions reals obtingudes d'Interactive Brokers, revelen que l'estratègia va cloure l'exercici amb un rendiment positiu de l'1,3% en 3 mesos. Si bé l'obtenció d'un benefici net positiu valida la capacitat del model per batre la fricció del mercat en determinades condicions, es va decidir comparar aquesta magnitud amb la mètrica de referència passiva de comprar i mantenir les accions. Durant el mateix interval temporal del 2025, el mercat subjacent va experimentar una revalorització significativament superior a l'1,3%, del 3,9% (veure Figura 4).

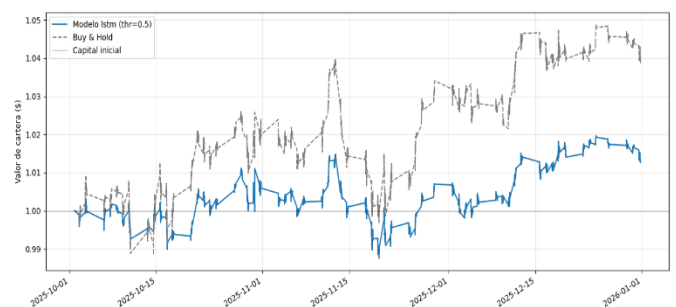


Fig. 4. Comparació del rendiment agregat de les 10 accions del *backtesting* entre l'estratègia de comprar i mantenir i usant les prediccions de l'arquitectura CNN-LSTM.

En conseqüència, es pot determinar que, des d'una perspectiva de gestió patrimonial, l'estratègia algorísmica proposada presenta una ràtio de rendibilitat-risc desfavorable: sotmetre un capital elevat a l'estrès i a l'exposició d'una operativa intradia d'alta freqüència per obtenir un marge tan estret no es tradueix en una inversió financerament eficient.

Aquest limitat rendiment pràctic troba la seva explicació en tres factors estructurals crítics.

En primer lloc, veiem que hi ha un gran impacte de les comissions de transacció: En una operativa d'alta freqüència amb intervals de 5 minuts, el nombre d'ordres llançades al mercat és massiu. L'aplicació de l'estructura tarifària de corretatge real actua com un fre sever sobre la corba d'equitat. Es va constatar que la viabilitat del sistema depèn absolutament de l'estructura de comissions; la utilització d'esquemes de comissions fixes (en lloc de variables o percentuals indexades per volum) redueix ràpidament els guanys degut al fet que molts dels intercanvis operen amb marges de preu molt estrets, convertint operacions teòrica-ment guanyadores en pèrdues netes de capital.

En segon lloc, l'absència d'un llindar de tal. Seguint el marc metodològic del paper de referència de J.P. Morgan, la variable objectiu d'aquest treball es va dissenyar com una classificació binària pura basada en el simple augment o disminució nominal del preu, sense establir un llindar mínim de variació percentual. En el context de la negociació intradia, aquesta decisió de disseny provoca que el model executi operacions davant de moviments alcistes microscòpics o marginals. Aquests guanys, en ser poc significatius, són incapaços de absorbir el cost imposat per la comissió de la plataforma de corretatge.

En tercer lloc, la gran diferència entre la precisió i la sensibilitat. El model, demostra una capacitat bona per predir quan hi haurà una baixada del mercat. En canvi, no és tant bo per pronosticar-ne les pujades. Per la nostra estratègia, en la que només podem aconseguir un benefici econòmic amb les pujades de les accions, aquest biaix en les prediccions ens suposa una limitació als possibles guanys.

En conclusió, la transició del laboratori al mercat mitjançant el backtesting confirma que la predictibilitat estadística del model CNN-LSTM és real, però el seu èxit econòmic en el món pràctic requereix un refinament en el disseny de l'estratègia. Per millorar la ràtio de rendibilitat, treballs futurs haurien d'incorporar un llindar de variació percentual mínima en l'etapa de classificació o restringir l'activació de les ordres, cosa que permet filtrar el soroll operatiu i amortitzar de manera eficient les comissions de transacció.

### 5.3 Comparativa amb l'Estat de l'art

Per tal de contextualitzar i validar de manera externa l'abast de les mètriques obtingudes, s'ha realitzat una comparació directa contra els resultats reportats per la divisió d'intel·ligència artificial de J.P. Morgan (veure Taula 5). En el seu treball de referència, els investigadors van proposar una arquitectura basada en un extractor convolucional acoblat a un bloc d'atenció (CNN-Transformer). Es contrasten els indicadors d'Accuracy i Accuracy\* obtinguts per les nostres dues configuracions enfront de la solució

oficial de J.P. Morgan.

TAULA 5  
Comparació d'arquitectures

| Arquitectura                                     | Exactitud | Accuracy* |
|--------------------------------------------------|-----------|-----------|
| CNN-LSTM<br>(Proposada en aquest treball)        | 67,90%    | 74,51%    |
| CNN-Transformer<br>(Proposada en aquest treball) | 66,30%    | 51,41%    |
| CNN-Transformer<br>(Proposada per J.P. Morgan)   | 56,70%    | 66,80%    |

L'anàlisi de la taula revela divergències de gran interès acadèmic. En primer lloc, s'observa que tant el nostre CNN-LSTM (67,90%) com el nostre CNN-Transformer (66,30%) assoleixen un nivell d'Accuracy global substancialment superior al registrat pel model de J.P. Morgan (56,70%). Aquest increment en l'exactitud base s'explica per una diferència estructural crítica en el disseny de l'experiment. L'estudi de J.P. Morgan es va executar utilitzant veles d'alta freqüència extrema d'1 minut per a una predicció immediata (a la següent espelma). Aquesta configuració sotmet el model a un volum massiu de soroll de microestructura de mercat i a fluctuacions aleatòries d'alta freqüència difícils de modelitzar. En contraposició, en aquest TFG s'ha optat per utilitzar veles de 5 minuts amb un horitzó de predicció d'1 hora vista. Aquest filtratge temporal i el major desplaçament de la variable objectiu actuen com un suavitzador natiu, reduint dràsticament el soroll inherent de la sèrie i permetent que les nostres xarxes identifiquin tendències estructurals més definides i netes.

En segon lloc, l'escrutini de la mètrica de l'Accuracy\* ens permet arribar a unes conclusions diferents. El model CNN-Transformer de J.P. Morgan demostra una estabilitat conceptual superior al nostre homòleg basat en la mateixa topologia; mentre que l'arquitectura de J.P. Morgan és capaç d'eleva la seva exactitud del 56,70% al 66,80% (+10,10%) en acotar el càlcul al percentil de màxima certesa, el nostre CNN-Transformer pateix un col·lapse sever, degradant el seu rendiment fins al 51,41%. Des de l'òptica de la ciència de dades, aquesta falta d'estabilitat en la nostra versió del Transformer respon, molt probablement, a la disparitat en el volum de dades d'entrenament. Aquests mecanismes, són coneguts per la seva baixa inducció de biaix estructural, la qual cosa significa que requereixen volums de dades massius (com els que s'han utilitzat en l'estudi de J.P. Morgan) per convergir correctament sense patir sobreajustament. Segurament, en disposar d'un volum de dades més contingut en el nostre cas, el mòdul d'atenció dissenyat és incapaç de calibrar de forma robusta les probabilitats de sortida en els extrems.

Finalment, la gran fita d'aquesta comparativa recau en el rendiment de la nostra arquitectura CNN-LSTM. Aquest model no només capitalitza la reducció de soroll de l'horitzó d'una hora per obtenir el millor Accuracy global (67,90%), sinó que demostra una resiliència i capacitat de filtratge excepcionals. Aquest fenomen confirma que,

davant de volums de dades mes reduïdes i usant predicció a més llarg termini, l'acoblament de memòries recurrents (LSTM) és molt més eficient i estable a l'hora de catalogar la jerarquia de la certesa que els models basats purament en atenció global.

## 6 CONCLUSIONS

Aquest estudi ha permès constatar amb rigor que, partint de dades d'alta freqüència degudament preprocessades del mercat nord-americà mitjançant l'índex Dow Jones, l'ús d'arquitectures híbrides que coordinen xarxes neuronals convolucionals amb mòduls de processament temporal profund constitueix un enfocament viable, sòlid i d'alta capacitat de modelització. L'acoblament frontal d'una capa CNN com a extractor de patrons locals ha demostrat ser una solució metodològica encertada per atenuar el soroll inherent de la freqüència intradia de cinc minuts. Aquesta hibridació s'ha manifestat altament capaç de desglossar la naturalesa no lineal i complexa de les sèries temporals financeres, captant dinàmiques volàtils i tendències subtils que sovint passen completament desapercebudes per als mètodes d'anàlisi quantitativa convencionals o de caràcter purament estadístic.

De l'escrutini comparatiu de laboratori executat entre ambdues aproximacions, es desprèn de manera inequívoca la superioritat de l'arquitectura CNN-LSTM per davant de l'alternativa CNN-Transformer en aquest context operatiu i sota els horitzons temporals fixats. L'estructura recurrent i seqüencial pròpia de la LSTM sembla adaptar-se amb molta més estabilitat a la naturalesa fluida i cronològica de les cadenes de cotització utilitzades, assolint mètriques de rendiment general superiors.

Així mateix, l'exploració del funcionament intern dels models mitjançant tècniques d'intel·ligència artificial explicable ha aportat una de les troballes més rellevants i disruptives d'aquesta investigació, revelant que ambdós models s'alineen de manera unànime a l'hora de prioritzar la matèria primera de la informació: els valors crus de les cotitzacions (OHLCV). Els mapes de calor de gradients integrats demostren de forma empírica que les xarxes profundes posseeixen prou potencial d'abstracció per formular les seves pròpies mètriques i senyals latents a partir dels preus bruts de mercat, devaluant sistemàticament el pes d'indicadors tècnics precalculats tradicionals com l'RSI. Aquesta conclusió obre un debat de gran interès en el camp de l'enginyeria financera, desmitificant la necessitat d'incorporar oscil·ladors clàssics en favor de l'autonomia de les xarxes profundes per dissenyar indicadors interns optimitzats.

Finalment, la translació d'aquests coneixements cap a la simulació de mercat real ens permet explicar quina utilitat en un camp real se'ls poden donar a aquests tipus d'arquitectures. L'obtenció d'un rendiment net de l'1,3% valida la capacitat predictiva real de la xarxa CNN-LSTM. No obstant això, en contrastar aquest guany amb el rendiment lliurat per l'estratègia passiva, es conclou que, per a aquest cas, el nivell de risc i l'esforç computacional associats a una operativa intradia binària pura no resulten econòmicament eficients. L'asimetria del model cap a la detecció de les tendències de caiguda i l'absència d'un llindar de tall

percentual en la variable objectiu adverteixen que la predictibilitat estadística requereix anar estrictament acompanyat d'un disseny d'estratègia financera sofisticat per tal de capitalitzar de forma neta i asseverada els avantatges obtinguts en el laboratori. També podria suggerir, que es necessita un volum més elevat de dades per poder extreure al màxim el valor d'aquests tipus d'arquitectures.

## 7 LÍNIES DE RECERCA FUTURA

A partir de les troballes i les limitacions identificades en aquesta investigació, es plantegen diverses línies de desenvolupament futur per elevar la viabilitat pràctica dels models. En primera instància, és imperatiu reformular la variable objectiu del sistema passant d'una classificació binària pura a una de caràcter condicional o multiclasse, on s'incorpori un llindar de variació percentual mínima del preu. Aquesta modificació conceptual garantirà que els models només emetin senyals operatius davant de moviments de mercat prou amplis com per absorbir l'impacte de les comissions reals de plataformes com Interactive Brokers, evitant la sobre operació en moments de consolidació lateral o micro-volatilitat. En aquest mateix sentit operatiu, es considera necessari el disseny d'una nova mètrica de seguretat derivada de l'*Accuracy*\* que esmeni el biaix direccional de confiança detectat. Aquesta nova mètrica hauria de segmentar el càlcul avaluant, de forma independent i aïllada, el 25% de les mostres amb major certesa quan el model prediu una pujada i, de manera paral·lela, el 25% de màxima confiança quan es prediu una baixada. D'aquesta manera s'evitaria que l'alta taxa d'encert o especialització de la xarxa en una sola de les direccions acabi emascarant el rendiment real global del sistema, assegurant una eina de filtratge completament neutral i lliure de distorsions estadístiques.

En segon lloc, atesa la gran capacitat demostrada pel model híbrid per detectar els vectors de correcció i caiguda del mercat (amb una sensibilitat superior al 94%), una línia de recerca d'alt valor financer consistiria a expandir la lògica operativa de Backtrader cap a una estratègia de caràcter bidireccional. Això inclouria la implementació d'operacions de venda en curt i l'ús d'instruments derivats per capitalitzar de forma neta el biaix baixista inherent que les xarxes han après a identificar en el Dow Jones.

Així mateix, s'estima altament recomanable dissenyar un experiment d'entrenament net que exclougui completament l'RSI, atès que l'anàlisi d'interpretabilitat mitjançant Gradients Integrats ha demostrat que aquesta mètrica precalculada aporta una rellevància gairebé nul·la al procés de decisió. Avaluar el rendiment de les xarxes utilitzant exclusivament els valors crus del preu i el volum (OHLCV) serviria per verificar si la reducció d'elements redundants optimitza el procés d'abstracció interna de la xarxa. No obstant això, la devaluació de l'RSI no ha d'implicar un abandonament absolut del camp de l'anàlisi tècnica quantitativa; donat que existeixen nombroses mètriques alternatives i oscil·ladors de diferent naturalesa matemàtica en la literatura financera, un treball futur d'interès consistiria a explorar sistemàticament si altres indicadors tradicionals sí que posseeixen la capacitat d'aportar informació

predictiva incremental i rellevant abans de ser processats per les capes convolucionals.

En paral·lel, es proposa ampliar substancialment el volum de dades utilitzat per a la fase d'entrenament, estenent l'històric recollit a través de múltiples cicles econòmics. Aquest increment d'escala obre la porta a analitzar de quina manera evoluciona el comportament de les dues arquitectures al llarg del temps, verificant si són capaces de consolidar un aprenentatge incremental robust que millori l'estabilitat de les seves mètriques. Disposar d'un conjunt de dades massiu esdevé un factor crític, especialment per al model CNN-Transformer, ja que oferiria la densitat d'informació requerida pels mecanismes d'autoatenció per estabilitzar les seves prediccions en condicions d'alta confiança i mitigar el sobreajustament que pateix en entorns més acotats.

Finalment, des de l'òptica de l'arquitectura de computació, es proposa testear l'impacte de la incorporació de dades alternatives de tipus exogen, com l'anàlisi de sentiment de notícies financeres en temps real, i explorar l'ús de mecanismes d'atenció localitzada en el model Transformer per avaluar si és capaç de mitigar el sobreajustament que ha patit en presència del soroll d'alta freqüència de la sèrie temporal.

## ÚS DE LA IA GENERATIVA

En el desenvolupament d'aquest treball, s'ha fet ús de la intel·ligència artificial generativa com a eina de suport i optimització de tasques. S'ha usat per funcions com el suport en la redacció i l'estil de l'informe, comparar diferents enfocaments tècnics per agilitzar la presa de decisions, configurar l'entorn d'execució i resoldre algunes tasques de programació puntuals. Tot i el suport d'aquestes eines, les decisions finals sobre el plantejament del projecte, la implementació dels models, l'anàlisi dels resultats i la interpretació de les dades, han estat preses manualment.

## AGRAÏMENTS

Vull expressar el meu més sincer agraïment al meu tutor, Albert Romero, per la seva guia constant, la seva paciència i el seu acompanyament al llarg de tot aquest procés. El seu suport expert en la part tècnica ha estat fonamental per superar els desafiaments d'aquest projecte, i els seus consells i dedicació m'han motivat en cada etapa del camí.

Així mateix, vull dedicar aquest treball a la meua família. El seu suport incondicional, la seva comprensió i els seus ànims constants, especialment en els moments de major exigència, han estat el pilar fonamental que m'ha permès tirar endavant i culminar amb èxit aquesta etapa acadèmica. Sóc una persona molt afortunada. Gràcies per creure sempre en mi.

## BIBLIOGRAFIA

- [1] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998, doi: 10.1109/5.726791.
- [2] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997, doi:

10.1162/neco.1997.9.8.1735.

- [3] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, vol. 30, 2017, pp. 5998–6008. [En línia]. Disponible a: <https://arxiv.org/abs/1706.03762>
- [4] Z. Zeng, R. Kaur, S. Siddagangappa, S. Rahimi, T. Balch, and M. Veloso, "Financial time series forecasting using CNN and Transformer," J.P. Morgan AI Research, New York, USA, arXiv:2304.04912v1 [cs.LG], 11 Apr. 2023. [En línia]. Disponible a: <https://arxiv.org/abs/2304.04912>
- [5] W. Lu, J. Li, Y. Li, A. Sun, and J. Wang, "A CNN-LSTM-based model to forecast stock prices," *Complexity*, vol. 2020, Article ID 6622927, 10 pages, Nov. 24, 2020, doi: 10.1155/2020/6622927. [En línia]. Disponible a: <https://doi.org/10.1155/2020/6622927>
- [6] Massive Interactive, Inc., "REST API Quickstart," Massive Documentation, 2024. [En línia]. Disponible a: <https://massive.com/docs/rest/quickstart>. [Accedit: 26-feb-2026]
- [7] J. W. Wilder Jr., *New Concepts in Technical Trading Systems*. Greensboro, NC, USA: Trend Research, 1978.
- [8] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimeshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, *PyTorch: An Imperative Style, High-Performance Deep Learning Library* [En línia]. Disponible a: <https://arxiv.org/pdf/1912.01703>
- [9] T. Akiba, S. Sano, T. Yanase, T. Ohta, i M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," 2024. [En línia]. Disponible a: <https://optuna.org/>. [Accedit: 1-maig-2026].
- [10] M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic attribution for deep networks," in *Proceedings of the 34th International Conference on Machine Learning*, vol. 70, 2017, pp. 3319–3328. [En línia]. Disponible a: <https://arxiv.org/abs/1703.01365>
- [11] Captum, "Introduction to Captum," 2024. [En línia]. Disponible a: <https://captum.ai/docs/introduction>. [Accedit: 3-maig-2026].
- [12] Backtrader, "Backtrader Documentation," 2024. [En línia]. Disponible a: <https://www.backtrader.com/docu/>. [Accedit: 25-abril-2026].

## APÈNDIX

### A1. EXPLICACIÓ DE LES ARQUITECTURES

En l'apèndix, s'explica d'una manera més detallada quin és el funcionament de les arquitectures. Per fer-ho, no es defineixen directament les arquitectures CNN-LSTM ni CNN-Transformer, s'explicarà cada component per poder entendre com funcionen un cop s'ajunten les dues arquitectures.

#### A1.1 Xarxes Neuronals Convolucionals (CNN)

Les xarxes neuronals convolucionals [1] constitueixen l'etapa frontal de processament en les dues arquitectures híbrides desenvolupades en aquest treball. Originalment dissenyades pel processament d'imatges (CNN 2D), la seva variant unidimensional (CNN 1D) es considera una eina eficient per a l'extracció de característiques locals en sèries temporals. En el context de l'alta freqüència financera, on les dades es presenten com una seqüència cronològica, la convolució opera fent lliscar un filtre o *kernel* al llarg de l'eix del temps. Aquest *kernel* aplica una operació matemàtica de multiplicació i suma ponderada sobre valors consecutius de la seqüència d'entrada, aconseguint capturar patrons geomètrics locals complexos, com ara ruptures de rangs, girs bruscos en el preu o repunts sobtats de la volatilitat que passarien desapercebuts en una anàlisi global.

Després de l'operació de convolució, les dades es processen mitjançant la funció d'activació GELU (*Gaussian Error Linear Unit*). A diferència de la clàssica funció ReLU, que anul·la a zero qualsevol valor negatiu i pot afavorir el problema de les anomenades neurones "mortes", la GELU permet que els valors lleugerament inferiors a zero continuïn contribuint de manera suau a la sortida. En un entorn altament sorollós i impredecible com la borsa intradia, aquesta transició suau pot ajudar el model a no descartar senyals subtils del mercat i a estabilitzar l'aprenentatge.

Directament com a resultat del processament convolucional combinat amb l'activació GELU, es transforma la matriu d'entrada en una seqüència de vectors de característiques, que anomenarem *tokens* per l'arquitectura CNN-Transformer, aquests alimentaran les capes posteriors del sistema.

Com es pot veure (Fig 5), de manera conceptual i simplificada, el mecanisme operatiu del bloc convolucional (CNN 1D) desenvolupat en aquest treball. En el diagrama es detalla l'estructura de la matriu d'entrada, la qual està integrada per les set característiques que s'han usat així com l'acció de la finestra lliscant. Per motius de claredat gràfica i síntesi visual, el *kernel* s'ha representat en la il·lustració amb una dimensió reduïda.

Específicament, l'esquema posa en relleu el procés de generació d'un vector de característiques individual: la regió emmarcada en color vermell delimita el camp receptiu actual del *kernel* on s'executen les operacions matemàtiques de convolució. Aquest operador desplaça de forma seqüencial al llarg de l'eix temporal i les operacions es realitzen exclusivament en cada característica de la matriu

d'entrada, fins a recórrer la seva totalitat, transformant així les dades brutes de la finestra en el conjunt complet de vectors de característiques densos requerits per a les següents etapes del sistema.

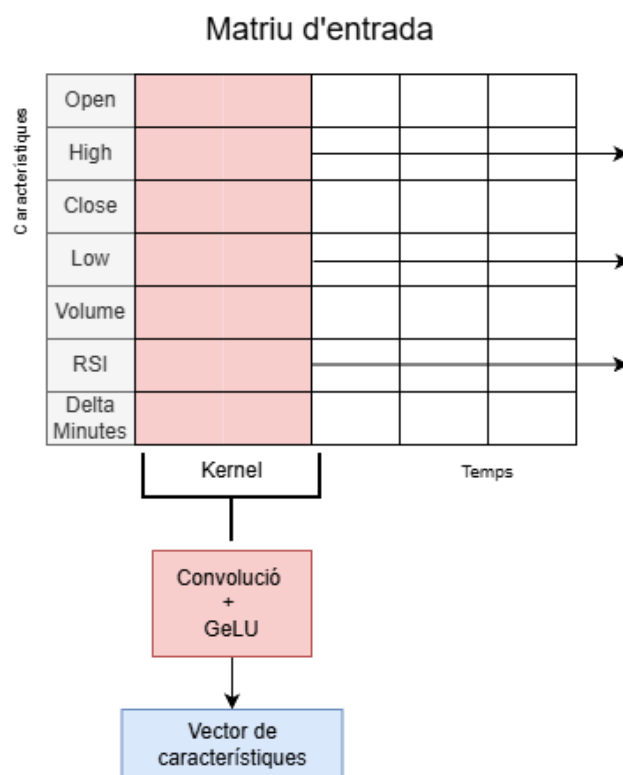


Fig. 5. Representació del funcionament del CNN i la generació de vectors de característiques.

#### A1.2 Xarxes de memòria a curt i llarg termini (LSTM)

Les xarxes de memòria a curt i llarg termini [2] (LSTM, per les seves sigles en anglès) són una variant especialitzada de les xarxes neuronals recurrents (RNN), dissenyades específicament per processar seqüències de dades i sèries temporals. Les RNN clàssiques pateixen un problema metodològic greu conegut com a desvaniment del gradient. A mesura que la finestra temporal es fa més llarga, la xarxa perd la capacitat de recordar o connectar esdeveniments passats amb el moment actual, fent que els senyals antics es dilueixin en el soroll. La LSTM resol aquesta limitació de manera elegant mitjançant una estructura interna capaç d'aprendre quina informació s'ha de retenir a llarg termini i quina s'ha de descartar per redundant.

El nucli conceptual de la LSTM és l'estat de la cel·la que es pot descriure com un flux d'informació que recorre linealment tota la seqüència temporal. Al llarg d'aquest flux, la xarxa afegeix o treu informació de manera molt controlada mitjançant tres comportes lògiques o portes, que s'activen de forma probabilística:

- Porta d'oblí: És el primer filtre de la cel·la. S'encarrega d'analitzar la informació que arriba del passat i decidir quin percentatge d'aquesta dada ja no és útil per al context actual de mercat i, per tant, s'ha d'oblidar per evitar acumulació de soroll.
- Porta d'entrada: Aquesta comporta avalua les

noves dades que entren en l'interval actual. Decideix quina part d'aquesta informació recent és prou rellevant per ser emmagatzemada i actualitzada en l'estat de la cel·la, modificant el flux d'informació.

- Porta de sortida: Finalment, aquesta porta determina quina part de la informació continguda en l'estat de la cel·la s'ha de manifestar cap a l'exterior. El resultat d'aquest filtre és l'estat ocult, que serveix tant com a predicció de la cel·la actual com a vector de memòria per a la següent iteració cronològica.

Gràcies a aquest mecanisme de control de flux (veure Fig 6), l'arquitectura LSTM esdevé una eina altament robusta per a l'anàlisi financera, ja que és capaç de mantenir un fil conductor net de les tendències subjacents del mercat al llarg de la finestra de visió, immunitzant el procés d'aprenentatge enfront de les alteracions aleatòries de curt termini.

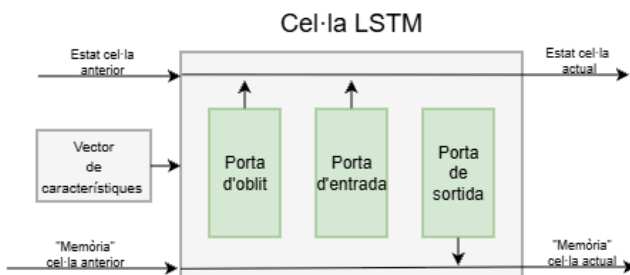


Fig. 6. Abstracció del funcionament d'una cel·la LSTM.

La figura presenta una extracció simplificada de l'estructura interna i del funcionament d'una cel·la LSTM, dissenyada per il·lustrar de manera clara com es gestiona el flux d'informació a través del temps per a la arquitectura.

El bloc operatiu s'alimenta de dues fonts d'entrada simultànies: d'una banda, els estats provinents de la cel·la anterior (tant l'estat de la cel·la com l'estat ocult o "memòria" a curt termini) i, de l'altra, el vector de característiques, que introdueix la informació nova i lliure de soroll que ha estat extreta prèviament per la xarxa convolucional (CNN). Un aspecte fonamental de la dinàmica interna de la LSTM és que aquest vector de característiques actualitzat i l'estat ocult anterior es distribueixen de manera paral·lela cap a totes i cadascuna de les tres portes lògiques (oblit, entrada i sortida). D'aquesta manera, cada comporta disposa de la visió completa del present i del passat immediat per decidir de forma autònoma quines són les dades que s'han d'esborrar, quines s'han de retenir i també les que formaran part de la sortida.

Finalment, un cop aplicades les respectives operacions matemàtiques de filtratge, els vectors resultants (l'estat de la cel·la actual i la "memòria" cel·la actual) s'envien directament cap a la següent cel·la de la seqüència cronològica. Allà, es repetirà exactament el mateix cicle d'aprenentatge, amb la particularitat que la nova cel·la rebrà un vector de característiques completament nou, corresponent a la següent finestra temporal de 5 minuts del mercat.

### A1.3 Transformers

L'arquitectura Transformer [3] representa un canvi de paradigma en la modelització de dades seqüencials i sèries temporals. A diferència de les xarxes recurrents tradicionals (com les LSTM), que avaluen la informació de manera lineal i interval a interval, el Transformer es basa en el principi del processament en paral·lel, analitzant tots els elements de la seqüència de manera simultània. Aquesta capacitat li permet capturar relacions i dependències de llarg abast amb una gran eficiència computacional.

Atès que aquest processament simultani fa que el model no incorpori de manera inherent la noció de l'ordre cronològic de la sèrie, és un requisit estructural imprescindible aplicar un mòdul de codificació posicional. Aquest mecanisme afageix configuracions posicionals, calculades mitjançant funcions sinusoidals de diferents freqüències a cada element de la seqüència d'entrada. D'aquesta manera, s'afegeix context d'ordre a la matriu de dades, assegurant que la xarxa reconegui quins elements precedeixen o succeeixen els altres al llarg de l'eix del temps.

El nucli operatiu del Transformer resideix en el mecanisme d'atenció amb varis caps (*Multi-head Attention*), el qual es fonamenta en el concepte d'autoatenció. Aquest procés es pot entendre analògicament com una cerca relacional, on cada element d'entrada genera tres vectors mitjançant matrius de pesos entrenables: una consulta, una clau i un valor. Mitjançant una operació de producte escalar entre les consultes i les claus de tots els elements, el model calcula de forma dinàmica una matriu de puntuacions que determina l'afinitat o rellevància estadística de cada part de la seqüència respecte a les altres. El format d'atenció amb varies parts, subdivideix aquest procés en subespais lineals paral·lels, permetent que el model aprengui simultàniament diferents tipus de relacions independents (com tendències macro o micro-volatilitats) en una mateixa iteració.

Finalment, cada bloc del codificador, o encoder del Transformer complementa el mòdul d'atenció amb una xarxa de connexió cap endavant, acompanyada de connexions residuals i capes de normalització per capes. Aquest disseny estructural interconnectat garanteix que la informació original no es degradi i optimitza el flux del gradient, facilitant un aprenentatge estable en arquitectures profundes.

### A1.4 Arquitectures híbrides

El processament de sèries temporals financeres altament sorolloses presenta un gran desafiament per a les arquitectures profundes no híbrides. Mentre que els models purament temporals (LSTM o Transformer) solen patir per convergir quan reben dades brutes a causa de la microestructura del mercat, les xarxes convolucionals aïllades manquen dels mecanismes de memòria necessaris per projectar tendències futures.

Per això, s'usen estratègies d'híbridació (veure Fig. 6) basada en dues etapes seqüencials ben diferenciades: la compressió espacial de patrons locals i la modelització de la dependència temporal macro.

Pel que fa a l'arquitectura CNN-LSTM, l'acoblament es realitza de manera lineal i cronològica. El flux de dades

s'estructura com es descriu a continuació.

En la primera etapa frontal, la matriu d'entrada, que conté les 7 característiques essencials del (OHLCV, RSI i delta\_minutes), és processada pel *kernel* convolucional. L'aplicació del filtre acoblat a l'activació GELU extreu les correlacions geomètriques immediates del mercat. La sortida d'aquesta etapa frontal es tradueix en una seqüència cronològica de vectors de característiques. Cada vector encapsula abstractament l'estat pur de la borsa en un instant determinat.

En la segona etapa, aquests tokens s'introdueixen de manera seqüencial interval per interval, en les cel·les de la xarxa LSTM. La LSTM rep cada vector de característiques com una representació més compacte i lliure de soroll, fet que li permet actualitzar el seu estat de cel·la i centrar-se exclusivament a registrar tendències subjacents d'alta fiabilitat.

Finalment, l'últim estat ocult de la LSTM es projecta cap a una capa lineal que emet la predicció binària final.

A diferència de l'aproximació recurrent, l'arquitectura CNN-Transformer aprofita el processament paral·lel del massiu substituint les cel·les recorrents per un bloc d'atenció global. L'arquitectura és similar a la comentada anteriorment. S'extreuen els tokens mitjançant el CNN, en lloc de processar-los d'un en un de manera lineal, aquests obtenen una codificació posicional perquè es disposi de la informació cronològica de l'ordre de les espelmes de 5 minuts abans d'entrar en el mòdul d'atenció. Després, el Transformer Encoder amb aquesta informació, avalua simultàniament com interaccionen tots els tokens generats per la CNN al llarg de tota la finestra temporal, calculant instantàniament quins esdeveniments passats tenen major rellevància sobre la direcció futura de l'índex Dow Jones.

Aquesta combinació permet que el sistema identifiqui relacions no lineals complexes distribuïdes de manera asíncrona en el temps, per bé que la manca d'un biaix inductiu estructural en el sistema d'atenció global fa que aquesta arquitectura requereixi corpus de dades massius per evitar caure en el sobreajustament sota el soroll de l'alta freqüència.

En la figura es representa el flux complet del processament, des de la introducció de les dades d'entrada fins a l'obtenció de la sortida final en forma de predicció binària (on el valor 1 indica la previsió d'un increment en el preu i el valor 0 reflecteix un moviment a la baixa). A partir del punt on es bifurca el corrent de dades, la secció codificada en color lila il·lustra la seqüència temporal de l'arquitectura CNN-LSTM, mentre que la branca de color verd esquematitza la ruta corresponent al model CNN-Transformer.

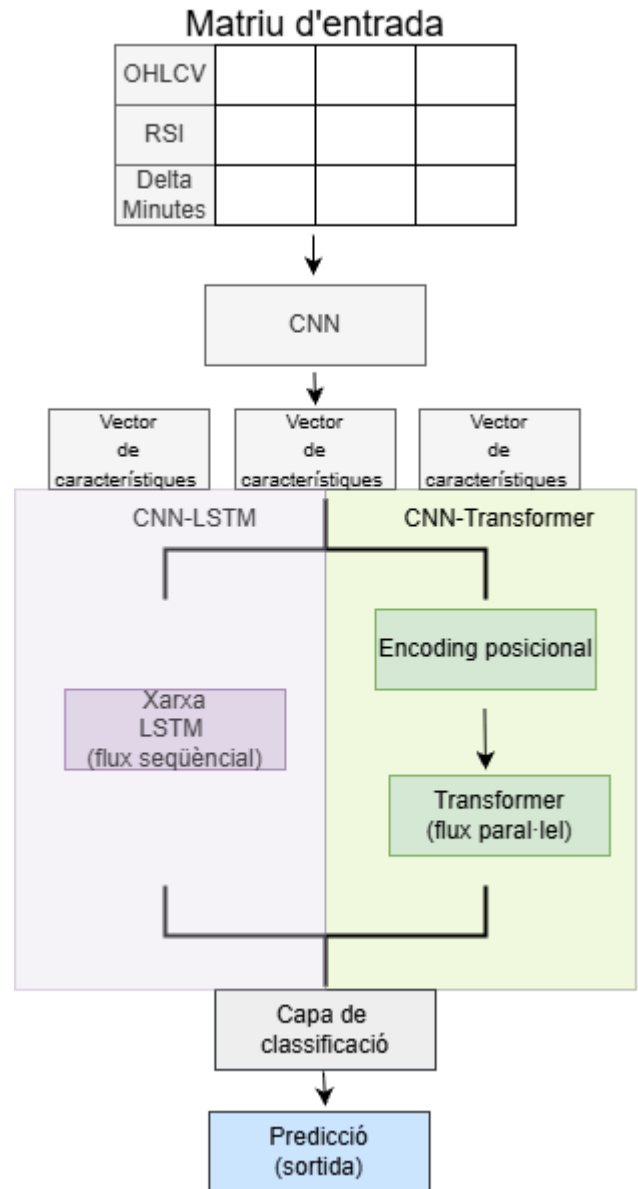


Fig. 7. Representació del flux de funcionament de les dues arquitectures.