INFORME FINAL PROJECTE MACHINE LEARNING INDIAN FOOD

Professor: Ramon Baldrich

Alumnes:
Pol Beltran Moncada- 1708836
Héctor Cervelló Navarro - 1708336
Oriol Ramos Navarro - 1708786
Pau Vidal Martín - 1671799

Grau en Enginyeria de Dades

Introducció

El desenvolupament d'un sistema de classificació d'imatges mitjançant tècniques de visió per computador és un procés complex que transcendeix l'aplicació directa d'algorismes estàndard. El present document recull i analitza els reptes tècnics addicionals identificats durant la implementació d'un sistema de reconeixement d'aliments basat en el paradigma Bag of Visual Words (BoVW), utilitzant descriptors SIFT i classificadors SVM. Aquests desafiaments, sorgits durant les diferents fases del projecte, han requerit una anàlisi sistemàtic, un diagnòstic precís i el disseny de solucions específiques que han permès optimitzar tant l'eficiència computacional com la qualitat dels resultats obtinguts.

Els treballs addicionals que es presenten s'articulen al voltant de sis àrees crítiques que han condicionat significativament el desenvolupament del projecte. En primer lloc, s'aborda la problemàtica de l'eficiència computacional i la gestió d'experiments de llarga durada, un aspecte fonamental quan es treballa amb volums considerables de dades visuals. En segon lloc, s'analitza la relació entre la densitat de descriptors i la qualitat del model, qüestionant la premissa intuïtiva que més informació comporta necessàriament millors resultats. En tercer lloc, s'examinen les limitacions intrínseques del sistema per discriminar entre classes visualment similars, identificant patrons d'error recurrents i les seves causes fonamentals. En quart lloc, es tracta la problemàtica de la generalització del model i l'escassetat de dades d'entrenament, un dels reptes més prevalents en l'aprenentatge automàtic aplicat. En cinquè lloc, s'explora una estratègia experimental de neteja del dataset basada en clustering, avaluant-ne la viabilitat i les condicions sota les quals aquesta tècnica podria resultar beneficiosa. Finalment, s'introdueix una estratègia alternativa basada en l'agrupació d'etiquetes i la creació de superclasses, plantejant una revisió crítica de la definició original de les classes com a possible via per reduir confusions estructurals i millorar l'estabilitat global del classificador.

Cadascun d'aquests reptes s'estructura seguint una metodologia sistemàtica que parteix de la identificació del problema observat, continua amb l'anàlisi de les causes subjacents, presenta les solucions implementades o proposades, i conclou amb una explicació detallada dels resultats experimentals i les implicacions tècniques. Aquest enfocament rigorós permet no només documentar els obstacles trobats, sinó també fonamentar les decisions tècniques adoptades i extreure'n aprenentatges transferibles a projectes futurs. L'objectiu és proporcionar una visió completa i transparent del procés de desenvolupament, posant de manifest que l'èxit d'un sistema de visió per computador no resideix únicament en l'elecció d'algorismes adequats, sinó en la capacitat d'identificar, diagnosticar i resoldre els múltiples desafiaments tècnics que emergeixen durant la seva implementació real.

Treballs Addicionals: Reptes, Diagnòstic i Solucions

1. Eficiència Computacional i Gestió d'Experiments

Ens vam trobar amb experiments de durada excessiva, amb alguns superiors a 7 hores d'execució, i això paral·lelitzant les execucions, repartint els mètodes extractors de descriptors entre diferents ordinadors. El repte més crític va ser un bloqueig en el procés de clustering: l'algorisme K-Means estàndard era completament inviable per al nostre cas d'ús.

Es van identificar dos colls d'ampolla principals. Primer, la manipulació d'imatges d'alta resolució estava alentint significativament el procés d'extracció de característiques. Segon, la complexitat computacional intrínseca del K-Means clàssic (que requereix calcular distàncies entre tots els punts i tots els centroides en cada iteració) no escalava adequadament amb els grans volums de dades del nostre projecte. A més, la durada extrema dels experiments posava en risc la integritat dels resultats davant de possibles fallades del sistema durant l'execució, fet que ja ens va passar en una nit d'execució.

Vam implementar una solució integral de tres components: (1) un sistema de checkpointing per a la persistència de resultats intermedis, (2) una funció de preprocessament de redimensionament d'imatges, i (3) una migració de l'algorisme estàndard a MiniBatchKMeans amb inicialització K-means++.

Explicació:

- Sistema de checkpointing: Es va desenvolupar un mecanisme que guarda incrementalment els resultats parcials en format JSON durant l'execució dels experiments. Això assegura la reproductibilitat dels resultats i protegeix contra la pèrdua d'informació en cas de fallades del sistema, permetent reprendre els càlculs des del darrer punt guardat. També vam fer servirt pickle per a la mateixa dinàmica.
- Optimització del preprocessament: Vam redimensionar totes les imatges a una resolució de 300×300 píxels abans de l'extracció de característiques. Aquesta modificació va reduir el temps d'extracció un 60% sense pèrdua apreciable de qualitat en els descriptors, accelerant significativament tot el pipeline i estandaritzant les mides de tot el dataset.
- Migració algorísmica: La substitució de Kmeans per MiniBatchKMeans va ser crítica: va reduir l'ús de RAM i el temps de càlcul a aproximadament 3 hores (naturalment, paral·lelitzant la feina). Analitzant, vam trobar per la documentació que aquest mètode comportava una pèrduda molt petita de precisió, pero el guany en velocitat de processament era enorme, i per tant vam decidir fer servir aquesta funció.

2. Qualitat dels Descriptors (Dense)

En un intent de millorar la precisió del model, es va realitzar un experiment dràstic augmentant la densitat dels descriptors. Es va passar d'un step_size estàndard de 20 píxels a un valor extrem de 5 píxels, incrementant així el nombre de keypoints detectats de manera substancial quan utilitzàvem el mètode extractor Dense. Contràriament a les expectatives, el rendiment del model no va millorar significativament malgrat l'augment massiu en la quantitat de descriptors extrets per imatge.

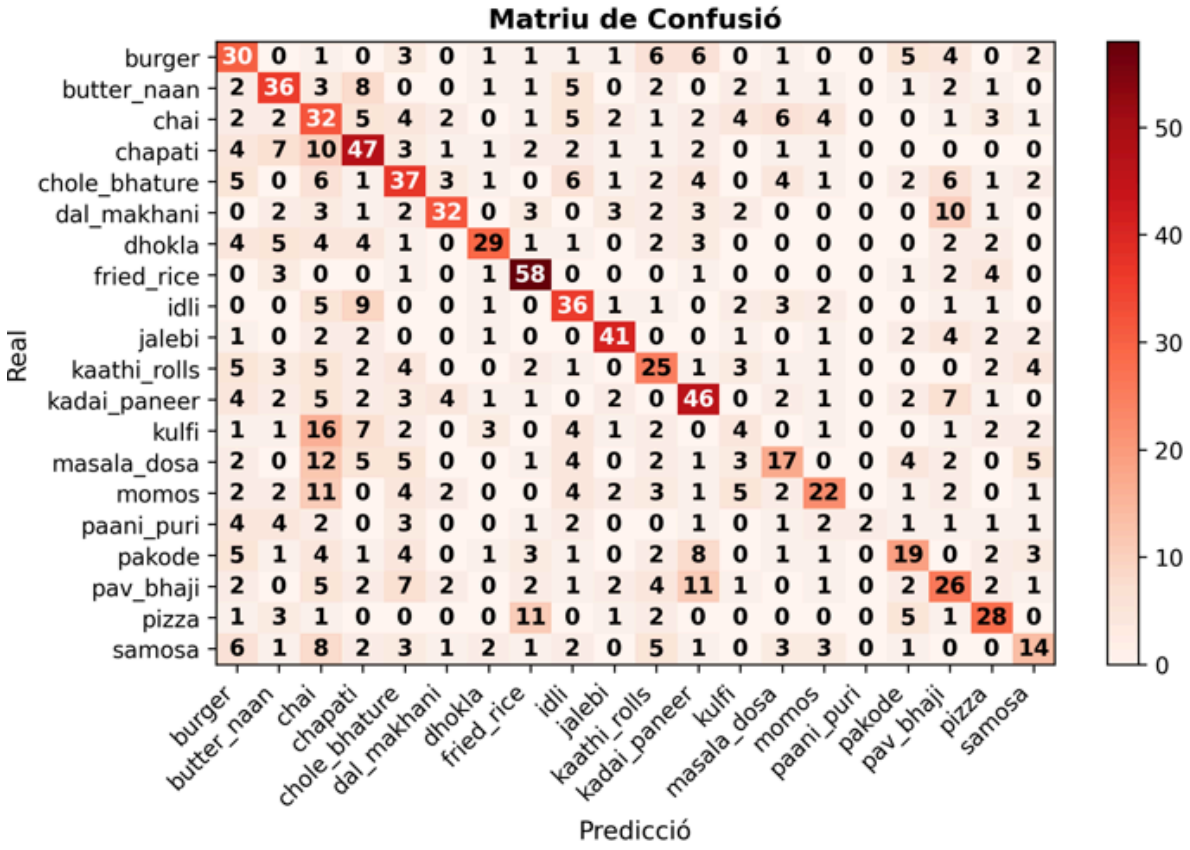
L'experiment va revelar un principi fonamental en feature extraction: la simple quantitat de dades no equival automàticament a millor informació. La qualitat i la rellevància dels keypoints per a la discriminació entre classes és més important que el seu nombre absolut. Un excés de descriptors densament agrupats pot introduir redundància sense aportar informació discriminativa addicional.

Després d'analitzar els resultats de l'experiment, es va optar estratègicament per mantenir la configuració estàndard (step_size=20), evitant així la càrrega computacional innecessària i significativa que implicava la configuració extrema (step_size=5) sense obtenir beneficis palpables en el rendiment.

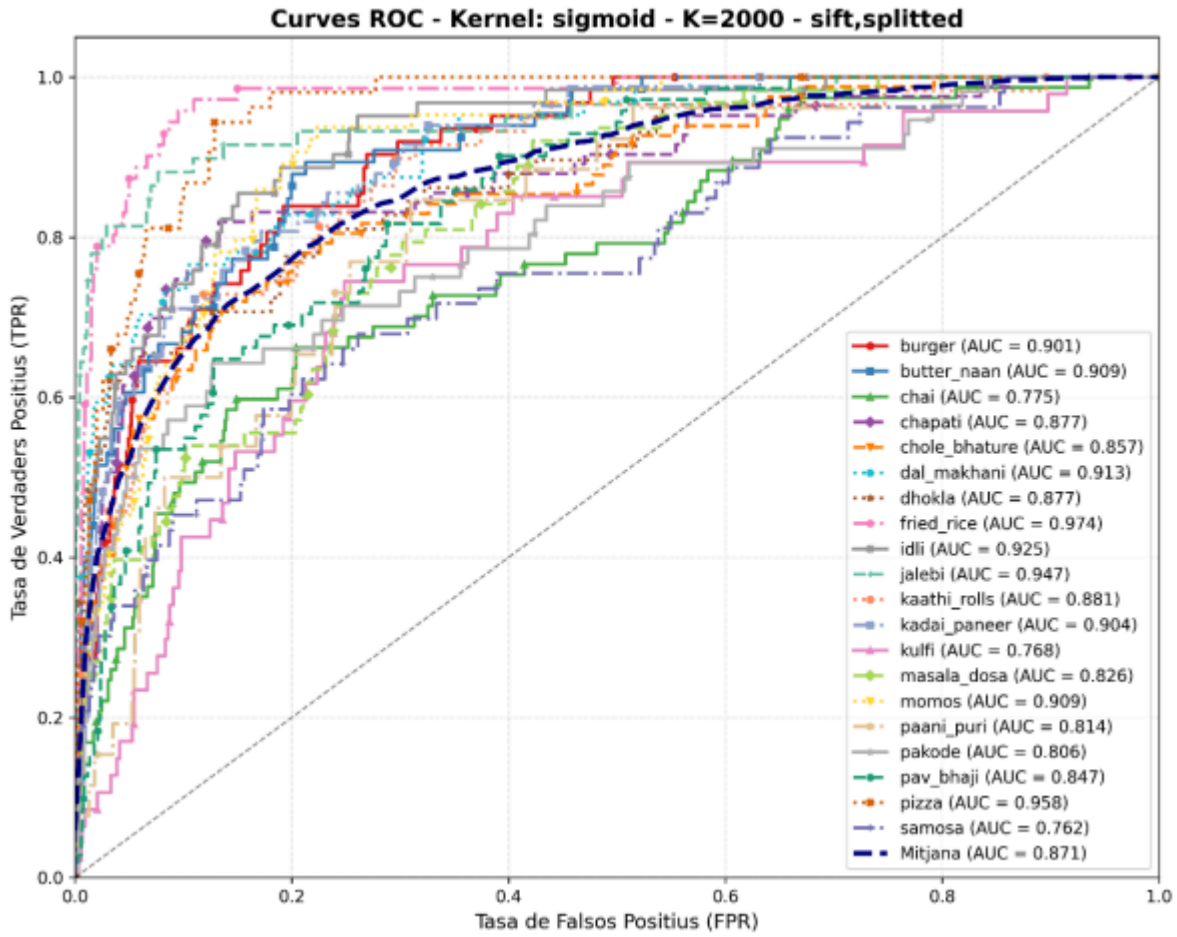
Explicació: Aquesta decisió confirma que el coll d'ampolla en el rendiment del nostre model no era la falta de descriptors o la manca de cobertura de les imatges, sinó la capacitat discriminativa intrínseca per diferenciar entre les classes del nostre problema específic. L'increment de densitat simplement va augmentar el cost computacional (temps d'extracció, memòria per emmagatzemar descriptors, temps de clustering) sense proporcionar informació nova rellevant per a la classificació.

3. Discriminació de Classes i Anàlisi d'Errors

El model presentava taxes d'error elevades i heterogènies en certes categories alimentàries. Les classes "paani puri" i “kulfi” mostraven el pitjor rendiment de totes les classes, i s'observaven patrons de confusions recurrents entre diferents plats en la matriu de confusió.



L'anàlisi detallada de la matriu de confusió i l'estudi de les corbes ROC individualitzades per classe van revelar dues causes principals. Primera, la similitud visual extrema entre determinats plats constitueix un repte fonamental per a qualsevol sistema de classificació basat en característiques visuals. Segona, la classe "Paani Puri" patia addicionalment d'escassetat de dades en el conjunt de test, disposant únicament de 26 imatges, una quantitat significativament inferior a altres classes i insuficient per avaluar robustament el rendiment.



Vam implementar una anàlisi exhaustiva multi-nivell: es van generar corbes ROC individualitzades per a cada classe per avaluar la capacitat discriminativa del model en cada categoria específica, i es va analitzar detalladament la matriu de confusió per mapar sistemàticament els patrons de fallada i identificar els parells de classes problemàtics.

Explicació:

- Classes amb bon rendiment: Les classes visualment distintives i amb característiques úniques (com pizza, burger) presenten un AUC (Area Under the ROC Curve) alt, indicant que són fàcils de classificar per al model gràcies a les seves propietats visuals clarament diferenciades.
- Patrons de confusió lògics: Les fallades del model no són aleatòries, sinó que es concentren en parells de classes amb sentit des del punt de vista visual i conceptual. Específicament segons la matriu de confusió:

1.Kulfi →chai (16 confusions)

Els dos tenen tonalitats marró-beix semblants, i pot generar confusió al passar-les al model a baixa resolució. El kulfi a moltes fotos surt servit en got/bol i al tenir una textura cremosa pot semblar una beguda espessa , i també porta condiments o cobertures (cardamom, pistatxos) pot confondre's amb les espècies del chai, pot comportar dificultat a la distinció per a algoritmes SIFT.

Kulfi <-> Chai

2.Pav_bhaji →dal_makhani (10 confusions)

ELs dos plats son purés espessos amb textura cremosa i amb una base de curry, tenen un color marró-vermellós. La mantega fosa a la superfície i els components vegetals triturats creen patrons visuals molt semblants que poden crear confusió, especialment quan es presenten en bols similars. A part la diferencia real és el gra de llegum, i això és difícil de captar pel model.

Pav_bhaji <-> Dal_makhani


3.Pav_bhaji →kadai_paneer (11 confusions)
La similitud principal ve donada que utilitzen una salsa vermella-taronjada molt semblant, la textura espessa i cremosa, i la presència de components vegetals, els trossos de verdures poden semblar-se als components del pav bhaji triturat, complicant l'extracció de característiques distintives per SIFT.

Pav_bhakji <-> Kadai_paneer


4.Chai →masala_dosa (12 confusions)
La confusió pot venir donada per el to del farcit marró-daurat uniforme similars al chai. En fotos llunyanes i redimensionades la dosa pot ocupar tot el frame, semblant una tassa de chai vista desde dalt. El problema es que es poden generar descriptors SIFT similars als patrons d'aquestes classes. Sobretot destacar la quantitat de marró que hi ha, tot molt monotòn. Els 2 menjars en si son marrons.

Chai <-> Masala_dosa


5.Chai →momos (11 confusions)

Els momos presenten una superfície blanquinosa-beix que pot recordar el chai amb abundant llet. La forma arrodonida dels momos i la presència de condensació o salses pot crear reflexos i textures que s'assemblen als patrons visuals del líquid del chai, a part les imatges del momos poden generar poc contrast ja que el chai sovint també està en entorns clars, i el model confon objectes petits clars sobre un fons neutre. A més, sovint el momos es presentat amb la taça de salsa com es veu en la foto, en la que la podria confondre.

Chai <-> Momos


6.Chapati → Chai (10 confusions)

El chapati normalment és d'un color beix clar/marró suau semblant al color del chai amb llet. En moltes imatges, el chai apareix en tasses sobre plats clars, creant superfícies circulars planes semblants a un chapati. El model pot relacionar el color que tenen en comú i la forma que comparteixen.

Chapati <-> Chai


7.Butter_naan → Chapati (8 confusions)
Esparàvem que aquestes dues classes tinguessin més confusió ja que visualment son casi idèntics, l'ùnica diferencia real és el gruix, cosa que li resulta complicat al model classificar plats semblants només en funció del gruix, ja que té en compte més descriptors.

Butter_naan <-> Chapati


8.Pizza → Fried_rice (11 confusions)
Els dos plats presenten ingredients dispersos i la pizza pot semblar un plat amb textures irregulars, com l'arròs fregit. I al redimensionar les imatges el model potser no les identifica amb molta claredat i les confon.

Pizza <-> Fried_rice


Aquesta anàlisi demostra que els errors del model són sistemàtics i relacionats amb les limitacions intrínseques de les característiques visuals per diferenciar plats amb aparences similars. Recalcar, com s'ha dit abans, que hi ha una predominancia del color marró generalitzada sovint a bastants plats, en la qual el extractor en mode splitted que fem doncs no acaba de treure tot el seu potencial.

4. Generalització i Escassetat de Dades

Hi havia un risc clar d'overfitting (sobreajustament) en el model entrenat, i una limitació significativa en la quantitat d'imatges d'entrenament disponibles que impedia que el model aprengués patrons generalitzables.

L'avaluació amb el kernel RBF va revelar un problema greu de generalització. La diferència (gap) entre l'accuracy en train (0.9141, és a dir, 91.41%) i test (0.4758, és a dir, 47.58%) era excessivament alta: 0.4383 (43.83 punts percentuals). Aquest gap tan ampli indica clarament que el model estava memoritzant els exemples d'entrenament en lloc d'aprendre patrons generals aplicables a dades noves, una evidència inequívoca d'overfitting.

Vam realitzar una avaluació comparativa de diferents configuracions de kernels, comparant específicament RBF versus Sigmoide per entendre millor el comportament de generalització. Addicionalment, vam dissenyar una estratègia completa de Data Augmentation per combatre l'escassetat de dades, encara que aquesta va quedar com a descartada tal com el professor ens va indicar a la reunió.

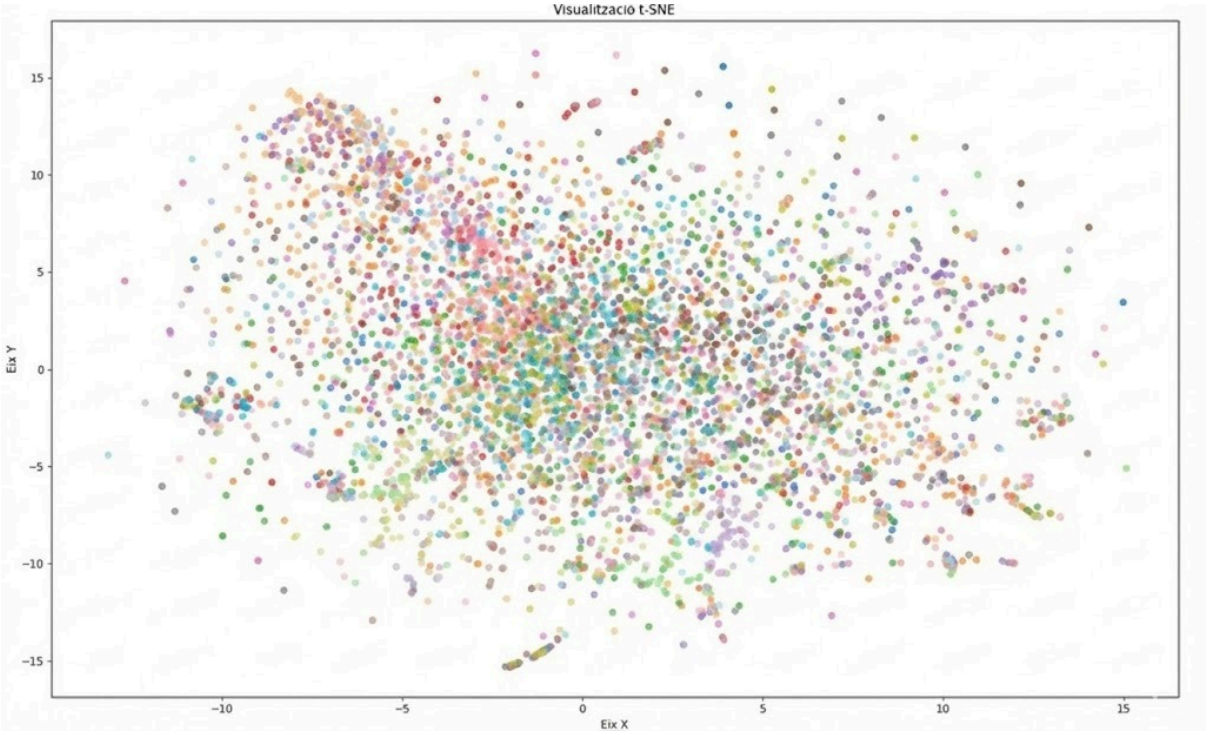
Explicació:

- Comparació de kernels: El kernel Sigmoide amb paràmetre $C=1.5$ va demostrar una capacitat de generalització significativament millor que el kernel RBF. Malgrat tenir un accuracy absolut en test lleugerament inferior al millor RBF, el gap entre train i test es va reduir dràsticament a només 0.2109 (21.09 punts percentuals), gairebé la meitat del gap observat amb RBF. Això indica que el kernel Sigmoide estava aprenent patrons més generalitzables i menys específics del conjunt d'entrenament.
- Estratègia de Data Augmentation (descartada): Es va dissenyar una proposta detallada d'augmentation que incloïa: operacions de zoom (variacions d'escala), transformacions de mirall horitzontal, rotacions lleugeres, i perturbacions de color (ajustos de brillantor, contrast i saturació). Aquesta estratègia hauria permès generar entre 10 i 20 variants per cada imatge original, multiplicant efectivament el dataset i proporcionant al model una major diversitat d'exemples per aprendre.

5. Neteja del Dataset i Selecció d'Imatges Òptimes

El model presentava dificultats en l'establiment de fronteres de decisió clares, i es va plantejar la hipòtesi que algunes imatges del dataset podrien estar introduint soroll i dificultant la classificació en lloc de contribuir positivament a l'aprenentatge.

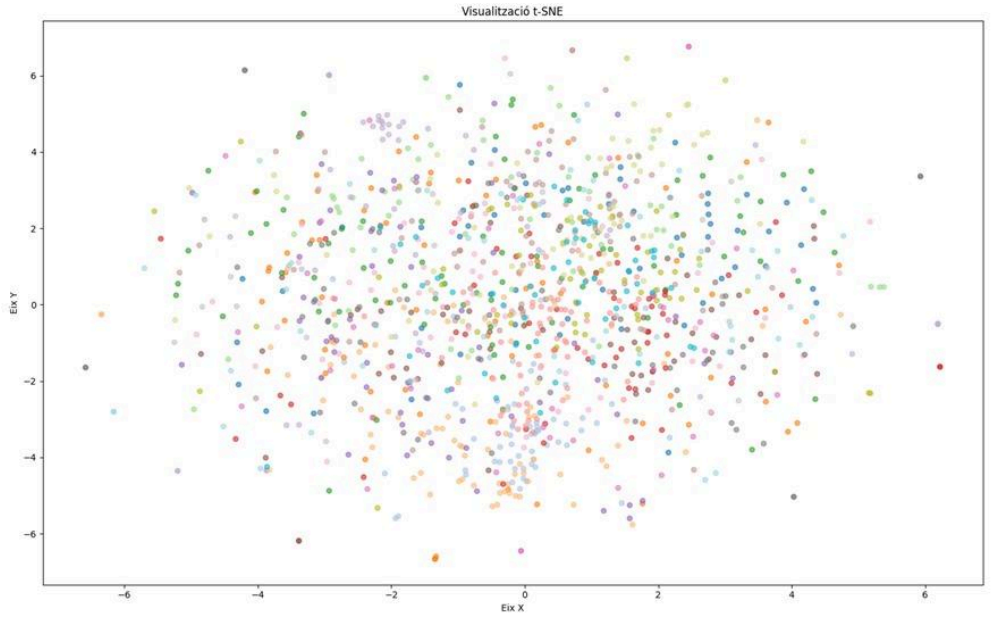
L'anàlisi de visualització de les dades en l'espai de característiques va revelar una manca de diferenciació clara entre classes, suggerint la presència d'imatges amb baixa qualitat representativa o outliers que distorsionaven les fronteres de decisió. La hipòtesi era que existien imatges en el dataset que "molestaven" al procés d'entrenament, creant fronteres dolentes. Aquesta va ser la visualització aplicant T-sne:



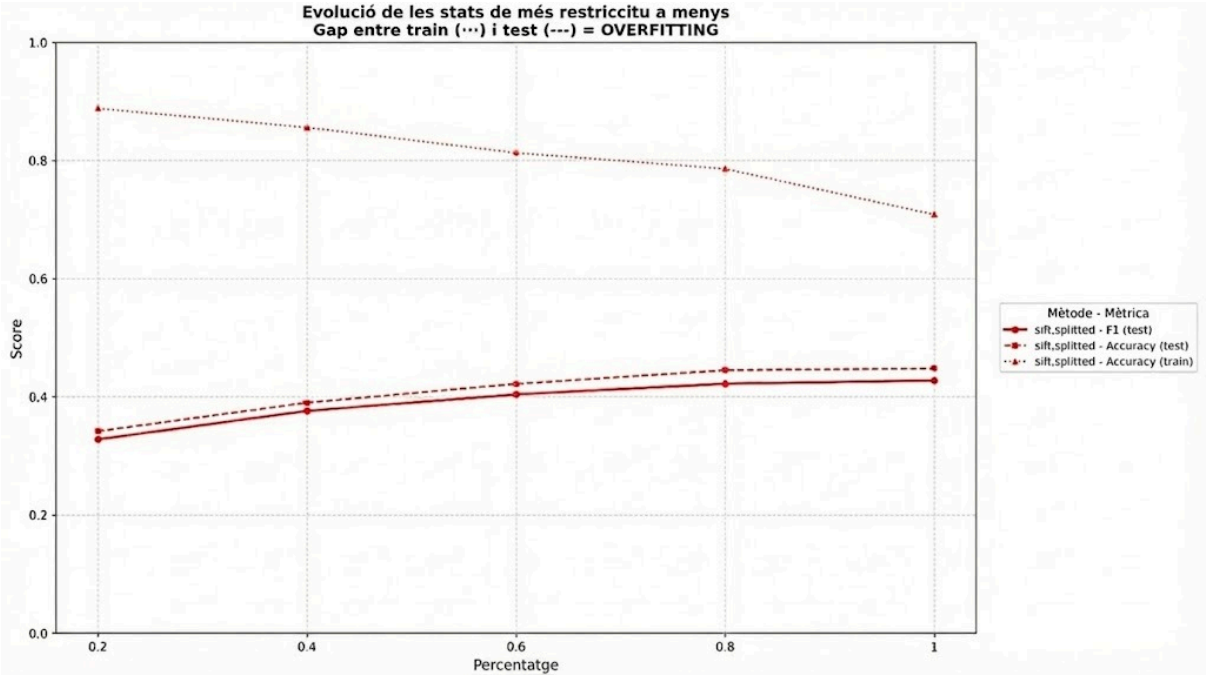
Vam implementar un sistema de neteja basat en clustering utilitzant K-means amb K=1 aplicat de manera independent a cada classe. La metodologia seguia tres passos principals: (1) calcular el centroide de cada classe en l'espai de característiques, (2) computar la distància de totes les imatges de la classe respecte del seu centroide corresponent, i (3) seleccionar únicament un determinat percentatge de les imatges més properes al centroide, considerades les més representatives. Es van dur a terme diversos experiments amb diferents percentatges de retenció, des del 20% fins al 100%, amb l'objectiu d'identificar el punt òptim.

Explicació:

- Visualització inicial: Les gràfiques de dispersió de les classes sense cap tipus de filtratge mostraven distribucions amb un solapament significatiu i absència d'agrupacions nítides, fet que confirmava la manca d'una diferenciació clara entre etiquetes i anticipava la dificultat d'establir fronteres de decisió robustes.
- Experiment amb filtratge restrictiu: Quan es va aplicar un filtratge estricte, mantenint únicament el 20% de les imatges més properes al centroide de cada classe, les visualitzacions obtingudes no van mostrar una millora significativa en la nitidesa ni en la separació dels grups.
- Resultats de l'experiment: L'anàlisi de l'evolució de l'accuracy per a diferents percentatges de retenció va revelar un patró clarament contraproduent: com més restrictius érem —és a dir, com menys imatges es retenien—, pitjor era el rendiment del model. Aquest comportament indica que la reducció del nombre d'imatges d'entrenament penalitzava el model en major mesura del que no pas el beneficiava.
- Interpretació i conclusions: Malgrat que la hipòtesi de la presència de soroll en el dataset fos probablement correcta, la reducció dràstica del volum de dades d'entrenament va resultar més perjudicial que beneficiosa. Es va determinar que, en un escenari d'escassetat de dades, mantenir el dataset complet era una opció preferible. En conseqüència, es pot afirmar que les imatges no presenten una separabilitat clara entre etiquetes, fet que limita significativament la capacitat de classificació del model. Així, fins i tot després d'eliminar possibles outliers i conservar únicament el 20% de les imatges més properes al centroide de cada classe, la visualització continuava mostrant una distribució caòtica i sense estructures diferenciades clares.



Evolució de les estadístiques a mida que borràvem menys soroll:



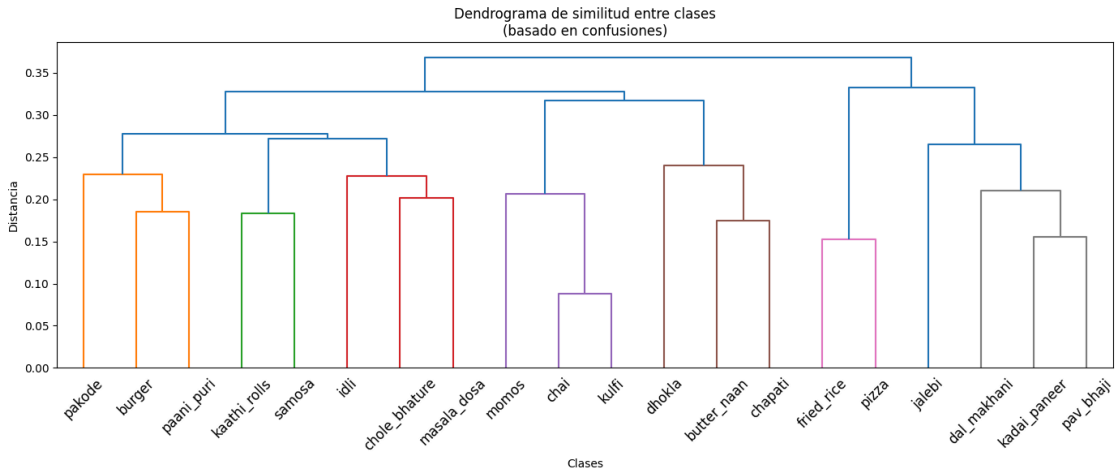
Podem observar, a més, que com més restrictius érem en la selecció d'imatges, pitjors resultats s'obtenien en el conjunt de test, mentre que el rendiment en train millorava. Aquest comportament indica clarament un fenomen d'overfitting. Així mateix, mitjançant la visualització del segon t-SNE utilitzant només el 20% de les imatges, es va constatar que les mostres de cada etiqueta apareixien fortament disperses, sense una agrupació clara ni una consistència interna apreciable.

6. Agrupació d'Etiquetes i Creació de Superclasses

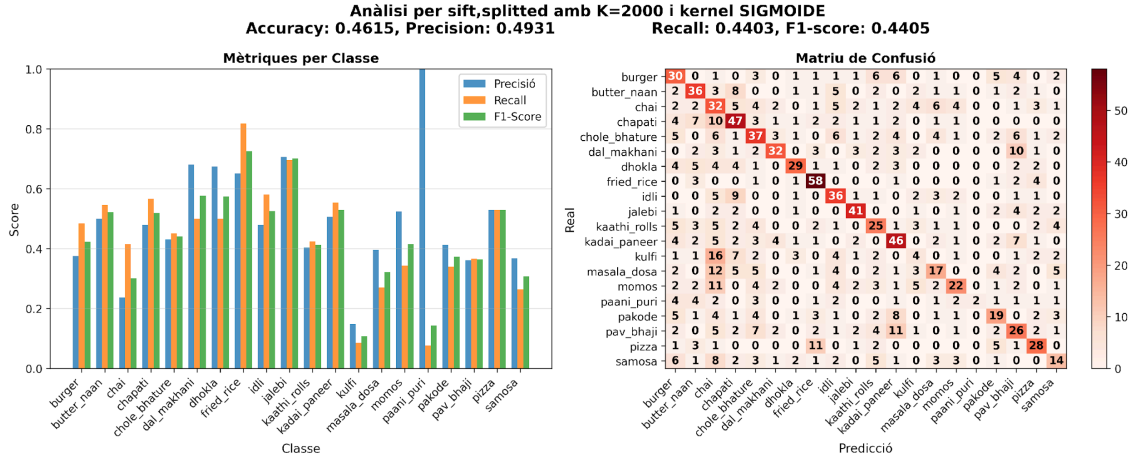
El model continuava presentant dificultats significatives per discriminar entre determinades classes visualment molt similars, fet que es traduïa en taxes d'error elevades i patrons de confusió persistents. Malgrat les optimitzacions prèvies aplicades sobre el pipeline, algunes classes mostraven un rendiment sistemàticament baix, indicant que el problema podia no residir únicament en el model o en les característiques, sinó en la pròpia definició de les etiquetes.

La causa principal identificada és que algunes classes del dataset presenten una similitud visual extrema, fins al punt que, fins i tot per a un observador humà, la seva distinció pot resultar ambigua en determinades imatges. Aquesta manca de separabilitat intrínseca en l'espai visual provoca que el model sigui incapaç d'establir fronteres de decisió clares, generant confusions recurrents que penalitzen el rendiment global, especialment en mètriques com el recall i l'F1-score.

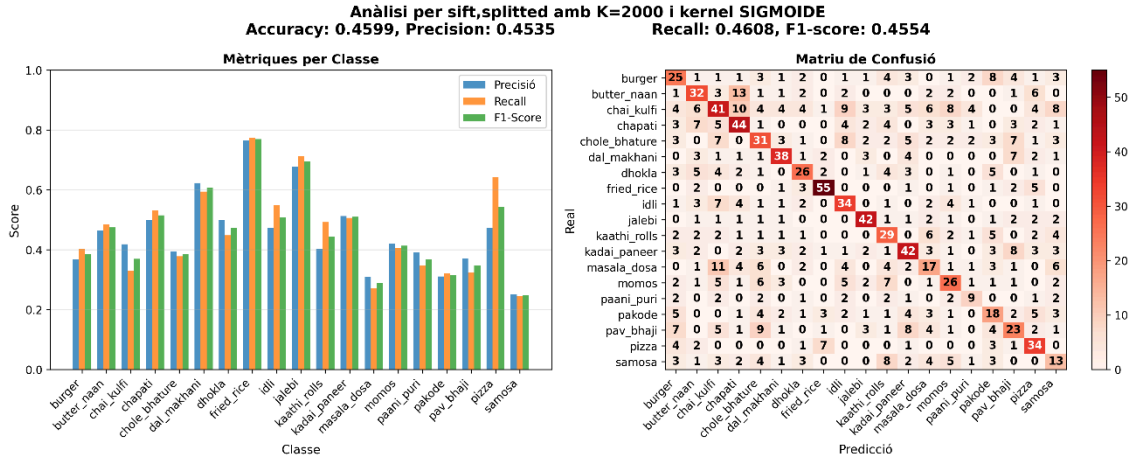
Com a estratègia alternativa, es va plantejar una nova hipòtesi: agrupar etiquetes visualment molt properes per crear superclasses més compactes. Tot i que inicialment aquesta idea havia estat descartada en la presentació del projecte, es va considerar que, des d'un punt de vista experimental, podria permetre obtenir fronteres de decisió més clares i un model globalment més estable. Per validar aquesta hipòtesi, es va construir una matriu de distàncies entre classes a partir de la matriu de confusió, utilitzant-la com a base per identificar quines etiquetes eren sistemàticament confoses entre si.



Abans d'agrupar:



Després d'agrupar:

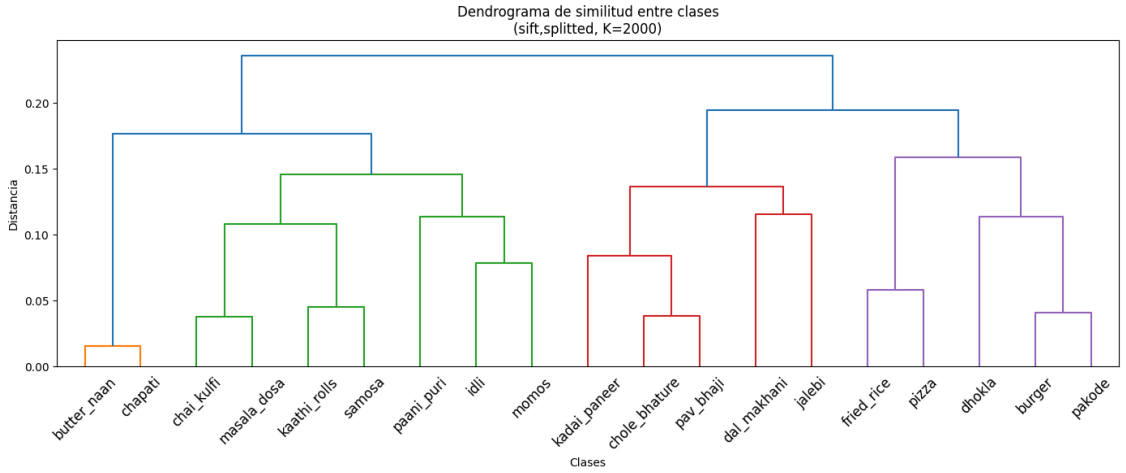


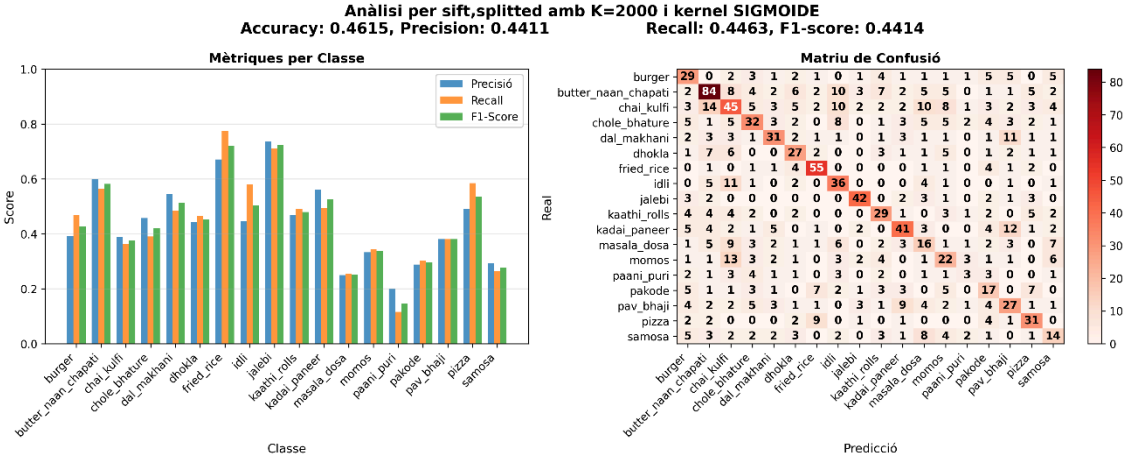
Explicació:

● Primera agrupació (chai + kulfi):

Per aprofundir en aquesta estratègia, es va generar un dendrograma a partir de les distàncies entre classes, que permet visualitzar jeràrquicament possibles agrupacions addicionals. L'anàlisi del primer dendrograma va revelar que les classes més properes eren *chai* i *kulfi*. Es va procedir a ajuntar-les en una única superclasse i es va tornar a entrenar el model. Els resultats van mostrar una reducció del valor extremadament alt de precisió que anteriorment presentava *paani_puri*, però, en contrapartida, va desaparèixer el mal rendiment associat a *kulfi*. En conjunt, el model es va mostrar més estable, amb una millora del recall i de l'F1-score, mètrica clau del projecte. Aquests resultats suggereixen que aquesta agrupació concreta permet reduir l'impacte de confusions estructurals i millorar el comportament global del classificador.

● Anàlisi jeràrquica amb dendrograma:
Per tant potser aquesta podia ser una bona solució. En estirar una mica més la corda, vàrem veure que el dendrograma que es generava amb aquesta agrupació suggeria agrupar *butter_naan* i *chapati*, dues classes que ja s'havia observat que eren visualment molt similars.





● Segona agrupació (butter_naan + chapati):
En aplicar aquesta nova agrupació, els resultats van empitjorar de manera generalitzada. Es va observar una davallada en les mètriques principals, indicant que, en aquest cas, la fusió d'etiquetes eliminava informació discriminativa rellevant i perjudicava la capacitat del model per diferenciar correctament entre altres classes. Això demostra que no totes les similituds visuals justifiquen una agrupació efectiva des del punt de vista de la classificació.

Aquest experiment posa de manifest que l'agrupació d'etiquetes pot ser una estratègia vàlida de manera puntual, però no generalitzable. En el cas de *chai* i *kulfi*, la creació d'una superclasse redueix confusions estructurals i millora l'estabilitat del model. En canvi, altres agrupacions aparentment lògiques, com *butter_naan* i *chapati*, resulten contraproduents. Per tant, aquesta tècnica ha de ser aplicada de manera selectiva, guiada per una anàlisi empírica rigorosa, i no com una solució universal als problemes de discriminació entre classes.

Tot i això, el model que vam fer nosaltres amb un predictor especialitzat de doble capa funciona millor que l'agrupament de les classes (que té un 46% d'accuracy).

Conclusió

Els treballs addicionals desenvolupats durant aquest projecte han proporcionat una comprensió profunda dels reptes inherents a la construcció de sistemes de classificació d'imatges basats en característiques visuals locals. L'anàlisi exhaustiva dels sis àmbits problemàtics identificats ha permès no només resoldre obstacles tècnics concrets, sinó també establir principis metodològics aplicables a projectes similars de visió per computador i aprenentatge automàtic.

En l'àmbit de l'eficiència computacional, la implementació d'un sistema integral de checkpointing, el preprocessament optimitzat d'imatges i la migració a MiniBatchKMeans amb inicialització intel·ligent han demostrat que la viabilitat pràctica d'un projecte depèn tant de la qualitat de les solucions algorísmiques com de la seva eficiència temporal i espacial. La reducció del temps d'execució a aproximadament 3 hores i l'eliminació del bloqueig de memòria han transformat un sistema inicialment inviable en una solució operativa i escalable. Pel que fa a la qualitat dels descriptors, l'experiment amb Dense ha evidenciat un principi fonamental: la redundància d'informació no equival a informació discriminativa rellevant, i l'augment de la densitat de keypoints sense criteris selectius només incrementa el cost computacional sense beneficis palpables en el rendiment.

L'anàlisi detallada de la discriminació de classes ha revelat que els errors del model no són aleatoris, sinó sistemàtics i coherents des del punt de vista de la similitud visual entre plats. Les confusions recurrents entre determinades parelles de classes reflecteixen les limitacions intrínseques dels descriptors SIFT per capturar diferències discriminatives en contextos de gran similitud cromàtica i textural. Aquesta constatació ha conduït a explorar estratègies alternatives que van més enllà de l'optimització del model, posant el focus en la pròpia definició de les etiquetes.

En aquest context, l'experiment d'agrupació d'etiquetes i creació de superclasses ha aportat una visió crítica i rellevant: en determinats casos, la reformulació del problema de classificació pot ser més efectiva que la introducció de noves tècniques o ajustos hiperparamètrics. Els resultats mostren que algunes agrupacions, com la fusió de classes visualment extremadament similars, poden reduir confusions estructurals i millorar la stabilitat global del model, mentre que altres agrupacions aparentment lògiques poden resultar contraproduents. Això demostra que aquesta estratègia no és universal, sinó que ha de ser aplicada de manera selectiva i guiada per una anàlisi empírica rigorosa.

Pel que fa a la problemàtica de generalització, la comparació entre kernels RBF i Sigmoide ha posat de manifest que l'elecció de la funció kernel no només afecta l'accuracy absolut, sinó també la capacitat del model per aprendre patrons generalitzables. El kernel Sigmoide, malgrat oferir un rendiment lleugerament inferior en termes absoluts, ha demostrat un gap train-test significativament menor, indicant una menor tendència a l'overfitting. Aquesta observació suggereix que, en contextos d'escassetat de dades, la prioritat hauria de ser la selecció d'arquitectures i hiperparàmetres que maximitzin la generalització, fins i tot a costa d'un lleuger sacrifici en l'accuracy d'entrenament.

Seguidament, l'experiment de neteja del dataset basat en clustering ha proporcionat una lliçó valuosa sobre l'equilibri entre qualitat i quantitat de dades. Tot i que la hipòtesi de presència de soroll en el dataset era plausible, la reducció dràstica del volum d'entrenament ha resultat més perjudicial que beneficiosa en el context actual d'escassetat de dades. No obstant això, aquesta estratègia romandria com una tècnica prometedora en escenaris futurs on es disposés d'un dataset substancialment més gran o es pogués aplicar data augmentation, permetent eliminar el soroll mantenint un volum suficient de dades representatives.

I, finalment, la prova d'avaluar si l'accuracy millorava mitjançant la creació de superclasses no va donar els resultats esperats. Ens trobem davant d'un dataset amb poca consistència interna i, en aquest context, l'estratègia de crear superclasses només podria tenir sentit si es reduís el problema a molt poques categories. A causa d'això, considerem que l'estratègia prèviament implementada amb el predictor de dues capes és més òptima que aquesta aproximació, ja que, en realitzar una segona iteració d'agrupació de les classes més properes, el rendiment del model comença a disminuir de manera clara.

En síntesi, els treballs addicionals han transformat obstacles aparentment insuperables en oportunitats d'aprenentatge i millora. Les solucions implementades han demostrat la importància d'un enfocament iteratiu i experimental en el desenvolupament de sistemes de machine learning, on la monitorització constant, l'anàlisi crítica dels resultats i la disposició a qüestionar assumpcions inicials són claus per a l'èxit. Les lliçons extretes d'aquest projecte —la primacia de l'eficiència, el valor de la qualitat sobre la quantitat en descriptors, la comprensió dels patrons d'error, la prioritat de la generalització, i l'equilibri entre neteja i volum de dades— constitueixen un corpus de coneixement tècnic directament transferible a futurs projectes de visió per computador i aprenentatge automàtic.