

Detecció automàtica de defectes de soldadura mitjançant l'anàlisi de *visual transformers* i tècniques de *time series imaging*

Christian Velasco Gallego

Grau d'enginyeria informàtica

Àrea d'intel·ligència artificial

Nom Consultor/a: David Isern Alarcón

Nom Professor/a responsable de l'assignatura: *Friman Sánchez Castaño*

16 de juny de 2024



Aquesta obra està subjecta a una llicència de
Reconeixement-NoComercial [3.0 Espanya de
Creative Commons](https://creativecommons.org/licenses/by-nc/3.0/es/)

FITXA DEL TREBALL FINAL

Títol del treball:	Detecció automàtica de defectes de soldadura mitjançant l'anàlisi de <i>visual transformers</i> i tècniques de <i>time series imaging encoding</i>
Nom de l'autor:	<i>Christian Velasco Gallego</i>
Nom del consultor/a:	<i>David Isern Alarcón</i>
Nom del PRA:	<i>Friman Sánchez Castaño</i>
Data de lliurament (mm/aaaa):	06/2024
Titulació:	<i>Grau d'enginyeria informàtica</i>
Àrea del Treball Final:	<i>Intel·ligència artificial</i>
Idioma del treball:	Català
Nombre de crèdits:	12
Paraules clau	<i>Identificació de defectes; aprenentatge profund; codificació de sèries temporals en imatges</i>
Resum del Treball (màxim 250 paraules): <i>Amb la finalitat, context d'aplicació, metodologia, resultats i conclusions del treball.</i>	
<p>Nombrosos defectes, com ara danys al casc, fallades del motor i mal funcionament de l'equip, poden ocórrer en operacions marítimes, les quals poden afectar la seguretat i fiabilitat dels vaixells. En conseqüència, la detecció i classificació de defectes en el sector marítim és de vital importància per a garantir el seu correcte funcionament. Aquest estudi introdueix un nou enfocament de classificació de defectes mitjançant l'ús de tècniques de <i>time series imaging</i> amb <i>vision transformers</i>. En concret, s'analitzen en aquest estudi els mètodes de codificació de sèries temporals a imatges <i>Gramian Angular Field</i> (GAF), <i>Recurrence Plot</i> (RP) i <i>Markov Transition</i> (MT). També es presenta un cas d'estudi sobre la soldadura per arc metàl·lic per analitzar el rendiment de la metodologia proposada i avaluar la viabilitat d'implementar aquest tipus de mètodes per a la identificació de defectes. Els resultats indiquen que <i>Gramian Difference Angular Field</i> és el mètode que millor efectua la tasca proposada, obtenint una exactitud de més del 70%.</p>	

Abstract (in English, 250 words or less):

Numerous defects, such as hull damage, engine failures, and equipment malfunctions, can occur in maritime operations, which may affect the safety and reliability of ships. Accordingly, the detection and classification of defects in ships is of paramount importance to guarantee its adequate functioning. This paper introduces a new time series imaging approach for defect identification by combining distinct time series imaging approaches with a vision transformer. Specifically, the time series imaging methods Gramian Angular Field (GAF), Recurrence Plot (RP), and Markov Transition Field (MTF) are analysed in this study. A case study on metal arc welding is also presented to highlight the performance of the proposed methodology and assess the feasibility of implementing this type of methods for defect identification. The results of this case study indicate that *Gramian Difference Angular Field* is the most feasible encoding method for the task defined, as this method achieved an accuracy of over 70%.

Índex

1.	Resum	1
2.	Introducció	2
2.1.	Context i justificació del Treball	2
2.2.	Objectius del Treball	3
2.3.	Enfocament i mètode seguit	4
2.4.	Planificació del Treball	4
2.5.	Breu sumari de contribucions i productes obtinguts	5
2.6.	Breu descripció dels altres capítols de la memòria	6
3.	Estat de l'art	7
4.	Metodologia	10
4.1.	Preprocessament de les dades	10
4.1.1.	<i>Downsampling</i> de la sèrie temporal	11
4.1.2.	Divisió del conjunt de dades en set d'entrenament i set d'avaluació	11
4.1.3.	Normalització de la sèrie temporal	12
4.2.	Codificació de les dades en imatges	13
4.2.1.	<i>Gramian Angular Field</i> (GAF)	13
4.2.2.	<i>Markov Transition Field</i> (MTF)	14
4.2.3.	<i>Recurrence Plot</i> (RP)	14
4.2.4.	<i>Markov Transition Matrix</i> (MTM)	15
4.3.	Anàlisi de classificació amb <i>Visual Transformer</i>	15
4.3.1.	<i>Patch Embedding Layer</i> i <i>Positional Encoding</i>	16
4.3.2.	<i>Transformer Encoder</i>	16
4.3.3.	<i>Classification Head</i>	17
5.	Resultats	18
5.1.	Cas d'estudi	18
5.2.	Anàlisi exploratori de les dades	19
5.3.	Anàlisi dels resultats de la fase de preparació de les dades i de codificació de sèries temporals en imatges	20
5.4.	Avaluació del model <i>Vision Transformer</i> implementat per a la classificació de defectes en soldadures	23
5.5.	Estudi comparatiu	25
6.	Discussió	27
7.	Conclusions	28
7.1.	Conclusions	28
7.2.	Línies de futur	30
7.3.	Seguiment de la planificació	30
8.	Glossari	32
9.	Bibliografia	33
	Annexos	35

Llista de figures

Figura 1. Diagrama de Gantt del present TFG.	5
Figura 2. Representació gràfica de la metodologia proposada.	10
Figura 3. Representació gràfica d'un <i>Vision Transformer</i> .	16
Figura 4. Representació gràfica de dos exemples de senyals d'amperatge i voltatge associades a una mala qualitat de soldadura	20
Figura 5. Representació gràfica de dos exemples de senyals d'amperatge i voltatge associades a una bona qualitat de soldadura	20
Figura 6. Representació gràfica de l'abans i després d'aplicar <i>downsampling</i> a una senyal de voltatge i amperatge.	21
Figura 7. Representació gràfica dels resultats de codificar un exemple de senyals associades a mala qualitat i bona qualitat mitjançant l'ús de JRP	21
Figura 8. Representació gràfica de dos exemples de codificació de sèries temporals del paràmetre voltatge en imatge.	22
Figura 9. Representació gràfica de dos exemples de codificació de sèries temporals del paràmetre amperatge en imatge.	22
Figura 10. Representació gràfica dels resultats de codificar un exemple de senyals associades a mala qualitat i bona qualitat mitjançant l'ús de MTM.	23

Llista de taules

Taula 1. Planificació del treball.	5
Taula 2. Característiques principals del conjunt de dades utilitzat per a l'avaluació de la metodologia implementada.	18
Taula 3. Categorització de les senyals d'amperatge i voltatge en funció de la qualitat de la soldadura.	19
Taula 4. Configuració de l'arquitectura del ViT.	23
Taula 5. Configuració dels paràmetres d'optimització i entrenament.	24
Taula 6. Resultats principals dels diferents mètodes de codificació de sèries temporals a imatges en combinació amb ViT.	25
Taula 7. Resultats principals de l'estudi comparatiu.	26

1. Resum

Nombrosos defectes, com ara danys al casc, fallades del motor i mal funcionament de l'equip, poden ocórrer en operacions marítimes, les quals poden afectar la seguretat i fiabilitat dels vaixells. En conseqüència, la detecció i classificació de defectes en el sector marítim és de vital importància per a garantir el seu correcte funcionament. Aquest estudi introdueix un nou enfocament de classificació de defectes mitjançant l'ús de tècniques de *time series imaging* amb *vision transformers*. En concret, s'analitzen en aquest estudi els mètodes de codificació de sèries temporals a imatges *Gramian Angular Field* (GAF), *Recurrence Plot* (RP) i *Markov Transition*. També es presenta un cas d'estudi sobre la soldadura per arc metàl·lic per analitzar el rendiment de la metodologia proposada i avaluar la viabilitat d'implementar aquest tipus de mètodes per a la identificació de defectes. Els resultats indiquen que *Gramian Difference Angular Field* és el mètode que millor efectua la tasca proposada, obtenint una exactitud de més del 70%.

2. Introducció

En aquest capítol s'introdueix la situació actual del sector naval en relació amb la detecció de defectes de soldadura. Concretament, es fa èmfasis en la situació que aquest sector presenta en relació amb els recents avanços en intel·ligència artificial. En base a aquest context, es justifica la necessitat del present treball així com els objectius principals i específics d'aquest estudi. Segons la definició d'aquests objectius, s'estableix la planificació del projecte a efectuar i s'analitza l'enfocament a seguir i les contribucions i productes derivats de la implementació d'aquest treball.

2.1. Context i justificació del Treball

Les estructures dels cascos dels vaixells estan formades per milers de peces d'acer unides entre si mitjançant connexions soldades (Amirafshari et al., 2021). Arran les dimensions de les embarcacions i la complexitat del procés de soldadura en aquest entorn, es considera que el procés de soldadura és una de les tecnologies industrials més significatives en la construcció de vaixells, el qual representa aproximadament el 70% de la càrrega de treball durant el procés de construcció i el 40% dels costos totals (Liu et al., 2024).

Per això, cal garantir la qualitat de soldadura¹ per tal d'evitar possibles defectes que puguin alterar el correcte funcionament de les embarcacions en entorns operacionals. D'aquí que sorgeixi la necessitat de desenvolupar tècniques "més intel·ligents" per al correcte control de la qualitat de la soldadura a fi d'evitar possibles defectes durant el procés. En aquest aspecte, s'han realitzat diversos estudis on s'han considerat diferents mètodes d'aprenentatge computacional i aprenentatge profund per a detectar automàticament defectes de soldadura. Per exemple, Raoofi et al. (2024) va dur a terme una anàlisi de classificació amb l'objectiu d'identificar soldadures incompletes a través de la utilització d'un classificador basat en el detector YOLOv4² i anàlisi de vídeos. O bé Schwarz et al. (2022), el qual va implementar regressió lineal i el perceptró multicapa per a predir la resistència a la cisalla per tracció³.

Malgrat tots els estudis identificats i la importància que el procés de soldadura presenta en el sector transport, s'ha identificat que l'estat de l'art actual en relació amb l'aplicació de la intel·ligència artificial per a la detecció automàtica de defectes en certes indústries, com ara la naval, és limitat. Arran això, en aquest estudi s'introdueix una metodologia basada en la codificació de sèries temporals amb imatges i l'ús de *Vision Transformers*⁴ (ViT) per a detectar defectes de soldadura.

Segons la revisió de l'estat de l'art actual, el qual s'introdueix en el tercer capítol d'aquesta memòria, no s'ha pogut identificar cap estudi on s'analitzi l'ús de tècniques de *Time Series Imaging*⁵ (TSI) amb ViT per a la detecció de defectes de soldadura, encara que TSI ha demostrat en altres àrees de coneixement la

seva capacitat d'identificar patrons que no es poden observar en el format habitual en què es representen les sèries temporals (Fahim et al., 2021). A més, ViT ha demostrat tenir un rendiment superior a altres models més tradicionals, com ara les *Convolutional Neural Networks* (CNNs), en tasques de classificació d'imatges arran la seva capacitat per a capturar segments d'imatges no adjacents (Chen et al., 2024).

Arran tots els aspectes mencionats anteriorment, es considera que aquest estudi és necessari a fi d'analitzar altres maneres alternatives d'analitzar aquestes tasques de detecció de defectes per a comparar-les amb models més tradicionals que s'apliquen en l'actualitat.

2.2. Objectius del Treball

L'objectiu principal d'aquest treball és el de desenvolupar un *framework* mitjançant tècniques de *time series imaging* i aprenentatge profund per a la detecció de defectes de soldadura. A continuació, s'indiquen els objectius específics:

Objectius específics primaris

- Dur a terme una anàlisi crítica d'estudis existents relacionats amb la utilització de la intel·ligència artificial per a la identificació de defectes de soldadura dins del sector marítim.
- Investigar la importància d'aplicar la intel·ligència artificial per a la detecció de defectes de soldadura.
- Estudiar la viabilitat d'implementar les tècniques de *time series imaging* encoding *Gramian Angular Field*⁶ (GAF), *Recurrence Plots*⁷ (RP) i *Markov Transition Fields*⁸ (MTF) per a facilitar la implementació de *image classifiers*⁹.
- Analitzar el potencial d'un dels *image classifiers* desenvolupats més recentment, conegut com a *vision transformer*, en tasques de classificació.
- Avaluar l'exactitud de la metodologia implementada a través d'un cas d'estudi.
- Extreure conclusions a partir dels resultats obtinguts del cas d'estudi implementat.

Objectius específics secundaris

- Combinar tècniques de *time series imaging* amb *vision transformers* a fi d'analitzar si és plausible aquests tipus de mètodes per a classificar defectes de soldadura a partir de sèries temporals de paràmetres que faciliten la comprensió de la qualitat de la soldadura, com ara el voltatge i l'amperatge.
- Efectuar un estudi comparatiu entre la metodologia desenvolupada i altres xarxes neuronals convencionals que normalment s'utilitzarien per a tractar la problemàtica plantejada en aquest treball (*Long Short-Term Memory Neural Network*¹⁰ (LSTM), per exemple).

- Establir línies de treball futur per a millorar les actuals estratègies de detecció automàtica de defectes de soldadura.

2.3. Enfocament i mètode seguit

L'estratègia principal d'aquest estudi és adaptar i combinar metodologies ja existents amb l'objectiu de millorar el procés de detecció de defectes durant el procés de soldadura. S'ha decidit seleccionar aquesta estratègia arran la disponibilitat de *frameworks* que permeten la implementació de xarxes neuronals de manera simple i robusta, com ara *Tensorflow*, *Keras* i *Pytorch*. A més, també existeixen llibreries que permeten aplicar diferents mètodes de TSI per a la codificació de sèries temporals en imatges i per al tractament d'imatges, com ara *scikit-image* i *pyts*.

2.4. Planificació del Treball

Per al correcte desenvolupament del present treball, s'han definit un total de dotze tasques. Les primeres tres tasques fan referència a la definició del TFG, a la definició del context i justificació, i a la definició dels objectius a fi de determinar l'abast del projecte, la necessitat d'aquest i les contribucions principals. La quarta tasca es relaciona amb l'elaboració de l'estat de l'art a fi d'analitzar l'estat actual de l'àrea de coneixement a tractar i, així, justificar la necessitat d'aquest treball.

Les tasques 5-9 fan referència a la part principal d'aquest projecte, que és el de l'elaboració de la metodologia, la qual combina tècniques de codificació de sèries temporals en imatges amb *visual transformers* a fi d'avaluar si es millora l'exactitud de la tasca de classificació a implementar. Per a això, primerament es definirà la metodologia a desenvolupar (tasca 5). Un cop definida, es procedirà amb l'anàlisi de les dades a fi d'entendre les seves característiques així com identificar possibles reptes amb la preparació de les dades que s'hagin de tractar en la següent fase (tasca 6). D'acord amb els reptes relacionats amb les característiques de les dades, es procedeix amb la preparació de les dades per tal de garantir els següents tres aspectes: 1) les dades presenten la qualitat necessària, 2) les dades estan en el format necessari, i 3) les dades són les necessàries (tasca 7). A continuació, es procedeix amb la codificació de les dades en imatges. Per a això, s'analitzaran diferents tècniques de codificació d'imatges, com ara *Gramian Angular Field* i *Recurrence Plots* (tasca 8). Per finalitzar la metodologia, es desenvoluparà el *visual transformer* per a classificar els defectes en la tasca 9.

Un cop desenvolupada la metodologia necessària per a la correcta detecció automàtica de defectes en el procés de soldadura, cal avaluar la seva exactitud. Per a això, s'han definit les tasques 10 i 11. La tasca 10 fa referència al desenvolupament d'un estudi comparatiu a fi d'analitzar l'exactitud de la metodologia implementada amb altres xarxes neuronals més convencionals, com ara les *Convolutional Neural Networks*¹¹ i les *Long Short-Term Neural*

Networks. La tasca 11 fa referència a la utilització de mètriques i representacions gràfiques per a avaluar els resultats obtinguts.

Finalment, es defineix l'última tasca, la tasca 12, per a la redacció del TFG. Aquesta es duu a terme durant el transcurs de tot el TFG. La taula 1 presenta una descripció breu de les tasques planificades així com la respectiva data d'inici i dies de duració. En canvi, la figura 1 representa gràficament el diagrama de Gantt definit.

Taula 1. Planificació del treball.

Tasca	Descripció tasca	Inici	Dies
T1	Definició TFG	15/03/2024	10
T2	Definició context i justificació	15/03/2024	31
T3	Definició objectius	15/03/2024	31
T4	Elaboració de l'estat de l'art	01/04/2024	20
T5	Definició metodologia	17/04/2024	5
T6	Anàlisi de les dades	22/04/2024	15
T7	Preprocessament de les dades	01/05/2024	15
T8	Codificació de les dades a imatges	05/05/2024	20
T9	Desenvolupament del <i>visual transformer</i>	20/05/2024	7
T10	Desenvolupament estudi comparatiu	28/05/2024	7
T11	Avaluació dels resultats	04/06/2024	5
T12	Redacció del TFG	09/06/2024	83

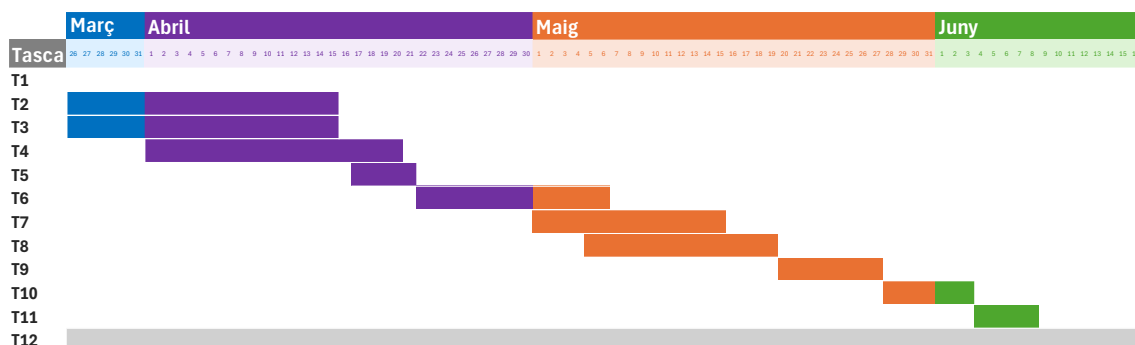


Fig. 1. Diagrama de Gantt del present TFG.

2.5. Breu sumari de contribucions i productes obtinguts

El producte principal obtingut serà una nova metodologia per a detectar defectes en el procés de soldadura de manera automàtica a través de la combinació de mètodes ja existents, els quals es basen en tècniques de codificació de sèries temporals en imatges i en models d'aprenentatge profund.

Les contribucions principals seran les següents:

- Identificar millores en les estratègies actuals de classificació de defectes de soldadura mitjançant la utilització de la intel·ligència artificial.
- Analitzar el potencial d'implementar *image classifiers* en tasques de classificació on l'entrada de dades principal són sèries temporals que han estat codificades en imatges per a millorar les tasques de detecció.
- Adaptar el tipus de metodologia indicat en el punt anterior per a desenvolupar una metodologia per a la classificació de defectes de soldadura amb l'objectiu que aquest pugui ser aplicat en sectors on la qualitat de la soldadura sigui crítica.
- Efectuar un estudi comparatiu per a investigar si la metodologia plantejada en aquest treball millora en termes d'exactitud els resultats obtinguts quan s'apliquen xarxes neuronals més tradicionals (LSTM, per exemple).
- Detectar línies futures que permetin l'avanç d'aquest tipus de tecnologies dins el sector marítim.

2.6. Breu descripció dels altres capítols de la memòria

A continuació, es detalla com s'estructura la resta de capítols de la present memòria:

- **Capítol 3: Estat de l'art.** A fi de justificar la necessitat d'aquest estudi, s'analitzen els treballs més rellevants dins de l'àrea de coneixement a tractar amb l'objectiu de definir de manera crítica possibles millores que es poden efectuar en relació amb la identificació automàtica de defectes de soldadura mitjançant l'ús de la intel·ligència artificial.
- **Capítol 4: Metodologia.** En aquest capítol es representa la metodologia desenvolupada a través de la combinació de tècniques de *time series imaging encoding* i *vision transformers* per a identificar defectes en soldadura.
- **Capítol 5. Resultats.** Per a garantir el correcte anàlisi que determini la viabilitat de la metodologia implementada, s'efectuarà un cas d'estudi, els resultats del qual es detallaran en aquest cinquè capítol.
- **Capítol 6. Discussió.** A partir dels resultats presentats en el capítol anterior, s'avaluarà la viabilitat de l'estudi, les millores obtingudes, les limitacions de la metodologia implementada i possibles desavantatges que s'han pogut identificar en aquest capítol.
- **Capítol 7. Conclusions.** En aquest capítol es detallaran les conclusions del treball i una reflexió crítica de l'assoliment dels objectius definits en la secció 2.2. *Objectius del treball*. Segons aquesta reflexió, es determinaran les línies de treball futur i una anàlisi exhaustiu del seguiment de la planificació considerada per al projecte definit en aquest estudi.
- **Capítol 8. Glossari.** En aquest capítol es defineixen els termes i acrònims més rellevants dins la present memòria.
- **Capítol 9. Bibliografia.** En l'últim capítol de la memòria s'introdueix les referències bibliogràfiques més rellevants que han estat utilitzades per a efectuar aquest treball.

3. Estat de l'art

A fi de determinar la situació actual en relació amb la implementació de la intel·ligència artificial per a la detecció automàtica de defectes durant el procés de soldadura, cal fer una anàlisi de l'estat de l'art en aquesta àrea de coneixement. Per a això, s'han identificat els estudis científics més rellevants de les següents bases de dades principals: Scopus i IEEE. Primerament, s'ha efectuat una primera cerca a fi de considerar únicament estudis dins del sector naval. Malgrat això, arran que el nombre d'estudis és limitat, s'ha decidit ampliar aquesta cerca a estudis dins del sector transport i fabricació. En total, s'han analitzat un total de set estudis científics, els quals es resumeixen a continuació:

Jager et al., (2008) va desenvolupar un nou *framework* per a la classificació automàtica de defectes mitjançant la utilització de *Hidden Markov Models* (HMMs). Per a validar la metodologia implementada, es va introduir un cas d'estudi relacionat amb un procés de soldadura de làser industrial. Concretament, les dades experimentals utilitzades van ser recollides a través d'una càmera CMOS que es situava en una de les línies de producció. En total, es van obtenir un total de 99 peces de soldadura amb defectes. De totes les seqüències analitzades, només es va produir un 1,8% de fals positius. Aquests resultats demostren el potencial d'utilitzar HMMs en aquest tipus de tasques, atès que HMMs permeten la consideració de dependències temporals en comparació amb altres models com ara els *Gaussian Mixture Models* (GMMs).

Köhler et al. (2020) va implementar les Convolutional Neural Networks (CNNs) a fi de monitoritzar la qualitat del procés de soldadura. Per a això, es van considerar diferents qualitats de soldadura i diferents tipus d'aliatges a fi d'analitzar les capacitats de generalització de les xarxes neuronals. Per a avaluar l'efectivitat d'aquest tipus de xarxa neuronal, es va aplicar a un cas d'estudi on es va considerar la màquina de soldadura EJOT RES 38/155 RRS C2, la qual presenta velocitat màxima de rotació de 8000 r/min i una força de soldadura màxima de 6 kN. Els experiments es van realitzar amb una força de soldadura de 4.5 kN, una velocitat de rotació de 6000 r/min i una profunditat de 0.2 mm. En total, es van generar 105 conjunt de dades. L'exactitud obtinguda per a aquest cas d'estudi va ser del 88,5%.

Dorbane et al. (2022) va desenvolupar una metodologia per a predir el comportament mecànic d'un aliatge d'alumini per al tipus de soldadura per fricció-agitació. Concretament, van utilitzar les *Gated Recurrent Unit* (GRU), les quals són un tipus de xarxes neuronals recurrents. Es van fer ús d'aquest tipus de xarxes neuronals arran que es van considerar sèries temporals per a entrenar el model. Es van emprar làmines d'aliatge d'alumini Al6061-T6 de 3 mm de gruix com a materials d'unió. A l'implementar una tasca de regressió, es van utilitzar les mètriques *Root Mean Squared Error* (RMSE), *Mean Absolute Error* (MAE), *Mean Absolute Percentage Error* (MAPE) i el *Coefficient of Determination* (R^2). Els resultats van demostrar la viabilitat d'aquest tipus de xarxes neuronals per a predir el comportament mecànic d'aliatges, atès que el R^2 va ser d'1 per al cas d'estudi implementat. Malgrat això, els autors consideraren la necessitat

d'implementar més cas d'estudis amb altres materials i aliatges a fi de garantir la viabilitat d'aquests models.

Tyystjärvi et al. (2024) va modificar l'arquitectura de la xarxa U-Net per a efectuar segmentació d'imatges i localitzar la soldadura efectuada a fi de detectar defectes a través l'ús de radiografies industrials de soldadures de tipus làser. Es van considerar un total de 451 radiografies com a part del cas d'estudi a fi de validar la metodologia implementada. Els resultats obtinguts van demostrar la capacitat de la xarxa U-Net modificada per a la localització de porositats, la qual cosa facilita l'anàlisi de les dimensions i ocurrences al llarg del temps de les porositats en qüestió a fi de monitoritzar possibles defectes que s'originen arran la soldadura aplicada.

Zhang and Wen (2016) va emfatitzar la importància de la fase d'extracció de característiques per a la correcta detecció de defectes. Primerament, van considerar la idea d'implementar atenció visual per a detectar zones d'interès que puguin estar relacionada amb els defectes de soldadura. A continuació, van aplicar estratègies per a millorar la qualitat de la imatge, van aplicar una fase de filtratge i reducció de soroll i van aplicar segmentació d'imatges per a determinar característiques morfològiques de la soldadura, com ara la seva longitud i amplada. Per al cas d'estudi es va considerar alumini de tipus A506. Per a la recol·lecció de dades, es va emprar un sistema d'adquisició d'imatges a través de la utilització de una càmera CDD. Els resultats obtinguts va demostrar la importància d'aquest tipus d'estratègies per a la detecció de defectes de penetració i oxidació de la superfície.

Raoofi et al. (2024) va introduir el YOLOv4 amb l'objectiu de detectar les connexions d'acer a fi de detectar soldadures no completades en temps real. Per a analitzar el correcte funcionament de la metodologia proposada, es va realitzar un cas d'estudi en una de les línies de producció del Group Canam Inc. El cas d'estudi presentava un total de 15.356 imatges anotades amb sis classes diferents, les quals fan referència a estats òptims de soldadura i a estats defectuosos de soldadura. Els resultats generals indiquen el potencial d'aquest tipus d'arquitectures a fi d'efectuar una detecció automàtica de defectes en temps real, atès que el model va obtenir una exactitud de més del 98% per al cas d'estudi proposat.

Asadi et al. (2024) va fer ús de tècniques d'aprenentatge profund per a la detecció, anotació i segmentació de la soldadura en un entorn experimental. Per a això, es van fer ús dels diferents models de YOLO i CNN. Per al cas d'estudi, es van considerar dos conjunts de dades, els quals presentaven imatges abans, durant i després la deposició. En relació amb la segmentació, YOLOv8 va ser el model que va obtenir els millors resultats, atès que va oferir aportacions certament rellevants en relació amb la localització i definició de la soldadura. En addició, si s'avalua el CNN implementat, es pot observar resultats prometedors en la predicció de la geometria de la soldadura, obtenint un *Mean Squared Error* (MSE) de 0,082 per a la predicció de l'àrea, 0,111 per a la predicció de la longitud i 0,334 per a la predicció de l'amplada.

Al realitzar l'anàlisi de l'estat de l'art, es pot observar que la majoria d'estudis contemplen o bé imatges o bé vídeos per a detectar possibles defectes en la soldadura. De fet, dels set estudis analitzats, únicament un va considerar l'anàlisi de sèries temporals. Malgrat això, l'objectiu principal de l'estudi era analitzar el comportament del material durant la soldadura i no la detecció automàtica de defectes. Per això, es considera que l'estudi introduït en aquesta memòria pot ser d'interès en aquesta àrea de coneixement a fi d'analitzar altres mètodes que es puguin utilitzar en la detecció automàtica de soldadura a partir de l'ús de sèries temporals de paràmetres que podrien ser utilitzats per facilitar la comprensió de la qualitat de la soldadura, com ara el voltatge i l'amperatge. A més, es pretén analitzar models d'aprenentatge profund arran els resultats obtinguts en els estudis analitzats, atès que cinc d'aquests estudis van demostrar l'eficiència d'aquest subconjunt de la intel·ligència artificial en termes de precisió. Malgrat això, cap d'aquests cinc estudis consideren la implementació d'un ViT, el qual és considerat en aquest projecte en combinació amb la implementació d'un anàlisi de tècniques de codificació de sèries temporals en imatges a fi de conèixer si aquests tipus de metodologies milloren en la tasca de detecció de defectes de soldadura si es compara amb altres mètodes més tradicionals.

4. Metodologia

Una vegada identificada la necessitat i contribució principal d'aquest projecte, es presenta la metodologia dissenyada. Aquesta està constituïda d'un total de tres fases principals. La primera fase fa referència a la fase de preprocessament de les dades, la qual presenta com a objectius principals garantir la qualitat de les dades, garantir que les dades estan en el format necessari i garantir que les dades són les necessàries. Un cop assolits aquests tres objectius, es continua amb la implementació de la segona fase, la qual presenta com a objectiu principal analitzar les tècniques de codificació de sèries temporals en imatges més significatives. Concretament, s'analitzaran les següents: 1) *Gramian Angular Field* (GAF), 2) *Markov Transition Field*¹² (MTF), 3) *Recurrence Plot* (RP) i 4) *Joint Recurrence Plot*¹³ (JRP). Les imatges resultants d'aquesta segona fase seran els valors d'entrada de l'*image classifier* que permetrà efectuar l'anàlisi de classificació necessari per a detectar automàticament si una soldadura és defectuosa o no. De tots els models disponibles, s'ha seleccionat el *Vision Transformer* (ViT), atès que no s'ha detectat el seu ús en l'estat de l'art efectuat i la seva implementació ha demostrat millorar la precisió en la fase d'inferència per a tasques similars en altres àrees de coneixement.

En els següents paràgrafs d'aquest capítol s'exploraran en detall cadascun dels mètodes implementats a fi d'estudiar el seu funcionament.

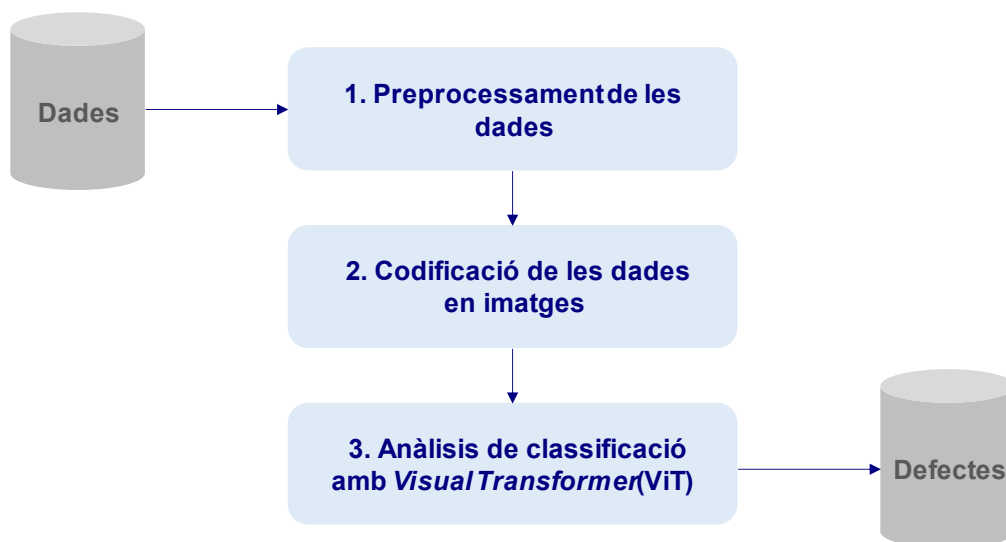


Fig. 2. Representació gràfica de la metodologia proposada.

4.1. Preprocessament de les dades

De totes les tasques de preparació de dades analitzades per al correcte desenvolupament d'aquest projecte, es mencionaran les tres més rellevants. Aquestes són les següents: 1) aplicació de *downsampling* de la sèrie temporal,

2) divisió del conjunt de dades en set d'entrenament i set d'avaluació i 3) normalització de la sèrie temporal.

4.1.1. *Downsampling* de la sèrie temporal

Normalment, les senyals recopilades de sensors contenen soroll que pot inferir negativament en l'entrenament del model. A més, en funció de la freqüència de recollida de les dades, és possible que s'incrementi de manera significativa les dimensions de la senyal sense l'addició d'informació rellevant. En conseqüència, per a reduir el cost computacional i eliminar el soroll en la senyal, en aquest estudi s'aplica *downsampling*, el qual permet reduir el nombre d'instàncies de la senyal.

En aquest cas, per a aplicar *downsampling*, s'ha dividit el nombre d'instàncies en n grups, el qual s'ha definit en funció d'un factor de *downsampling*, k (veure equació 1). Per a cada grup, s'ha agafat l'últim valor de la seqüència.

$$n = \frac{N}{k}, \quad (\text{Eq. 1})$$

on N és el nombre total d'instàncies de la senyal i k és el factor de *downsampling*,

4.1.2. Divisió del conjunt de dades en set d'entrenament i set d'avaluació

Per a la correcta avaluació del model, atès que es fan ús de dades per tal d'entrenar el model, és necessari dividir el conjunt de dades en dos sets, els quals són el set d'entrenament i el set d'avaluació.

El set d'entrenament es fa ús per a entrenar el model, és a dir, són les dades que es consideraran per a actualitzar els diferents pesos i biaixos que derivarà en una millora en la classificació de defectes.

Malgrat això, si es considera el set d'entrenament per a avaluar el model, probablement els resultats no seran realistes, atès que les prediccions seran probablement precises arran que s'estan considerant les dades que han estat utilitzades per a millorar el rendiment del model. En conseqüència, cal fer ús d'un set nou que no hagi estat emprat en la fase d'entrenament i, aquest, es el que es denomina com a set d'avaluació.

D'aquesta manera, és a dir, amb aquesta divisió, s'esperen que els resultats siguin més realistes, atès que s'utilitzen dades que no han estat considerades en cap moment per al model a fi de garantir la capacitat de generalització del model.

A més, al no considerar dades del set d'entrenament en la fase d'avaluació, es minimitzen els biaixos que poden sorgir en aquesta fase en qüestió.

Per últim, cal mencionar que el set d'entrenament s'ha dividit novament en set d'entrenament i set de validació, a fi d'utilitzar aquest últim set per a la configuració dels hiperparàmetres del model implementat.

La selecció dels registres per a considerar-los o bé en un set o bé en un altre s'ha efectuat de manera aleatòria. S'ha considerat que el 70% del conjunt de dades serà part del set d'entrenament i el 30% del set de test. El set de validació serà el 20% del set d'entrenament.

4.1.3. Normalització de la sèrie temporal

Per a finalitzar aquesta primera fase, s'ha dut a terme la normalització de les sèries temporals. Aquesta normalització és necessària perquè els diferents paràmetres utilitzats presenten rangs diferents i, per tant, si no es gestiona aquesta diferència, pot provocar que un paràmetre tingui més influència sobre el model que l'altre simplement perquè el rang de valors és major, encara que el paràmetre en qüestió sigui menys rellevant. A més, la normalització de valors presenta certs beneficis quan es consideren models d'aprenentatge profund, atès que facilita la convergència, presenta efectes de regularització i simplifica la inicialització dels paràmetres, per exemple.

De totes les possibles normalitzacions, s'ha considerat la normalització per mínim i màxim. D'aquesta manera, es garanteix que totes les senyals d'entrada considerades presenten el mateix rang de valors (0, 1). Per a efectuar aquesta normalització, s'ha fet ús de la següent equació:

$$x_{normalised} = \frac{x - x_{min}}{x_{max} - x_{min}}, \quad (\text{Eq. 2})$$

on x és la senyal a normalitzar, x_{min} és el valor mínim de la senyal i x_{max} és el valor màxim de la senyal.

És necessari considerar que normalment es requereix d'altres fases per a garantir la qualitat i el format de les dades, com ara la implementació d'imputació de dades faltant i la detecció de valors atípics. Malgrat això, arran que les dades ja estaven preprocessades, les subfases a implementar en aquesta fase de preparació de les dades és mínima.

4.2. Codificació de les dades en imatges

Aquest estudi presenta com a objectiu principal el d'analitzar el potencial que poden presentar els mètodes de codificació de sèries temporals d'imatges en termes de precisió i cost computacional. La selecció d'aquest mètodes d'anàlisi ha estat basat en la seva capacitat d'extreure les característiques principals de les sèries temporals i la possibilitat d'utilitzar xarxes neuronals que permeten la classificació d'imatges (*imatge classifiers*) amb l'objectiu d'analitzar si aquesta combinació permet millorar la precisió en l'elaboració de la tasca de detecció.

4.2.1. Gramian Angular Field (GAF)

A fi de codificar la sèrie temporal en imatge mitjançant el mètode de *Gramian Angular Field* (GAF), és necessari representar la sèrie temporal en coordenades polars a través de la codificació del valor com el cosinus angular i el *timestamp* com el radi, és a dir:

$$\begin{cases} \phi = \arccos(\tilde{x}_i), & -1 \leq \tilde{x}_i \leq 1, \tilde{x}_i \in \tilde{X} \\ r = \frac{t_i}{N}, & t_i \in \mathbb{N} \end{cases} \quad (\text{Eq. 3})$$

on, t_i és el *timestamp* i N és un factor constatat, l'objectiu del qual és regularitzar l'*span* del sistema de coordenades polar. Per a estimar \tilde{x}_i , que és la sèrie temporal analitzada normalitzada en un rang $[a, b]$, es fa ús de la següent equació:

$$\tilde{x}_i = a + (b - a) \cdot \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (\text{Eq. 4})$$

Un cop transformades la sèrie temporal, es pot procedir a la codificació de la imatge aplicant o bé el *Gramian Angular Summation Field*¹⁴ (GASF) o bé el *Gramian Difference Angular Field*¹⁵ (GADF).

Per a codificar la sèrie temporal en imatge mitjançant l'ús de GASF, s'efectua la suma trigonomètrica entre cada instància amb l'objectiu d'identificar la correlació temporal entre els intervals analitzats. Per a això, s'aplica la següent equació:

$$GASF = [\cos(\phi_i + \phi_j)] \quad (\text{Eq. 5})$$

De manera similar, es pot obtenir la imatge codificada segons el mètode GADF a través de la següent equació:

$$GADF = [\sin(\phi_i + \phi_j)] \quad (\text{Eq. 6})$$

4.2.2. Markov Transition Field (MTF)

Una altra manera de codificar la sèrie temporal en imatge és mitjançant la determinació de la matriu de transició de Markov. Aquesta matriu de transició es defineix tal i com s'indica a continuació:

$$M = \begin{bmatrix} w_{ij|x_1 \in q_i, x_1 \in q_j} & \cdots & w_{ij|x_1 \in q_i, x_n \in q_j} \\ w_{ij|x_2 \in q_i, x_1 \in q_j} & \cdots & w_{ij|x_2 \in q_i, x_n \in q_j} \\ \vdots & \ddots & \vdots \\ w_{ij|x_n \in q_i, x_1 \in q_j} & \cdots & w_{ij|x_n \in q_i, x_n \in q_j} \end{bmatrix}, \quad (\text{Eq. 7})$$

on w_{ij} es la freqüència donada considerant aquelles instàncies que pertanyen al quartil q_j que son seguides per instàncies que pertanyen al quartil q_i .

4.2.3. Recurrence Plot (RP)

Per a obtenir el *Recurrence Plot* (RP), primer cal determinar les diferents trajectòries, les quals s'obtenen a través de l'aplicació de la següent equació:

$$\vec{x}_i = (x_i, x_{i+\tau}, \dots, x_{i+(m-1)\tau}), \forall i \in \{1, \dots, n - (m-1)\tau\} \quad (\text{Eq. 8})$$

on m és el nombre de dimensions de la trajectòria donada i τ és *time delay*. A l'obtenir les trajectòries, es possible obtenir el RP, atès que aquest es defineix com la *pairwise distance* entre trajectòries, distància que es determina seguint la següent equació:

$$R_{i,j} = \Theta(\varepsilon - \|\vec{x}_i - \vec{x}_j\|), \forall i \in \{1, \dots, n - (m-1)\tau\} \quad (\text{Eq. 9})$$

on Θ és la funció de *Heaviside* i ε és el llindar.

En relació amb el *Joint Recurrence Plot* (JRP), aquest es tracta d'una extensió del RP, el qual permet implementar anàlisis multivariant. Per a això, es considera el producte de *Hadamard* per a cada RP resultant de cadascuna de les variables independents analitzades.

4.2.4. Markov Transition Matrix (MTM)

Per últim, també s'analitza l'obtenció de la matriu de transició mitjançant *first-order Markov chain* a través del següent procediment:

1. Es determina el nombre d'estats a considerar. En aquest cas, aquests es determinen mitjançant avaluació heurística.
2. Es determinen totes les possibles probabilitats de transició entre estats, les quals s'indiquen en una matriu de $r \times r$, és a dir:

$$\mathbf{P} = (p_{ij})_{1 \leq i, j \leq r} = \begin{pmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,r} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,r} \\ \vdots & \vdots & \ddots & \vdots \\ p_{r,1} & p_{r,2} & \cdots & p_{r,r} \end{pmatrix} \quad (\text{Eq. 10})$$

i que satisfà

$$0 \leq p_{ij} \leq 1, \quad 1 \leq i, j \leq r$$

$$\sum_{j=1}^r p_{ij} = 1, \quad 1 \leq i \leq r$$

4.3. Anàlisi de classificació amb *Visual Transformer*

La xarxa neuronal per excel·lència per a efectuar la tasca de classificació d'imatges és la *Convolutional Neural Network* o, al menys, fins que es va originar el *Vision Transformer* (ViT). El ViT és un tipus de xarxa neuronal, l'arquitectura del qual es basa en la implementació de *transformers* i considera les imatges com a una seqüència de pedaços (*patches*, en anglès), tal i com es mostra en la Figura 3. Amb l'objectiu d'entendre millor la seva arquitectura, en les següents subseccions es detallen les seves parts principals.

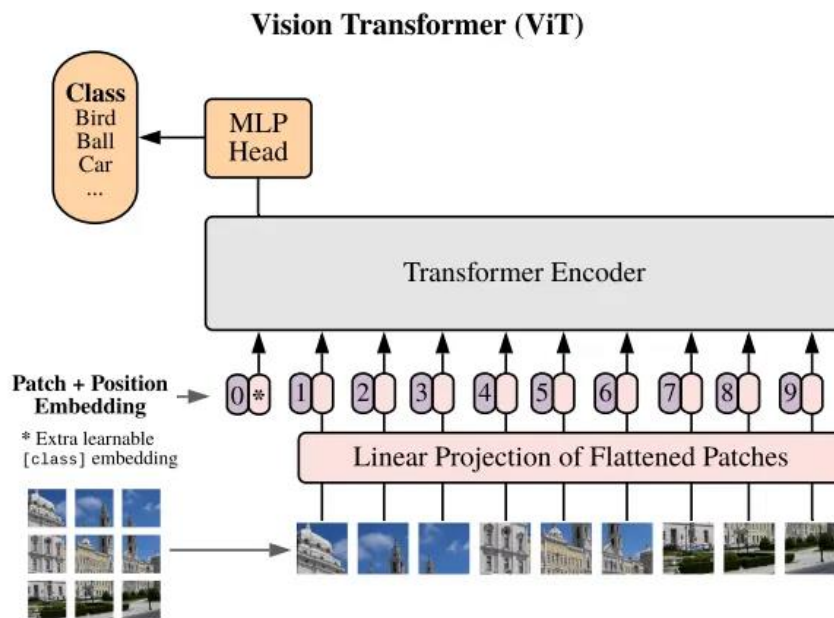


Fig. 3. Representació gràfica d'un *Vision Transformer* (font: Dosovitskiy et al., 2021)).

4.3.1. Patch Embedding Layer i Positional Encoding

Primerament, cal considerar l'input d'aquesta xarxa neuronal. Arran que per a aquest estudi es tracten imatges en escala de grisos, les imatges presenten unes dimensions de $H \times W \times C$, és a dir, *altura* \times *amplada* \times *nombre de canals*, on, en aquest cas, el nombre de canals és 1. Aquestes imatges es divideixen en *patches* de $P \times P$. El nombre de *patches* resultant (N) serà $\frac{H}{P} \times \frac{W}{P}$.

A continuació, s'aplica una capa de *Flattening* per tal de transformar cada *patch* resultant en un vector 1D de dimensió $P \times P \times C$. A cada vector resultat se li aplica una capa lineal a fi d'obtenir un vector de dimensió D amb l'objectiu d'obtenir un vector d'incrustació de mida fixa. Per finalitzar, s'aplica *positional encoding* a la seqüència resultant per tal de presentar la informació de posició necessària a cada *patch* a fi de retenir l'estructura espacial de la imatge.

4.3.2. Transformer Encoder

El *transformer encoder* està constituït per *multi-head self-attention mechanisms*, els quals són mecanismes que permeten al model "focalitzar la seva atenció" en les parts més rellevants de la imatge. Aquesta atenció s'obté a través de la consideració d'una puntuació d'atenció per a cada *patch* resultant de la fase anterior. D'aquesta manera, també es poden determinar aquells *patches* que presenten les característiques principals que permeten resoldre la tasca

desitjada a resoldre per la xarxa neuronal. A fi d'obtenir aquestes puntuacions, cal considerar els següents tres components: *Query* (Q), *Key* (K) i *Value* (V). Aquests tres components s'obtenen de la consideració dels *input embeddings*. A continuació, s'indica l'equació considerada per a l'estimació de les puntuacions d'atenció:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (\text{Eq. 11})$$

on d_k és la dimensió del vector de claus. Tant aquests tres components com la puntuació es determinen de forma simultània per a diferents parts de les característiques d'entrada. D'aquí que el mecanisme consideri el terme *multi-head* com a part de la seva definició. A més, amb l'objectiu d'estabilitzar i accelerar l'aprenentatge, també s'apliquen capes de normalització i connexions residuals, atès que aquestes han demostrat garantir el flux de gradient i la millora del rendiment del model.

Per a processar la informació resultant de les capes del mecanisme d'atenció i introduir no linearitat, també s'implementa una *feed-forward neural network* constituïda de dues capes lineals amb una funció d'activació no lineal entre aquestes capes.

4.3.3. Classification Head

Per finalitzar, s'implementa una *feed-forward neural network* per a efectuar l'anàlisi de classificació que, en aquest cas, tracta sobre la detecció de defectes en soldadura. Per tant, al tractar-se d'una classificació binària (defecte o no defecte), l'última capa d'aquesta xarxa presentarà dues unitats amb funció d'activació *softmax* i funció de pèrdua *Binary CrossEntropy*.

Tot el codi elaborat per a implementar aquesta metodologia i el procés d'avaluació d'aquesta es pot consultar en:

<https://github.com/cvegas00/WeldingDefectClassification>.

Per a la correcta implementació d'aquest codi, s'ha fet ús principalment de les llibreries **Pandas**, **Numpy**, **Pyts**, **Tensorflow** i **Keras**. El codi relacionat amb la implementació de les xarxes neuronals s'ha basat en codi ja existent dels manuals de Tensorflow i Keras.

5. Resultats

En aquesta secció s'introdueix el cas d'estudi implementat i els resultats derivats amb l'objectiu d'analitzar el rendiment de la metodologia considerada en aquest estudi.

5.1. Cas d'estudi

El conjunt de dades emprat per a avaluar els models implementats consisteix en un total de 209185 instàncies. Cada instància fa referència a un cicle de soldadura, del qual s'ha extret la senyal de voltatge i amperatge corresponent. La freqüència de recol·lecció és de 100 kHz i la dimensió de cada senyal és de 200. Cada procés de soldadura ha estat categoritzat segons la seva qualitat, és a dir, si la qualitat d'aquesta és bona o dolenta. Malgrat això, també s'han detectat instàncies no categoritzades, les quals han estat etiquetades amb el valor de -1. En la Taula 2 s'indiquen les característiques principals del conjunt de dades utilitzat.

La selecció d'aquest conjunt de dades ha estat principalment basat en la qualitat de les dades, les quals es troben adequadament preprocessades, no s'ha detectat cap valor faltant i el format d'exportació de les dades és simple i efectiu. A més, després d'una cerca exhaustiva, és un dels conjunt de dades més complets que s'han detectat que tracten sobre sèries temporals, atès que la majoria de conjunt de dades disponibles en obert són o bé imatges o bé vídeos efectuades durant el procés de soldadura. Per a obtenir una informació més detallada d'aquest conjunt de dades es recomana consultar (Hahn et al., 2023).

Taula 2. Característiques principals del conjunt de dades utilitzat per a l'avaluació de la metodologia implementada.

Nombre d'instàncies	209185
Nombre de paràmetres	2
Paràmetres	Voltatge i Amperatge
Tipus de dades	Numèrica continua
Freqüència de recol·lecció	100kHz
Màxim error permès	0.5%
Categorització de la qualitat de la soldadura	0: mala qualitat 1: bona qualitat -1: N/A

5.2. Anàlisi exploratori de les dades

A continuació, es descriu molt breument l'anàlisi exploratori de les dades que s'ha efectuat en relació amb el conjunt de dades seleccionat. Un dels primers passos que s'ha efectuat en aquest aspecte és el d'analitzar si les classes presentades en el conjunt de dades es troben balancejades, atès que si aquestes no ho estan poden afectar negativament en el rendiment del model. En aquest cas, tal i com es pot observar en la Taula 3, les dues classes principals d'anàlisi es troben balancejades (49742 instàncies de processos de mala qualitat i 46666 instàncies de processos de bona qualitat). A més es detecta, que hi ha un total de 112777 instàncies de processos de soldadura que es desconeix si es tracta d'un procés de bona qualitat o de mala qualitat. En conseqüència, aquestes instàncies són eliminades del conjunt de dades a fi de no ser incloses en l'estudi.

Taula 3. Categorització de les senyals d'amperatge i voltatge en funció la qualitat de la soldadura.

Categoria	Descripció	Nombre d'instàncies
0	Mala qualitat	49742
1	Bona qualitat	46666
-1	Es desconeix si la qualitat es bona o dolenta	112777

Tot seguit, es fa una representació gràfica d'exemples de senyals de voltatge i amperatge que fan referència a un procés de soldadura de mala qualitat (Figura 4) i exemples de senyals de voltatge i amperatge que fan referència a un procés de soldadura de bona qualitat (Figura 5).

Tal i com es pot observar en aquestes figures, cada cicle de soldadura es pot seccionar en les següents tres fases:

- **Fase 1: fase de pols.** La font de corrent augmenta amb l'objectiu d'escalfar i liquar el filferro utilitzat en el procés de soldadura. Per tal de facilitar l'energia suficient que es requereix per a assolir l'anterior objectiu, es requereix que aquesta energia es pugui transferir ràpidament. D'aquí que es pugui observar una forta pujada de corrent a l'inici del procés de soldadura. Aquesta fase finalitza quan el corrent arriba a un altiplà.
- **Fase 2: fase de desprendiment de gotes.** Quan la gota arriba a una certa dimensió provocada per l'escalfament del filferro, aquesta es desprèn del filferro a fi de ser expulsada cap a l'àrea on es desitja situar la soldadura. Per això, es produeix una reducció en la inducció d'energia, atès que llavors s'evita que la gota s'escalfi massa i es derivi en una deformació longitudinal d'aquesta.
- **Fase 3: fase de corrent.** Tal i com es poden observar en les figures, es produeix una segona reducció en la inducció d'energia a fi de garantir la correcta definició del següent cicle i garantir que l'estat energètic està definit de manera que permeti la correcta formació de la següent gota.

Inicialment, a través d'una inspecció visual ràpida dels exemples introduïts, no s'observa una diferenciació molt característica entre un procés de soldadura qualificat com a bo i un procés considerant com a dolent.

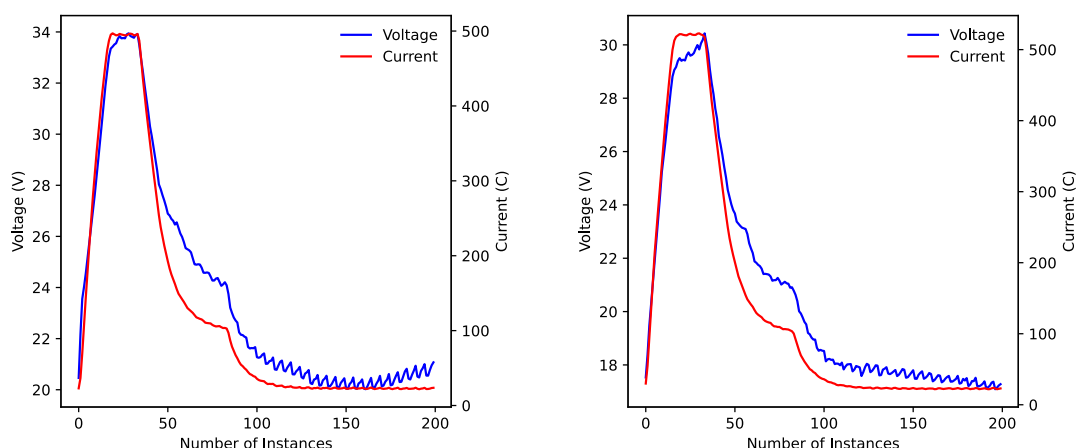


Fig. 4. Representació gràfica de dos exemples de senyals d'amperatge i voltatge associades a una mala qualitat de soldadura.

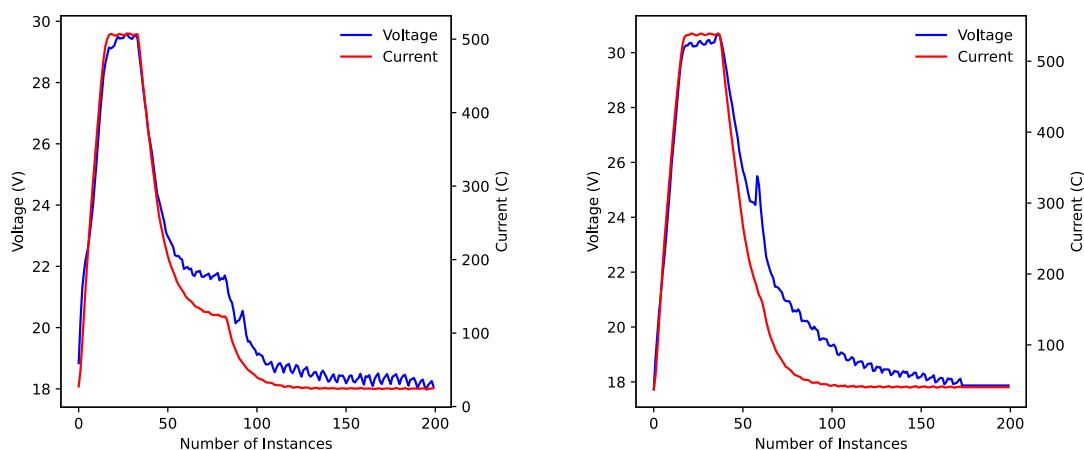


Fig. 5. Representació gràfica de dos exemples de senyals d'amperatge i voltatge associades a una qualitat bona de soldadura.

5.3. Anàlisi dels resultats de la fase de preparació de les dades i de codificació de sèries temporals en imatges

La primera fase de la preparació de les dades consisteix en aplicar *downsampling* a les senyals a fi de reduir la dimensió de la senyal i, en conseqüència, reduir el cost computacional. Malgrat això, cal tenir present que mitjançant l'aplicació d'aquesta tècnica és possible que hi hagi una pèrdua significativa de la informació que presenta les senyals i que això derivi en una pèrdua del rendiment del model. La figura 6 mostra un exemple de senyal abans d'aplicar la tècnica de *downsampling* i després d'aplicar *downsampling*. Es pot

observar que l'aplicació d'aquesta tècnica permet una suavització de la senyal i una reducció del soroll. No obstant això, també es pot observar una possible pèrdua de certes fluctuacions que poden ser rellevants per a la determinació de si la soldadura presenta un cert defecte. A més, també es pot observar una lleugera alteració de la senyal.

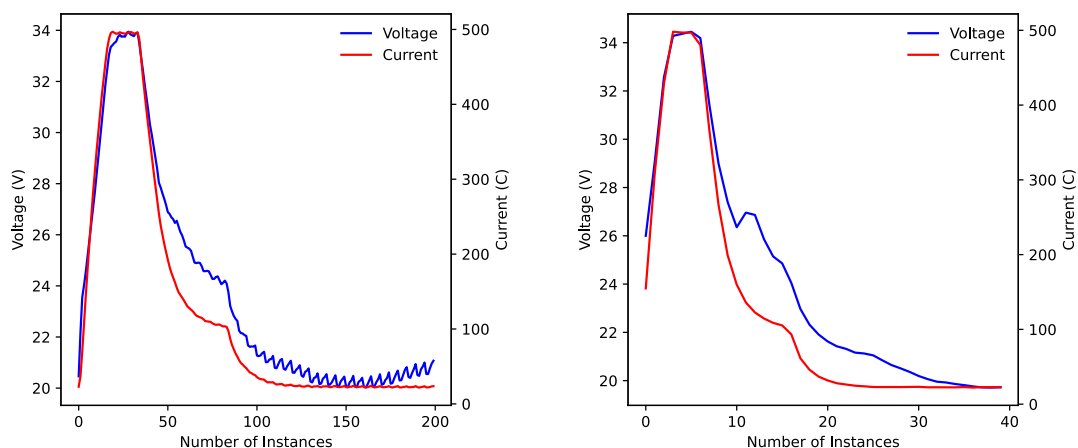


Fig. 6. Representació gràfica de l'abans i després d'aplicar *downsampling* a una senyal de voltatge i amperatge.

Un cop fet *downsampling* a totes les senyals a analitzar, s'ha implementat normalització a fi de que totes les senyals presentin un rang entre 0 i 1. Arran que les dades estaven preprocessades inicialment, no ha calgut aplicar cap altra tasca de preparació de dades addicional.

Per això, es pot procedir a la codificació de les senyals en imatges. Tal i com s'ha indicat a la part metodològica, s'han fet ús d'un total de sis mètodes: 1) GASF, 2) GADF, 3) MTF, 4) RP, 5) JRP i 6) MTM. Exemples d'aquestes codificacions es mostren en les Figures 7-10. Si s'observen aquests exemples en detall, es pot contemplar que les diferenciacions entre una RP de soldadura de bona qualitat i una RP de soldadura de mala qualitat és mínima. Aquest comportament és similar per a GASF i GADF. En canvi, es poden observar diferències molt significants si s'analitzen les imatges resultant d'aplicar MTF, JRP i MTM.

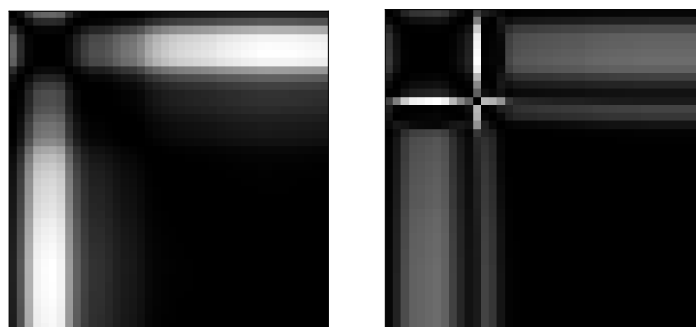


Fig. 7. Representació gràfica dels resultats de codificar un exemple de senyals associades a mala qualitat (figura de l'esquerra) i bona qualitat (figura de l'esquerra) mitjançant l'ús de JRP.

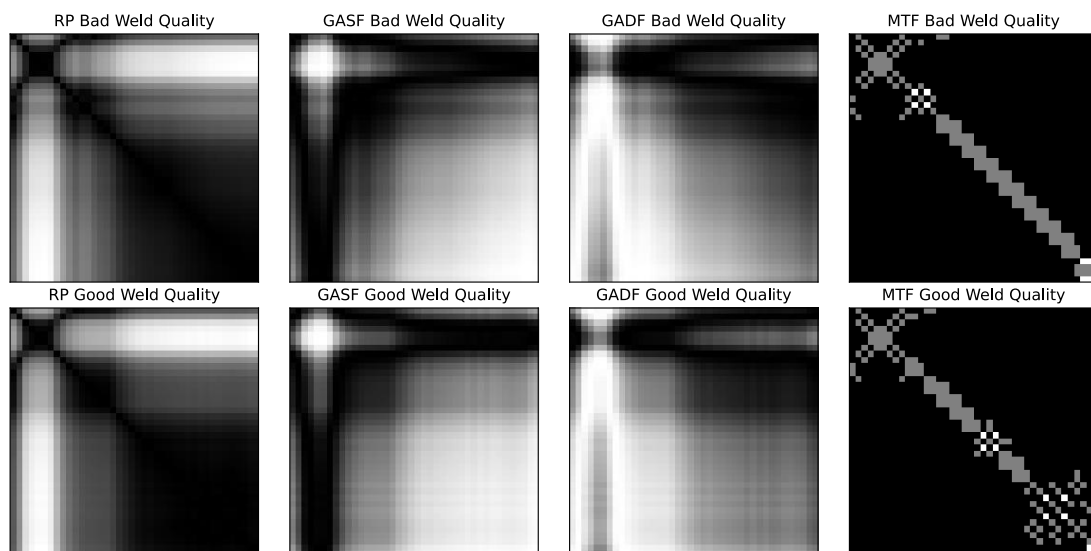


Fig. 8. Representació gràfica de dos exemples de codificació de sèries temporals del paràmetre voltatge en imatges. La primera fila d'imatges fa referència a un exemple de mala qualitat en la soldadura. La segona fila, en canvi, fa referència a un exemple de bona qualitat en la soldadura. En relació amb les columnes, aquestes fan referència a diferents tipus de metodologies utilitzades per a l'anàlisi de codificació de sèries temporals en imatges. Concretament, les següents: RP, GASF, GADF, MTF.

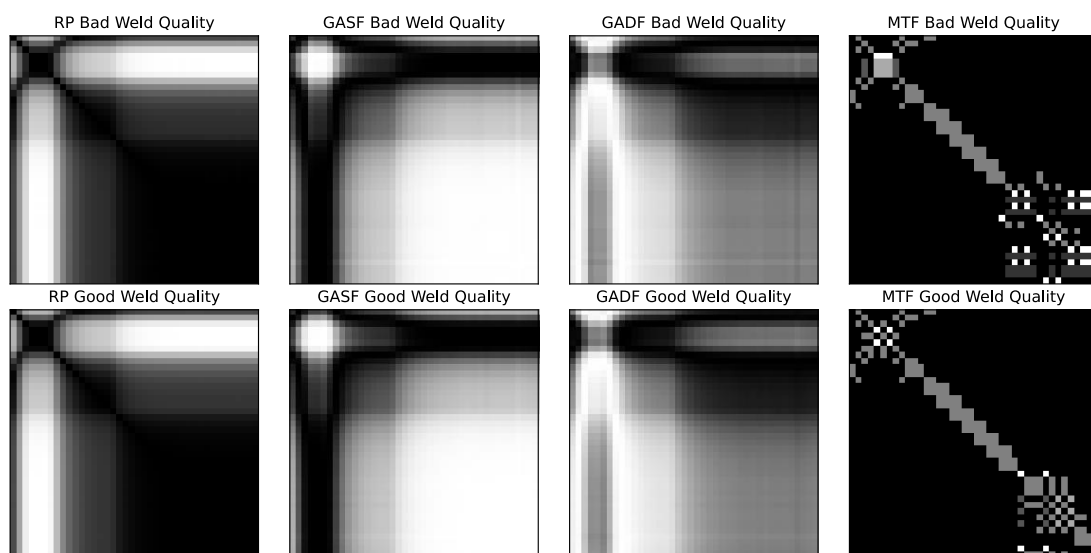


Fig. 9. Representació gràfica de dos exemples de codificació de sèries temporals del paràmetre amperatge en imatges. La primera fila d'imatges fa referència a un exemple de mala qualitat en la soldadura. La segona fila, en canvi, fa referència a un exemple de bona qualitat en la soldadura. En relació amb les columnes, aquestes fan referència a diferents tipus de metodologies utilitzades per a l'anàlisi de codificació de sèries temporals en imatges. Concretament, les següents: RP, GASF, GADF, MTF.

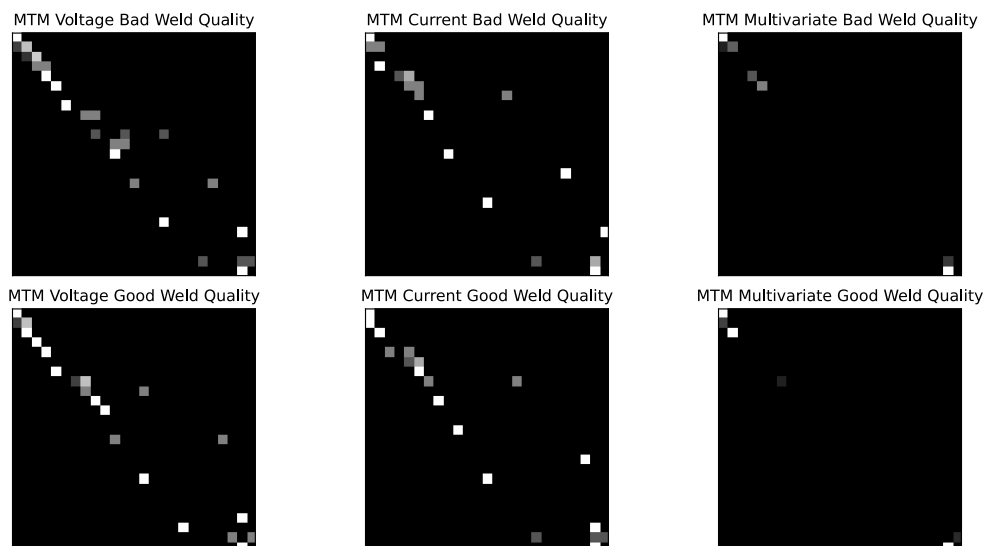


Fig. 10. Representació gràfica dels resultats de codificar un exemple de senyals associades a mala qualitat i bona qualitat mitjançant l'ús de MTM.

5.4. Avaluació del model *Vision Transformer* implementat per a la classificació de defectes en soldadures

Un cop produïda la codificació de sèries temporals a imatges, es pot iniciar la definició i entrenament del ViT amb l'objectiu de classificar els processos de soldadura en funció dels seus defectes i, en conseqüència, la seva qualitat.

La definició de l'arquitectura del ViT implementat s'indica a la Taula 4. Totes les imatges han estat codificades en unes dimensions d'amplada i altura de 40 x 40, excepte les MTM, les quals han pogut comprimir-se a 25 x 25 sense perdre informació rellevant. De manera similar, s'ha considerat una dimensió de *patch* de 4 x 4, a excepció de les MTM que s'ha considerat una dimensió de 5 x 5. Les dimensions de projecció han estat de 64 i el nombre de caps considerat 4. En total, s'han definit 8 *transformers*. Les unitats dels *MLP Head* són 2048 i 1024, és a dir, s'han considerat un total de dos capes de tipus *Dense*. Al tractar-se d'una tasca de classificació binària, el nombre d'unitats de l'última capa de la xarxa neuronal és 2 i la funció d'activació utilitzada és softmax.

Taula 4. Configuració de l'arquitectura del ViT.

Input Shape	n_images x 40 x 40
Patch Size	4
Projection Dimension	64
Nombre de Heads	4
Nombre de Transformer Layers	8
MLP Head Units	[2048, 1024]
Output Shape	2

En relació amb la configuració dels paràmetres d'optimització i entrenament, aquests es detallen en la taula 5. L'optimitzador considerat és l'Adam, el qual és un dels més comuns arran la seva eficiència degut a la consideració d'un *momentum* que evita caure en mínims locals de poca profunditat. El *learning rate* i *weight decay* són 0.001 i 0.0001, respectivament. La funció de pèrdua al tractar-se d'una tasca de classificació binària s'ha considerat la *binary crossentropy*. De manera similar, s'ha considerat la *binary accuracy* com a mètrica d'avaluació. Per últim, s'ha considerat un *batch size* de 256 i uns *epochs* de 10. El plantejament inicial era considerar un total de 100 *epochs* i aplicar *EarlyStopping*. Malgrat això, arran l'elevat cost computacional i el nombre de models que s'han implementat per a aquest estudi, s'ha hagut de reduir a 10 a fi d'obtenir tots els resultats desitjats. A més, s'ha observat que per a aquest estudi l'aprenentatge era limitat després de 10 *epochs* i, per tant, s'assumeix que els resultats no haguessin canviat de manera significativa si s'hagués considerat 100 *epochs*. Malgrat això, aquest aspecte es tractarà en treball futur.

Taula 5. Configuració dels paràmetres d'optimització i entrenament.

Optimiser	Adam
Learning Rate	0.001
Weight Decay	0.0001
Loss Functions	<i>Binary CrossEntropy</i>
Evaluation Metric durant la fase d'entrenament	<i>Binary Accuracy</i>
Batch Size	256
Epochs	10

Els resultats d'aplicar ViT per a cada imatge obtinguda es detallen en la Taula 6. Primerament, si s'analitzen els resultats segons el paràmetre utilitzat, es pot observar que el paràmetre voltatge proporciona informació més rellevant que el paràmetre corrent. Per exemple, si es considera el mètode GASF, hi ha una millora del 0.11 si es considera el paràmetre voltatge (*accuracy* obtingut 0.66) si es compara amb el paràmetre de corrent (*accuracy* obtingut 0.55). S'obtenen uns resultats similars si es consideren els mètodes GADF, MTF i MTM. També es pot observar que els mètodes multivariants en aquests cas no presenten un rendiment adequat. Una possible raó és la interferència del paràmetre corrent amb el paràmetre voltatge. Si es comparen tots els mètodes utilitzats, GADF és el mètode de codificació de sèries temporals a imatges que proporciona els millors resultats per aquests estudi en qüestions d'exactitud (*accuracy* del 0.71 sobre el set de test). Malgrat això, si es considera també el temps que es requereix per a efectuar l'entrenament del ViT, llavors, probablement el millor model seria el MTM, atès que pot assolir un *accuracy* del 0.63 amb un entrenament cinc vegades més ràpid, aproximadament (el temps d'entrenament del *GADF Voltage* és de 623 segons per *epoch*, mentre que el del *MTM Voltage* és de 110.5 segons per *epoch*). Cal dir que aquesta millora en el temps d'entrenament s'obté a una millor compressió de la imatge resultant pel MTM, el qual obté una informació similar amb menor dimensions.

Arran que el millor mètode en termes d'exactitud ha estat el GADF *Voltage*, aquest és el mètode que es considerarà en la segona fase del estudi, el qual consisteix en efectuar en estudi comparatiu amb xarxes neuronals més tradicionals que acostumen a ser emprades per aquest tipus d'estudis.

Taula 6. Resultats principals dels diferents mètodes de codificació de sèries temporals a imatges en combinació amb ViT.

Combinació d'imatges	<i>Accuracy</i>	<i>Training time [seconds/epoch]</i>
RP Current	0.52	570.4
RP Voltage	0.52	578.3
GASF Current	0.55	598.8
GASF Voltage	0.66	567.7
GADF Current	0.55	619.9
GADF Voltage	0.71	623.0
MTF Current	0.52	551.6
MTF Voltage	0.68	547.1
MTM Current	0.52	103.6
MTM Voltage	0.63	110.5
JRP (Multivariate)	0.52	506.4
MTM (Multivariate)	0.52	103.2

5.5. Estudi comparatiu

Tal i com es pot observar en la Taula 7, es presenten resultats similars en termes d'exactitud si es compara amb els resultats de la secció anterior. Per exemple, l'aplicació d'una FFN i LSTM no és la selecció més adequada per a aquest tipus d'estudi, atès que l'*accuracy* obtingut és relativament baix (menys del 0.6). A més, les LSTM presenten un cost computacional certament elevat. Per exemple, aquest tipus de xarxa ha presentat un temps mitjà per *epoch* de 957 segons, aproximadament. És a dir, gairebé 40 vegades més que el temps d'entrenament d'una 1D-CNN, la qual presenta un temps mitjà per *epoch* de 25 segons.

Els millors resultats s'obtenen amb l'aplicació de les CNN. Tant si s'aplica 1D-CNN com 2D-CNN amb la codificació d'imatge, els resultats són similars. Concretament, ambdós obtenen uns resultats que s'apropen al 0.8 d'*accuracy*. Això demostra que, tot i que els ViT han sorgit com a una alternativa a les CNN en tasques de classificació d'imatges, les CNN són encara la primera opció per a tasques molt específiques on el nombre de dades pot ser limitat. En addició, si es compara en termes de cost computacional, es pot observar que una ViT requereix de molt més temps d'entrenament que una CNN, atès que el temps mitjà per *epoch* de la GADF *Voltage* ViT en aquest cas ha estat de 623 segons. En canvi, la GADF *Voltage* 2D-CNN presenta un temps mitjà per *epoch* de 35 segons.

Per tant, un cop analitzats tots els resultats obtinguts, es pot considerar que el millor model per a resoldre la tasca del cas estudi plantejat és o bé el 1D-CNN o bé el GADF *Voltage* 2D-CNN.

La configuració de les xarxes neuronals utilitzades per a elaborar aquest estudi comparatiu està definida als annexos d'aquesta memòria.

Taula 7. Resultats principals de l'estudi comparatiu.

Combinació d'imatges	<i>Accuracy</i>
FFN (Voltatge)	0.55
FFN (Amperatge)	0.52
LSTM	0.52
1D-CNN	0.77
GADF <i>Voltage</i> 2D-CNN	0.75
Model proposat	0.71

6. Discussió

En aquest estudi es volien resoldre dos qüestions principals: té sentit codificar sèries temporals en imatges per tal d'emprar xarxes neuronals especialitzades en la utilització d'imatges per a la tasca de classificació de defectes de soldadura? Té sentit emprar ViT com a alternativa a altres xarxes neuronals més tradicionals?

En relació amb la primera qüestió, es considera que sí, que la codificació té sentit. Els resultats demostren que amb una correcta selecció del mètode de codificació el rendiment de certs models poden millorar. A més, aquestes codificacions permeten reduir les dimensions dels valors d'entrada mantenint la informació principal que permet diferenciar els diferents defectes durant el procés de soldadura. Aquestes codificacions també permeten extreure certs patrons que no poden ser identificats a través de la senyal inicial, la qual cosa permet millorar els rendiments del model. Per tant, es considera que té sentit codificar sèries temporals en imatges per tal d'emprar xarxes neuronals especialitzades en la utilització d'imatges per a la tasca de classificació de defectes de soldadura. Malgrat això, cal investigar més sobre el seu potencial. En l'apartat de treball futur s'indiquen algunes línies en aquest sentit que l'estudiant d'aquest estudi vol investigar.

En canvi, en relació amb la resposta de la segona qüestió, es considera que ViT ha proporcionat uns resultats satisfactoris, però que hi ha millors opcions per a elaborar aquest tipus de tasques. La millor alternativa és probablement les CNN, atès que el cost computacional associat al seu entrenament és molt menor que el cost de les ViT. A més, ha demostrat obtenir millors nivells d'exactitud per a tasques molt especialitzades on el nombre de dades és limitat. Per això, també caldria considerar les 1D-CNN, atès que és el model que millors resultats ha proporcionat en aquest estudi.

7. Conclusions

7.1. Conclusions

El procés de soldadura és un dels processos més rellevants en relació amb la construcció de vaixells, atès que les estructures dels cascos estan constituïdes per milers de peces, les quals estan unides entre si a través de connexions soldades. Arran d'això, és necessari garantir la qualitat de soldadura amb l'objectiu de prevenir possibles defectes que es poden produir per una incorrecta aplicació del procés.

L'objectiu d'aquest estudi era el d'analitzar la possibilitat de detectar de manera automàtica defectes en la soldadura a través de l'aplicació de la intel·ligència artificial. Concretament, mitjançant l'ús de tècniques de codificació de sèries temporals en imatges i *visual transformers*. En total, s'han considerat quatre mètodes de codificació diferents. Els mètodes en qüestió són els següents: *Gramian Angular Field*, *Markov Transition Field*, *Recurrence Plot* i *Markov Transition Matrix*.

Amb l'objectiu d'avaluar la metodologia plantejada, s'ha considerat un cas d'estudi que presentava un total de 209185 cicles de soldadura etiquetats segons el seu grau de qualitat. A més, també s'ha efectuat un estudi comparatiu de la metodologia proposada amb altres xarxes neuronals més tradicionals, com ara les *Feed Forward Neural Networks*, les *Convolutional Neural Networks* i les *Long Short-Term Memory Neural Networks*.

Els resultats indiquen que, de tots els mètodes de codificació analitzats, el *Gramian Angular Difference Field* és el mètode que ha proporcionat el millor resultat, obtenint, així, un *accuracy* superior al 70%. En addició, dels dos paràmetres utilitzats, el paràmetre de voltatge és el més característic. En relació amb l'estudi comparatiu, tot sembla indicar que la millor opció en relació amb la selecció de la xarxa neuronal per a implementar la tasca de classificació és la *Convolutional Neural Network*, la qual gairebé obté un *accuracy* del 80%.

En relació amb les lliçons apreses, es destaquen les següents:

- **Aprendre sobre soldadura.** Tot i que l'àrea de l'estudiant és manteniment predictiu, mai havia entrat en conceptes teòrics relacionats amb la soldadura. Malgrat això, aquest estudi ha permès conèixer els conceptes principals.
- **Aprendre com s'està aplicant la intel·ligència artificial en aquesta àrea de coneixement.** L'elaboració de l'estat de l'art ha permès a l'estudiant conèixer quin ús s'està fent de la intel·ligència artificial per a automatitzar la detecció de defectes i millorar la qualitat de soldadura.

- **Aprendre sobre mètodes de codificació i *visual transformers*.** Tot i que l'estudiant tenia coneixements d'aquest tipus de mètodes codificació, aquest estudi ha permès profunditzar en aquests. A més, ha permès conèixer de manera detallada com funcionen els *visual transformers* i quins són els seus fonaments teòrics.
- **Aprendre com implementar un cas estudi real.** Tot i que l'estudiant té experiència implementant aquest tipus de cas d'estudi, mai havia trobat tanta dificultat en trobar un conjunt de dades que s'adaptés a la metodologia que volia implementar. Arran que normalment es fa ús d'imatges per a analitzar defectes de soldadura, els recursos que consideren sèries temporals són limitats. Malgrat això, finalment es va poder trobar un conjunt de dades que s'adaptés a les necessitats de l'estudi.

En relació amb l'assoliment dels objectius plantejats, a continuació, es mostra la reflexió crítica de l'estudiant:

Objectius específics primaris

- **Dur a terme una anàlisi crítica d'estudis existents. Objectiu assolit.** Tal i com es mostra al capítol 3. *Estat de l'art*, s'ha fet una anàlisi crítica dels estudis més recents i rellevants que s'han identificat a través de l'ús de les bases de dades de Scopus i IEEE. Un cop analitzats tots els estudis seleccionats, es van extreure les limitacions i avantatges d'aquests estudis a fi de poder-los alinear amb la necessitat detectada que s'ha considerat en aquest estudi.
- **Investigar la importància d'aplicar intel·ligència artificial per a la detecció de defectes de soldadura. Objectiu assolit.** Aquest punt està alineat amb el punt anterior, atès que l'elaboració del capítol de l'estat de l'art ha permès conèixer a l'estudiant les actuals pràctiques que es duen a terme en relació amb l'aplicació de la intel·ligència artificial per a la detecció de defectes de soldadura.
- **Estudiar la viabilitat d'implementar GAF, RP, MTF amb *image classifiers*. Objectiu assolit.** No només s'ha assolit, sinó que també s'ha implementat un altre mètode de codificació que no estava previst. Aquest és el MTM i es detalla a la part de metodologia. En la secció resultats s'han analitzat els resultats obtinguts a fi de determinar si aquest tipus de metodologies són viables o no per a la tasca a implementar.
- **Analitzar *vision transformer* en tasques de classificació. Objectiu assolit.** Similar que al punt anterior, no només assolit, sinó que també s'ha analitzat les FFNs com a potencial *image classifiers*.
- **Avaluar l'exactitud de la metodologia implementada a través d'un cas d'estudi. Objectiu assolit.** S'ha utilitzat un cas d'estudi relacionat amb la qualitat de la soldadura. A més, la metodologia s'ha avaluat en termes d'exactitud i temps d'entrenament en dues fases. La primera fase fa referència a l'anàlisi dels resultats segons la metodologia de codificació emprada i, la segona, fa referència a l'anàlisi dels resultats de l'estudi comparatiu implementat.

- **Extreure conclusions a partir dels resultats obtinguts del cas d'estudi implementat. Objectiu assolit.** Aquestes conclusions s'introdueixen al capítol 6. *Discussió* i es finalitzen en el capítol 7. *Conclusions*.

Objectius específics secundaris

- **Combinar tècniques de *time series imaging* amb *vision transformers*. Objectiu assolit.** Aquesta combinació fa referència a la metodologia resultant que s'ha introduït en aquest estudi.
- **Efectuar un estudi comparatiu amb altres xarxes neuronals més convencionals. Objectiu assolit.** Inicialment es contemplava fer l'estudi comparatiu amb LSTM, únicament. Malgrat això, s'ha finalitzat fent un estudi comparatiu més elaborat al contemplar les FFNs, les LSTMs i les CNNs.
- **Establir línies de treball futur per a millorar les actuals estratègies de detecció automàtica de defectes de soldadura. Objectiu assolit.** Aquestes es presenten en la secció 7.2. *Línies de futur* d'aquesta memòria.

7.2. Línies de futur

A continuació, s'indiquen les línies de futur principals que no s'han pogut analitzar en aquest estudi:

- **Combinació de tècniques de codificació d'imatges.** En aquest estudi s'han analitzat els mètodes de codificació de manera independent. Malgrat això, és possible que una combinació d'aquestes permeti una millora en el rendiment del model.
- **Optimització dels hiperparàmetres de les xarxes neuronals.** En aquest estudi s'ha aplicat avaluació heurística per a determinar els diferents hiperparàmetres dels models. Malgrat això, és essencial emprar algoritmes d'optimització d'hiperparàmetres, com ara el *Particle Swarm Optimization* i el *Whale Optimisation Algorithm*.
- **Elaboració d'un model multimodal.** En aquest cas únicament s'ha fet ús d'un conjunt de dades d'una única font de recol·lecció. A més, únicament s'han tractat sèries temporals. Malgrat això, l'estudiant creu que seria útil fer ús també d'imatges de soldadura i crear un model multimodal que permeti elaborar la tasca mitjançant la consideració de diverses fonts.

7.3. Seguiment de la planificació

A mode general, l'estudiant considera que la planificació i la metodologia que s'ha dut a terme és la correcta. A més, l'estudiant considera que ha seguit la

planificació que va definir a l'inici d'aquest estudi. Com a possible millora, s'indiquen els següents aspectes:

- El temps total que l'estudiant va considerar per a l'elaboració de la memòria va ser incorrecte arran el desconeixement que aquest tenia de les entregues. L'estudiant es pensava que tenia fins a la PAC5b per a entregar la memòria. En canvi, aquesta va haver de ser entregada a la PAC4.
- Es va decidir ampliar el cas d'estudi i l'estudi comparatiu, la qual cosa va implicar un retard en l'entrega dels resultats. Malgrat això, l'estudiant considera que aquest canvi va ser el correcte a fi d'obtenir uns resultats més robustos que milloressin la qualitat de la memòria.

8. Glossari

¹**Soldadura.** Procés de fabricació que uneix metalls a través de la fusió per calor o pressió.

²**YOLOv4.** *You Only Look Once v4* (YOLOv4) és un algoritme d'aprenentatge profund per a la detecció d'objectes en temps real.

³**Cisalla per tracció.** Màquina utilitzada per a determinar la resistència d'un material a les forces de tracció.

⁴**Vision Transformer (ViT).** Model d'aprenentatge profund que utilitza els *transformers* per a dur a terme tasques de visió per computador.

⁵**Time Series Imaging (TSI).** Tipus de models que presenten com a objectiu principal la codificació de sèries temporals en imatges.

⁶**Gramian Angular Field (GAF).** Tipus de TSI que utilitza operacions trigonomètriques per a la codificació de sèries temporals en imatges.

⁷**Recurrence Plot (RP).** Tipus de TSI que estima trajectòries entre instàncies i la *pairwise distance* entre les trajectòries resultants per a codificar les sèries temporals en imatges.

⁸**Markov Transition Field (MTF).** Tipus de TSI que determina la matriu de transició de Markov per a codificar les sèries temporals en imatges.

⁹**Image classifiers.** Models que efectuen tasques de classificació mitjançant la consideració d'imatges com a variables d'entrada.

¹⁰**Long Short-Term Neural Network (LSTM).** Tipus de xarxa neuronal recurrent que considera les dependències temporals de les seqüències d'entrada.

¹¹**Convolutional Neural Network (CNN).** Tipus de xarxa neuronal comunament utilitzades per a l'anàlisi d'imatges.

¹²**Markov Transition Matrix (MTM).** Similar al MTF, és un tipus de TSI que determina la matriu de transició seguint les cadenes de Markov de primer ordre per a codificar sèries temporals en imatges.

¹³**Joint Recurrence Plot (JRP).** Tipus de TSI que considera el producte de *Hadamard* de RPs a fi d'efectuar una anàlisi multivariant.

¹⁴**Gramian Angular Summation Field (GASF).** Tipus de GAF que representa una sèrie temporal en imatge mitjançant la consideració de sumes angulars d'instàncies successives.

¹⁵**Gramian Angular Difference Field (GADF).** Tipus de GAF que representa una sèrie temporal en imatge mitjançant la consideració de diferències angulars d'instàncies successives.

9. Bibliografia

- Amirafshari, P., Barltrop, N., Wright, M., Kolios, A., 2021. Weld defect frequency, size statistics and probabilistic models for ship structures. *Int J Fatigue* 145, 106069. <https://doi.org/10.1016/j.ijfatigue.2020.106069>
- Asadi, R., Queguineur, A., Wiikinkoski, O., Mokhtarian, H., Aihkisalo, T., Revuelta, A., Ituarte, I.F., 2024. Process monitoring by deep neural networks in directed energy deposition: CNN-based detection, segmentation, and statistical analysis of melt pools. *Robot Comput Integr Manuf* 87, 102710. <https://doi.org/10.1016/j.rcim.2023.102710>
- Chen, T., Philippi, I., Phan, Q.B., Nguyen, L., Bui, N.T., daCunha, C., Nguyen, T.T., 2024. A vision transformer machine learning model for COVID-19 diagnosis using chest X-ray images. *Healthcare Analytics* 5, 100332. <https://doi.org/10.1016/j.health.2024.100332>
- Dorbane, A., Harrou, F., Sun, Y., 2022. Forecasting FSW Material's Behavior using an Artificial Intelligence-Driven Approach, in: 2022 International Conference on Decision Aid Sciences and Applications (DASA). IEEE, pp. 1553–1557. <https://doi.org/10.1109/DASA54658.2022.9765072>
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N., 2021. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, in: International Conference on Learning Representation. pp. 1–22.
- Fahim, S.R., Sarker, S.K., Mueen, S.M., Sheikh, Md.R.I., Das, S.K., Simoes, M., 2021. A Robust Self-Attentive Capsule Network for Fault Diagnosis of Series-Compensated Transmission Line. *IEEE Transactions on Power Delivery* 36, 3846–3857. <https://doi.org/10.1109/TPWRD.2021.3049861>
- Hahn, Y., Maack, R., Buchholz, G., Purrio, M., Angerhausen, M., Tercan, H., Meisen, T., 2023. Towards a Deep Learning-based Online Quality Prediction System for Welding Processes. *Procedia CIRP* 120, 1047–1052. <https://doi.org/10.1016/j.procir.2023.09.123>
- Jager, M., Knoll, C., Hamprecht, F.A., 2008. Weakly Supervised Learning of a Classifier for Unusual Event Detection. *IEEE Transactions on Image Processing* 17, 1700–1708. <https://doi.org/10.1109/TIP.2008.2001043>
- Köhler, T., Schiele, M., Glaser, M., Schricker, K., Bergmann, J., Augsburg, K., 2020. In-situ monitoring of hybrid friction diffusion bonded EN AW 1050/EN CW 004A lap joints using artificial neural nets. *Proceedings of the Institution of Mechanical Engineers, Part L: Journal of Materials: Design and Applications* 234, 766–785. <https://doi.org/10.1177/1464420720912773>
- Liu, J., Cheng, Y., Jing, X., Liu, X., Chen, Y., 2024. Prediction and optimization method for welding quality of components in ship construction. *Sci Rep* 14, 9353. <https://doi.org/10.1038/s41598-024-59490-w>
- Raofi, H., Sabahnia, A., Barbeau, D., Motamedi, A., 2024. Deep Learning Method to Detect Missing Welds for Joist Assembly Line. *Applied System Innovation* 7, 16. <https://doi.org/10.3390/asi7010016>
- Schwarz, E.B., Bleier, F., Guenter, F., Mikut, R., Bergmann, J.P., 2022. Improving process monitoring of ultrasonic metal welding using classical machine learning methods and process-informed time series evaluation. *J Manuf Process* 77, 54–62. <https://doi.org/10.1016/j.jmapro.2022.02.057>

- Tyystjärvi, T., Fridolf, P., Rosell, A., Virkkunen, I., 2024. Deploying Machine Learning for Radiography of Aerospace Welds. *J Nondestr Eval* 43, 24. <https://doi.org/10.1007/s10921-023-01041-w>
- Zhang, Z., Wen, G., 2016. An easy method of image feature extraction for real-time welding defects detection, in: 2016 13th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI). IEEE, pp. 615–619. <https://doi.org/10.1109/URAI.2016.7625790>

Annexos

Configuració CNN

A1. Configuració de l'arquitectura de la CNN (les dimensions del *kernel* i del *pool* s'han adaptat en funció de si es tractava de 1D-CNN o 2D-CNN).

Input Shape	n_images x 40 x 40 x 2
Nombre de <i>convolutional layers</i>	2
Unitats de la <i>convolutional layer</i>	[32, 16]
Dimensions del <i>kernel</i>	3
Nombre de <i>pooling layers</i>	2
Tipus de <i>pooling layer</i>	Max
Dimensions del <i>pooling</i>	2
Nombre de <i>dense layers</i>	2
Unitats de les <i>dense layers</i>	[32, 16]
Dropout	0.5
Funció d'activació	ReLU
Funció d'activació de l'última capa	Softmax
Output Shape	2

Configuració FNN

A2. Configuració de l'arquitectura de la FNN..

Input Shape	n_senyals x 1
Nombre de <i>dense layers</i>	4
Unitats de les <i>dense layers</i>	[30, 15, 15, 15]
Aplicació de Batch Normalization	True
Funció d'activació de les <i>dense layers</i>	ReLU
Funció d'activació de l'última capa	Softmax
Output Shape	2

Configuració LSTM

A3. Configuració de l'arquitectura de la LSTM.

<i>Input Shape</i>	n_senyals x 2
Nombre de <i>LSTM layers</i>	3
Unitats de les <i>LSTM layers</i>	[64, 32, 16]
Funció d'activació de les <i>LSTM layers</i>	ReLU
Nombre de <i>dense layers</i>	2
Unitats de les <i>dense layers</i>	[32, 16]
Funció d'activació de les <i>dense layers</i>	ReLU
Funció d'activació de l'última capa	Softmax
Output Shape	2