Tema1 IA

Manea Lidia-Elena - 341C4

30. November 2024

0.1 Flow-ul programului

Am inceput prin a citi datele si a le introduce in liste. Atat pentru setul de date cu fructe, cat si pentru cel mnist, am citit imaginile, le-am scalat si le-am introdus in liste. Am ales algoritmii de extragere de atribute pe baza rationamentului de mai jos. Am combinat subclasele de fructe in clase mai mari, iar apoi am facut graficele cu numarul de imgini ale claselor atat pe train, cat si pe test, respectiv graficele distributiei PCA. Am aplicat PCA cu 2, respectiv 3 componente pe ambele seturi de date si de asemenea am aplicat trasarea contururilor pentru a vizualiza rezultatele. Dupa toti acest pasi de vizualizare, am inceput sa scriu codul pentru preprocesare: am aplicat PCA cu 27 de componente pe imagini, rezultatele stocandu-le in vectori. Totodata, am aplicat hu moments pe imagini, iar apoi rezultatele de la PCA si hu moments le-am concatenat pentru a obtine vectorii finali. Dupa obtinerea acestor rezultate, am aplicat algoritmii cu diversi hiperparametri pentru a extrage cele mai bune optiuni.

0.2 Extragerea de atribute (Feature Extraction)

Pentru extragerea atributelor, am luat in considerare urmatoarele:
Pentru imagini referitoare la articole vestimentare, hainele sunt recunoscute in functie de conturul acestora, adica pantalonii difera prin rochii in primul rand prin forma acestora. Acest lucru se poate aplica si la fructe: o banana difera de ananas prin forma acestuia. De aceea, am ales sa utilizez hu moments pentru stabilirea conturului imaginilor. De asemenea, am ales sa utilizez PCA pentru ambele seturi de date pentru a extrage cele mai importante si distincte features din imagini.

0.3 Vizualizarea echilibrului intre clase

In setul de date cu fructe, exista un dezechilibru intre clase (de exemplu pentru clasa Cabbage White 1 sunt sub 200 de imagini, in timp ce pentru Pear Stone 1 sunt aproape 800 de imagini). Exista aceasta situatie atat pentru train, cat si pentru test la setul de imagini cu fructe. Dupa ce am combinat subclasele, am constatat ca inca exista un dezechilibru, diferand foarte mult prin numarul imaginilor. Pentru setul de date fashion mnist, clasele sunt perfect echilibrate, existand acelasi numar de imagini pentru fiecare clasa (cate 6000 de imagini pentru fiecare clasa train, cate 1000 de imagini pentru fiecare clasa test).

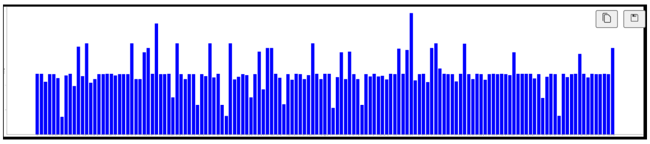

Abbildung 1: Folder train

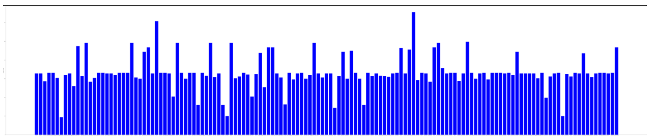

Abbildung 2: Folder train

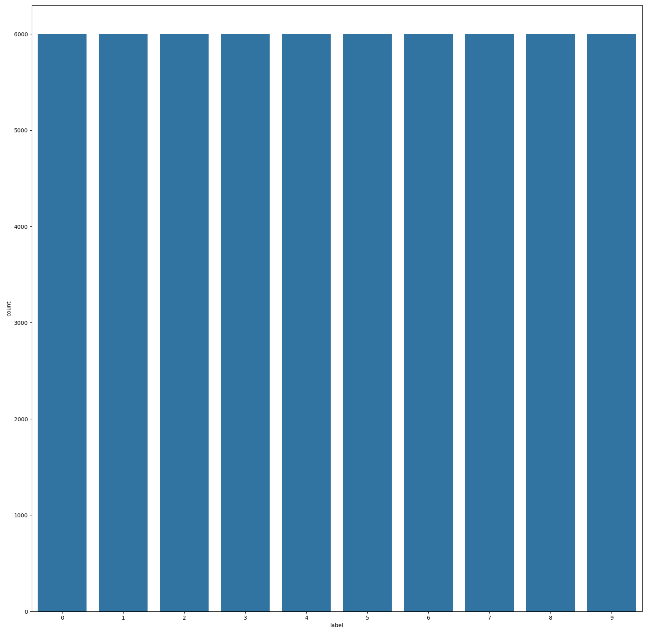

Abbildung 3: Folder train

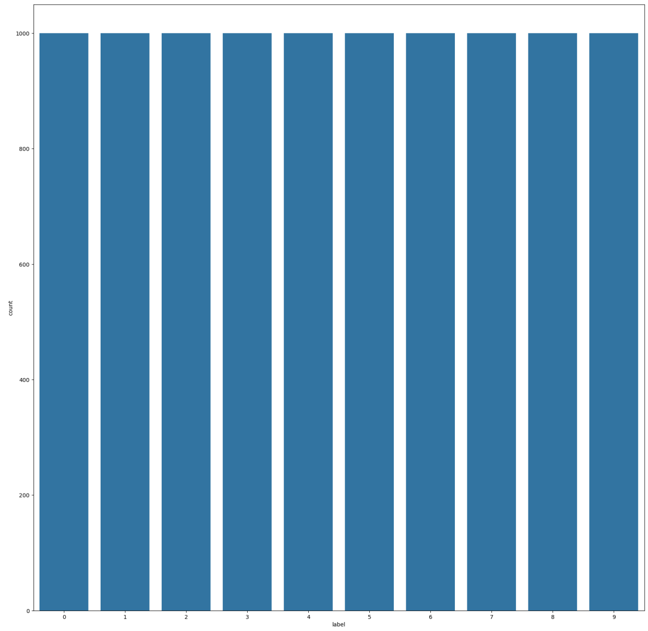

Abbildung 4: Folder train

0.4 Vizualizarea cantitativă a efectului de extragere a atributelor

Deoarece am aplicat PCA peste setul de date, am facut graficul de varianta cumulativa al componentelor PCA pe ambele seturi de date. Asadar, pentru ambele seturi de date, dupa ce am aplicat PCA pe 50 componente, observ ca varianta componentelor scade treptat, majoritatea fiind redundante. De aceea, am ales sa iau primle 27 de componente in ambele cazuri. 27 este un numar care acopera componente care sunt foarte utile, dar si din componentele redundante.

0.5 Vizualizarea calitativa a efectului de extragere a atributelor

Deoarece metoda mea de extragere a atributelor se bazeaza pe PCA si pe Hu moments, am aplicat PCA cu 27 componente pe diverse imagini si am aplicat si trasarea conturului pe un set de imagini. Imaginile rezultate, alaturi de labelurile lor, sunt prezente mai jos.
Se observa ca la PCA, 27 componente pe setul de ambele seturi de date sunt relativ suficiente pentru a distinge forma si culoarea obiectelor. De asemenea, conturul este foarte bine trasat, descoperind si particularitati ale fructelor (de exemplu contururi pe suprafata acestora).

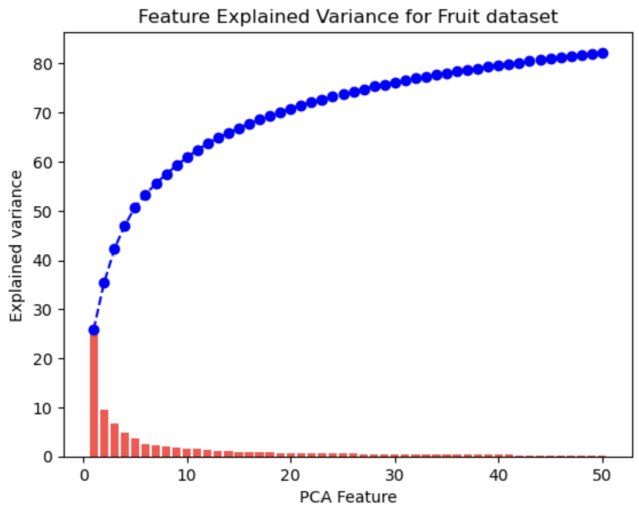

Abbildung 5: Folder train

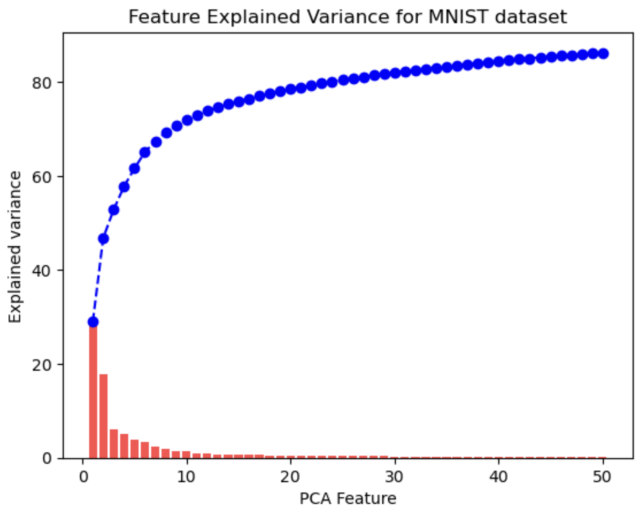

Abbildung 6: Folder train

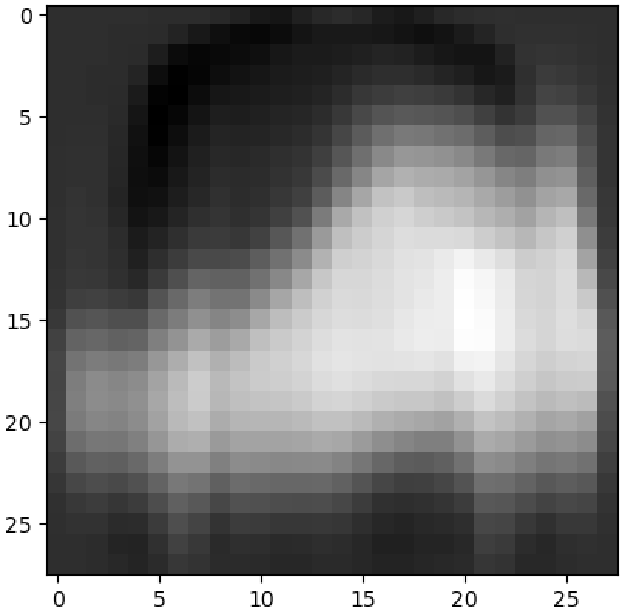

Abbildung 7: Ankle boots pca

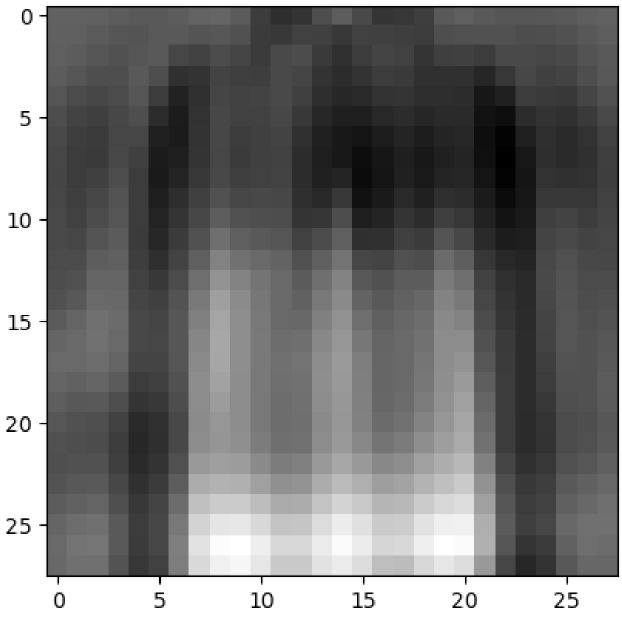

Abbildung 8: Bag pca

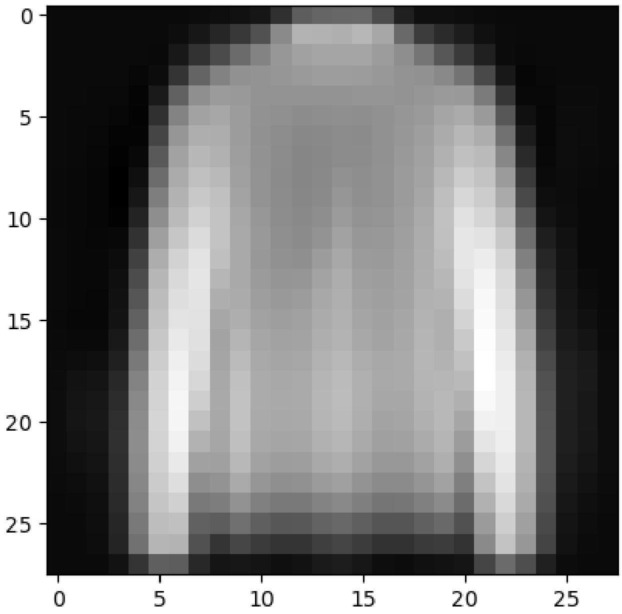

Abbildung 9: Coat pca

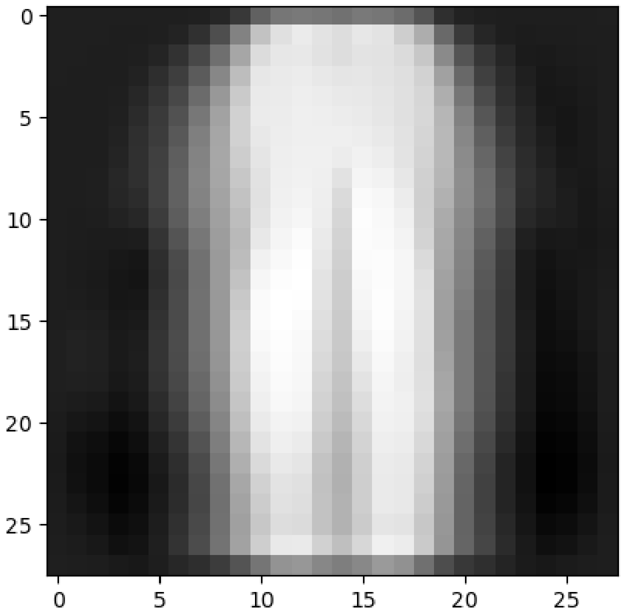

Abbildung 10: Dress pca

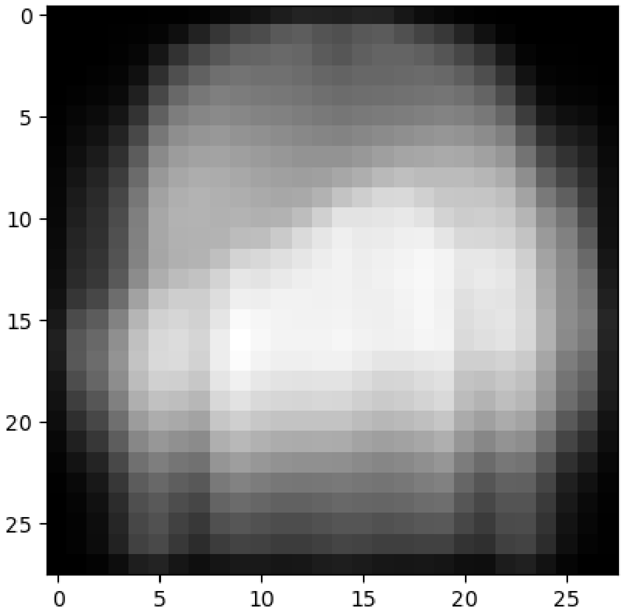

Abbildung 11: Pullover pca

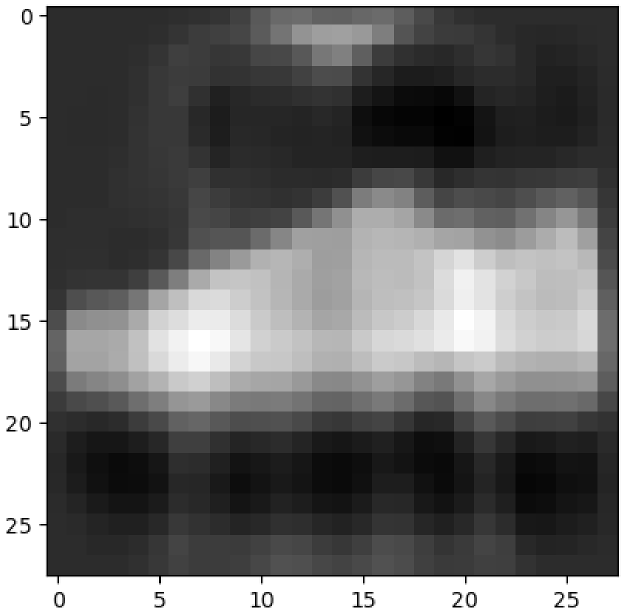

Abbildung 12: Sandal pca

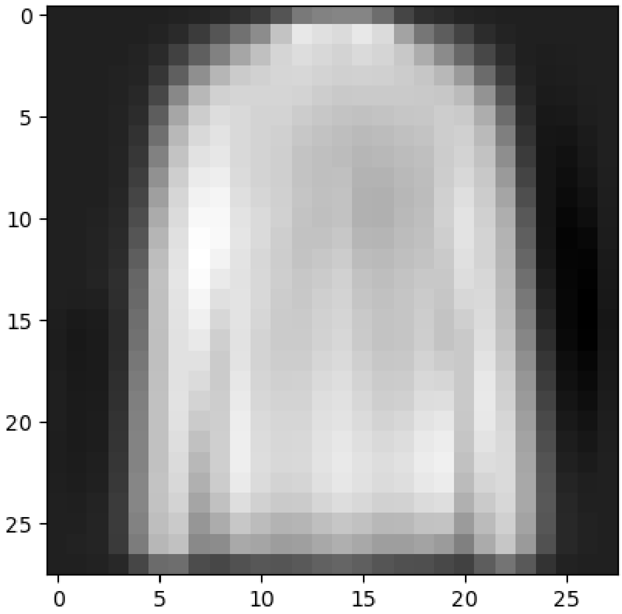

Abbildung 13: Shirt pca

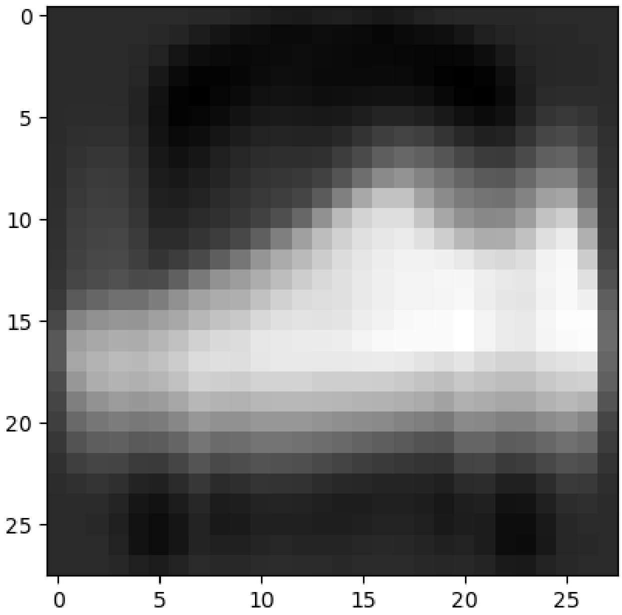

Abbildung 14: Sneaker pca

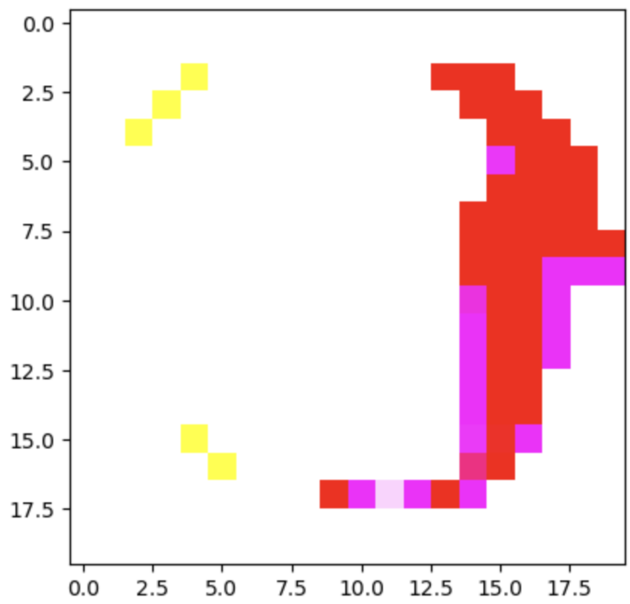

Abbildung 15: Apple pca

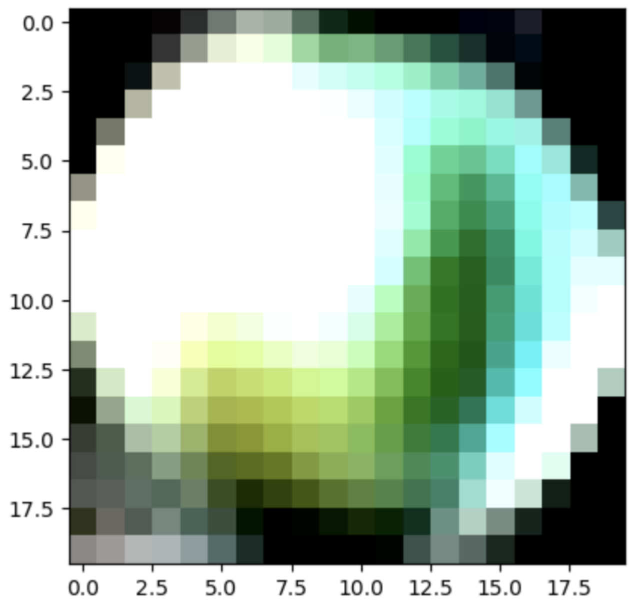

Abbildung 16: Banana pca

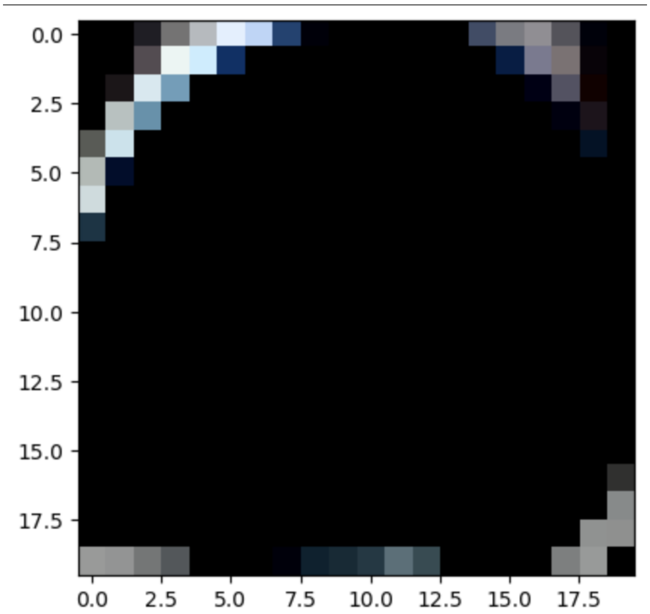

Abbildung 17: Cherry pca

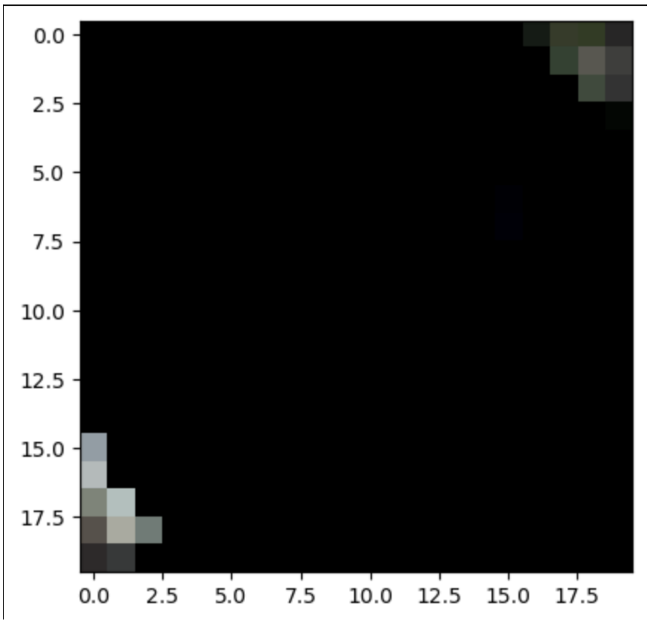

Abbildung 18: Grape pca

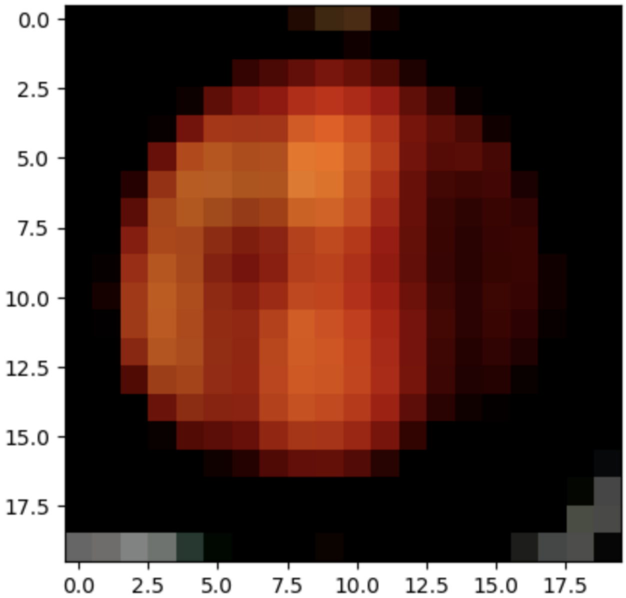

Abbildung 19: Peach pca

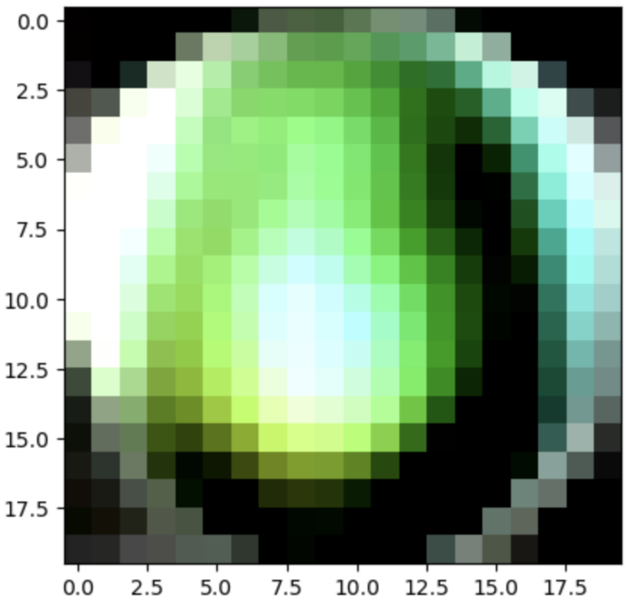

Abbildung 20: Pear pca

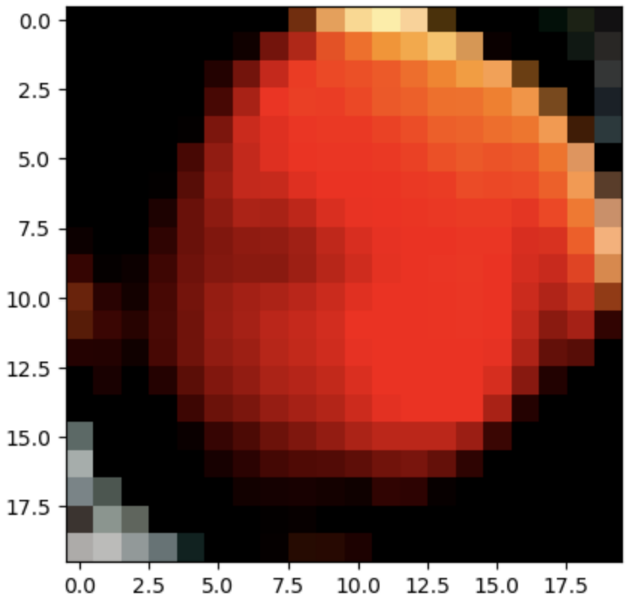

Abbildung 21: Pepper pca

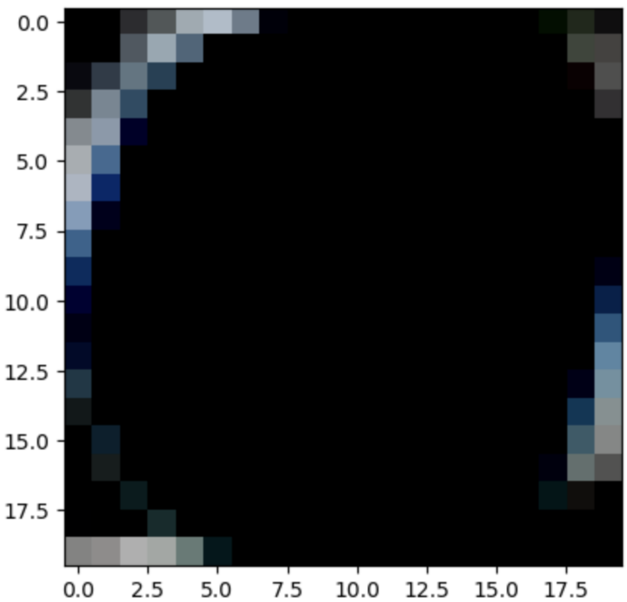

Abbildung 22: Plum pca

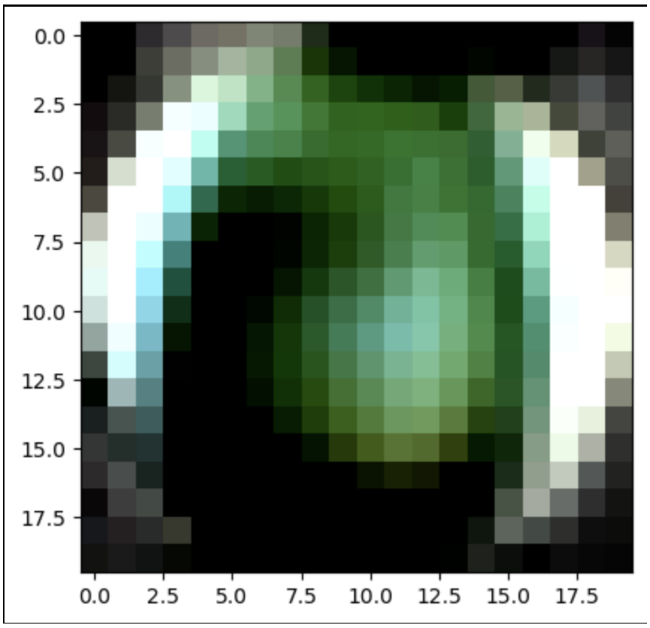

Abbildung 23: Potato pca

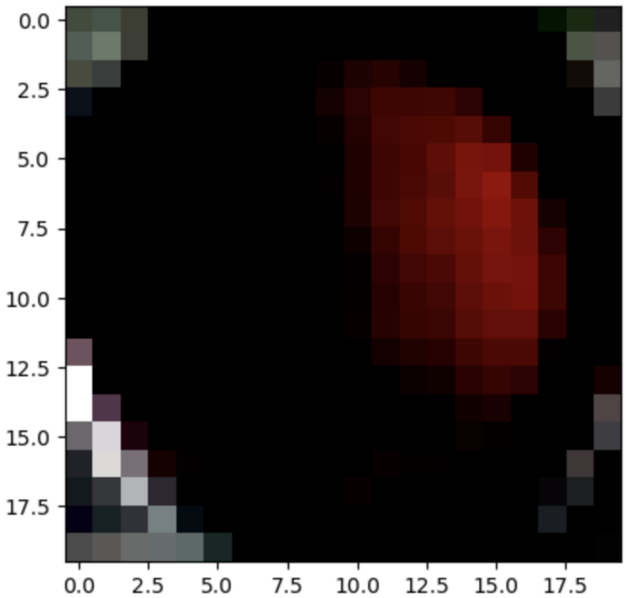

Abbildung 24: Tomato pca

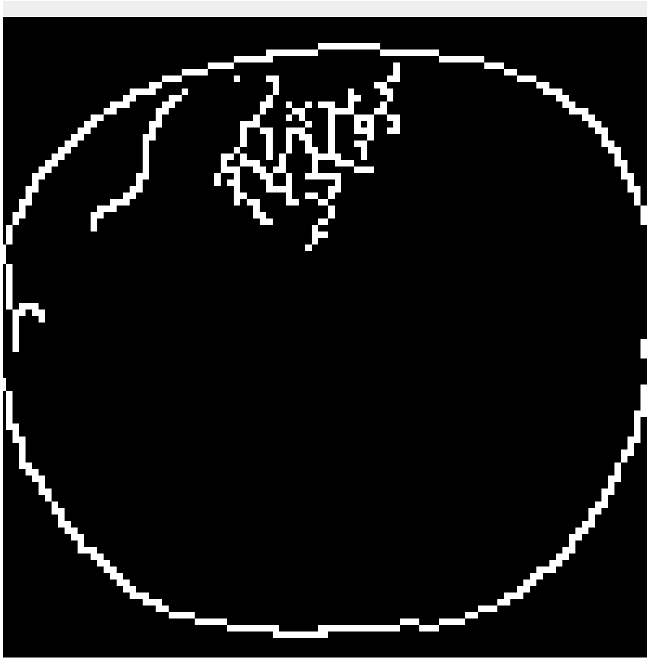

Abbildung 25: Apple contour

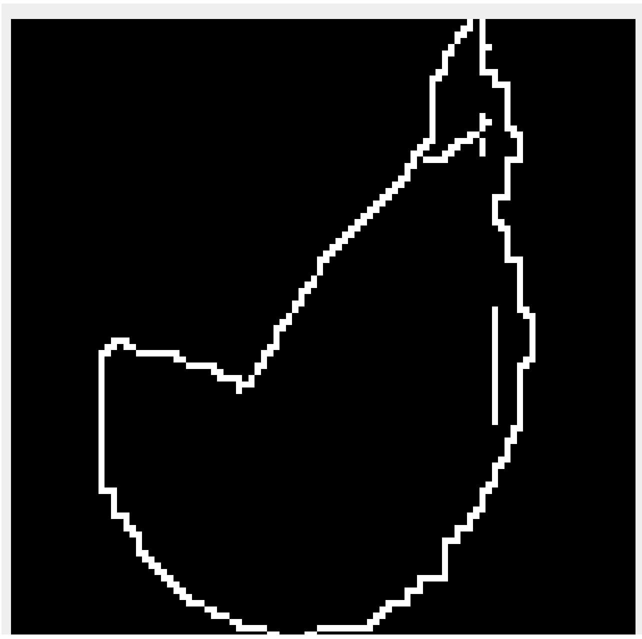

Abbildung 26: Banana contour

Abbildung 27: Cherry contour

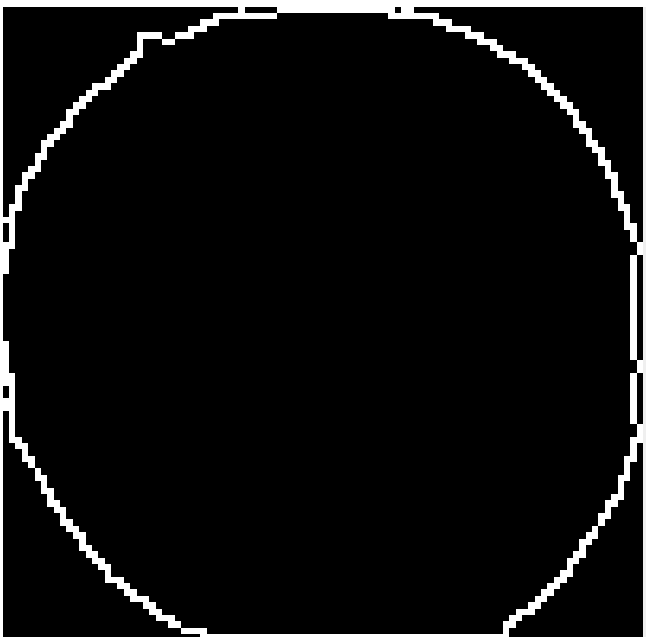

Abbildung 28: Grape contour

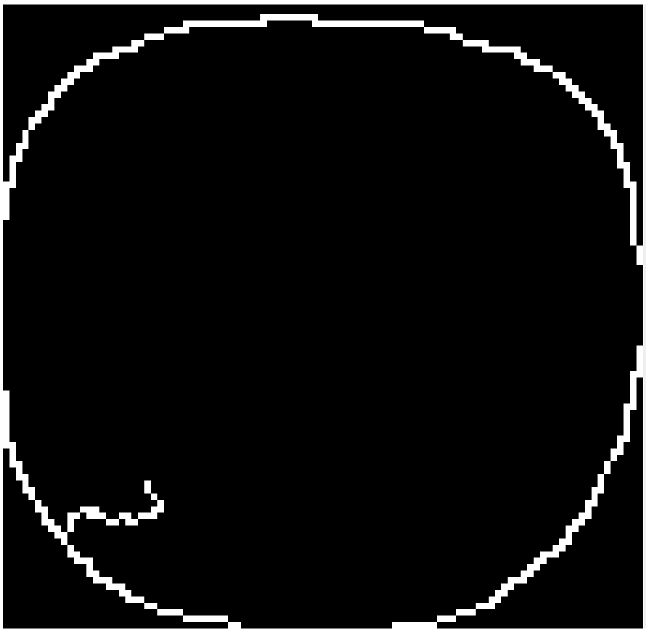

Abbildung 29: Peach contour

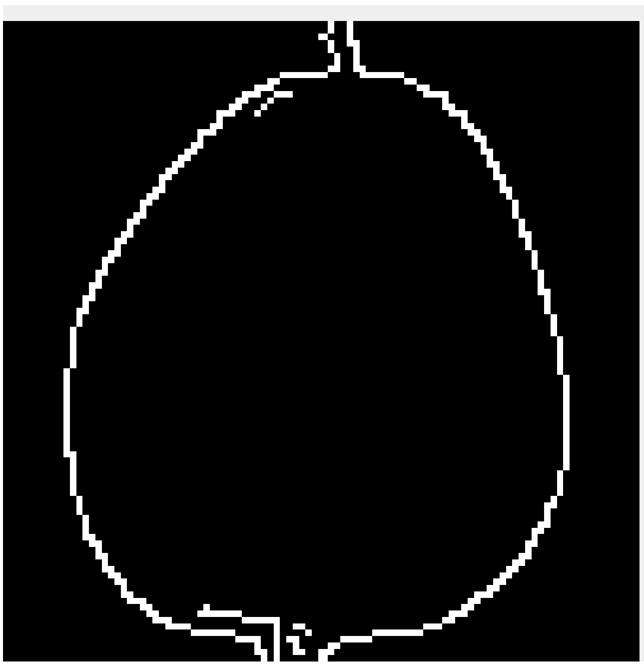

Abbildung 30: Pear contour

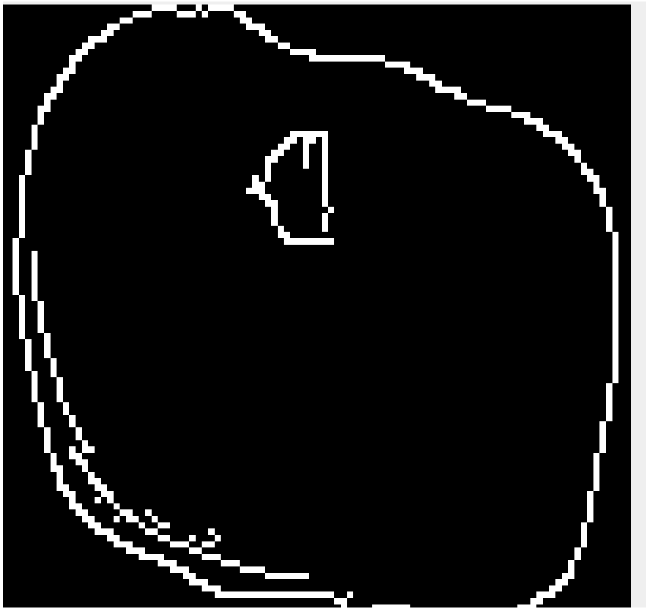

Abbildung 31: Pepper contour

Abbildung 32: Plum contour

Abbildung 33: Tomato contour

0.6 Rezultate algoritmi pentru setul de date Fruits

Pentru setul de date Fruits am aplicat cei 4 algoritmi cu diversi hiperparametri. Pentru logistic regression am luat: C = [0.1, 1, 10], max iter = [100, 1000, 10000], multi class = [auto, ovr, multinomial], solver = [newton-cg, lbfgs, sag, saga]. Pe masura ce rata de invatare crestea, crestea si acuratetea. Intr-un final am obtinut rezultate foarte bune pentru C=0.1, max iter=1000, multi class= auto, solver = newton-cg. Pe setul de train am obtinut o acuratete de 77.53, iar pe setul de test am ontinut o acuratete de 65.1. Pentru Random Forest am luat number trees = [5, 10, 50, 100, 200], max depth = [5, 10, 50, 100, 500, None], max features = [0.2, 0.5, 0.8, 1]. Am obtinut scoruri diferite radical, spre deosebire de logistic regression, care parea dependent mai mult de rata de invatare. Random forest performeaza mai bine pentru max samples care sunt diferite de 1, adica mai mici decat 1. Acuratetea pe train este suspect de mare, iar pe test este buna si ea. Pentru 99 acuratete pe train am obtinut 94 pe test, cu hiperparametri n estimators=200, max depth=50, max samples=0.8. Pentru svm, am luat urmatoarele: C = [0.01, 0.1, 1, 10], kernel = [linear, poly, rbf, sigmoid]. Se observa ca modelul are acuratete mare pe train in general pentru kernelul liniar, insa pe masura ce creste rata de invatare, creste si acuratetea pe restul kernelelor. Cea mai buna acuratete am ontinut-o pentru c=10 si kernel = rbf. Pe train am 99, iar pe test am 94. Pentru xgboost, am luat urmatorii hiperparametri: N = [10, 50, 100, 150, 200], max depth = [3, 5, 10, 20], lr = [0.1, 0.01, 0.001]. Se observa ca pe masura ce numarul de arbori creste, creste si acuratetea, iar pe masura ce creste si adancimea maxima, creste si acuratetea (tendinta de overfit). Pentru 200 de arbori, adancimea 5, learning rate 0.1, am obtinut 99 acuratete pe train si 85 pe test.

0.7 Rezultate algoritmi pentru setul de date MNIST

Am aplicat algoritmii pe setul de date MNIST, cu aceiasi hiperparametri si aceleasi valori ale acestora. Pentru Random Forest, cele mai bune valori le-am obtinut pentru n estimators=200, max depth=None, max samples=0.8; atat pe test, cat si pe train am obtinut acuratete de 85. Pentru xgboost, am obtinut o acuratete pe train si pe test de 87, cu hiperparametri max depth=20, learning rate=0.1, numar arbori=200. Pentru logistic regression, am obstinut acuratete de 81 pentru train si test, pentru hiperparametri C=10, max iter=1000, multi class=auto, solver = newton-cg. Pentru svm, cu hiperparametri C=10 si kernel=rbf am obtinut acuratete de 88 atat pe train cat si pe test.

0.8 Prediction, Recall, F1-score Fruits

Pentru logistic regression:
precision recall f1-score support

Apple 0.5159 0.6808 0.5870 2525
Apricot 0.7244 0.5610 0.6323 164
Avocado 0.6684 0.8414 0.7450 309
Banana 0.7104 0.5930 0.6464 484
Beetroot 0.5781 0.2467 0.3458 150
Blueberry 0.8455 0.6753 0.7509 154
Cabbage 1.0000 0.9574 0.9783 47
Cactus 0.5122 0.5060 0.5091 166
Cantaloupe 0.9752 0.9573 0.9662 328
Carambula 0.4242 0.2530 0.3170 166
Carrot 0.8772 1.0000 0.9346 50
Cauliflower 0.5964 0.7009 0.6444 234
Cherry 0.5595 0.5366 0.5478 1148
Chestnut 0.5631 0.3791 0.4531 153
Clementine 0.9822 1.0000 0.9910 166
Cocos 0.6893 0.7349 0.7114 166
Corn 0.5874 0.4309 0.4972 304
Cucumber 0.6725 0.6451 0.6585 417
Dates 0.9362 0.7952 0.8599 166

Eggplant 0.6108 0.4322 0.5062 236
Fig 0.8987 0.6068 0.7245 234
Ginger 0.2941 0.0505 0.0862 99
Granadilla 0.6538 0.6145 0.6335 166
Grape 0.7121 0.8290 0.7661 1146
Grapefruit 0.9617 0.9879 0.9746 330
Guava 0.9000 0.8675 0.8834 166
Hazelnut 0.8063 0.8217 0.8139 157
Huckleberry 1.0000 0.9699 0.9847 166
Kaki 0.7970 0.9458 0.8650 166
Kiwi 0.6610 0.5000 0.5693 156
Kohlrabi 0.7500 0.4204 0.5388 157
Kumquats 0.8768 0.7289 0.7961 166
Lemon 0.7827 0.7424 0.7621 330
Limes 0.8912 0.7892 0.8371 166
Lychee 0.9595 1.0000 0.9794 166
Mandarine 0.4800 0.5060 0.4927 166
Mango 0.5709 0.5487 0.5596 308
Mangostan 0.7595 0.5882 0.6630 102
Maracuja 0.9487 0.8916 0.9193 166
Melon 0.7461 0.7764 0.7610 246
Mulberry 0.9154 0.7256 0.8095 164
Nectarine 0.2481 0.0988 0.1413 324
Nut 0.5581 0.3763 0.4495 396
Onion 0.5392 0.6408 0.5856 451
Orange 0.8989 1.0000 0.9467 160
Papaya 0.9324 0.8415 0.8846 164
Passion 0.9938 0.9699 0.9817 166
Peach 0.4423 0.3206 0.3717 574
Pear 0.4186 0.4191 0.4188 1761
Pepino 0.6162 0.6867 0.6496 166
Pepper 0.6008 0.5521 0.5754 826
Physalis 0.7409 0.7409 0.7409 328
Pineapple 0.7586 0.8693 0.8102 329
Pitahaya 0.7296 0.8614 0.7901 166
Plum 0.7426 0.7538 0.7481 597
Pomegranate 0.4926 0.4085 0.4467 164

Pomelo 0.9329 1.0000 0.9653 153
Potato 0.4541 0.3295 0.3819 601
Quince 0.9294 0.9518 0.9405 166
Rambutan 0.7887 0.9329 0.8547 164
Raspberry 1.0000 0.9639 0.9816 166
Redcurrant 1.0000 0.7134 0.8327 164
Salak 0.5249 0.8457 0.6478 162
Strawberry 0.5765 0.3951 0.4689 410
Tamarillo 0.6946 1.0000 0.8198 166
Tangelo 0.8508 0.9277 0.8876 166
Tomato 0.6139 0.6930 0.6511 1707
Walnut 0.7367 1.0000 0.8484 249
Watermelon 0.8931 0.7452 0.8125 157
Zucchini 0.7960 1.0000 0.8864 160

accuracy 0.6505 23619
macro avg 0.7271 0.6982 0.7032 23619
weighted avg 0.6482 0.6505 0.6422 23619

Pentru SVM:
precision recall f1-score support

Apple 0.9011 0.9774 0.9377 2525
Apricot 0.9939 1.0000 0.9970 164
Avocado 0.9035 1.0000 0.9493 309
Banana 0.6423 0.8347 0.7260 484
Beetroot 0.8976 0.7600 0.8231 150
Blueberry 0.9935 1.0000 0.9968 154
Cabbage 0.9388 0.9787 0.9583 47
Cactus 0.9855 0.8193 0.8947 166
Cantaloupe 0.9909 0.9970 0.9939 328
Carambula 0.9750 0.9398 0.9571 166
Carrot 1.0000 1.0000 1.0000 50
Cauliflower 0.9331 0.9530 0.9429 234
Cherry 0.9665 0.9556 0.9610 1148
Chestnut 0.9481 0.8366 0.8889 153
Clementine 1.0000 1.0000 1.0000 166

Cocos 0.9866 0.8855 0.9333 166
Corn 0.8708 0.5987 0.7096 304
Cucumber 0.9713 0.8921 0.9300 417
Dates 0.9934 0.9096 0.9497 166
Eggplant 0.9029 0.7881 0.8416 236
Fig 1.0000 0.9957 0.9979 234
Ginger 0.9103 0.7172 0.8023 99
Granadilla 1.0000 1.0000 1.0000 166
Grape 0.9617 0.9860 0.9737 1146
Grapefruit 1.0000 1.0000 1.0000 330
Guava 1.0000 1.0000 1.0000 166
Hazelnut 0.9474 0.9172 0.9320 157
Huckleberry 1.0000 1.0000 1.0000 166
Kaki 1.0000 1.0000 1.0000 166
Kiwi 1.0000 1.0000 1.0000 156
Kohlrabi 1.0000 0.9682 0.9838 157
Kumquats 1.0000 1.0000 1.0000 166
Lemon 1.0000 1.0000 1.0000 330
Limes 0.9708 1.0000 0.9852 166
Lychee 1.0000 0.9699 0.9847 166
Mandarine 1.0000 0.9337 0.9657 166
Mango 0.9965 0.9351 0.9648 308
Mangostan 0.9515 0.9608 0.9561 102
Maracuja 0.9345 0.9458 0.9401 166
Melon 0.9959 0.9878 0.9918 246
Mulberry 1.0000 0.9939 0.9969 164
Nectarine 0.8586 0.7870 0.8213 324
Nut 0.9915 0.8838 0.9346 396
Onion 0.8756 0.8736 0.8746 451
Orange 0.9302 1.0000 0.9639 160
Papaya 1.0000 1.0000 1.0000 164
Passion 1.0000 1.0000 1.0000 166
Peach 0.9450 0.9878 0.9659 574
Pear 0.9091 0.8802 0.8944 1761
Pepino 0.9920 0.7470 0.8522 166
Pepper 0.8894 0.9540 0.9206 826
Physalis 0.8814 0.9970 0.9356 328

Pineapple 0.9792 1.0000 0.9895 329
Pitahaya 0.9708 1.0000 0.9852 166
Plum 0.9241 0.9581 0.9408 597
Pomegranate 0.9671 0.8963 0.9304 164
Pomelo 1.0000 1.0000 1.0000 153
Potato 0.9564 0.9118 0.9336 601
Quince 1.0000 1.0000 1.0000 166
Rambutan 0.9939 1.0000 0.9970 164
Raspberry 1.0000 0.9940 0.9970 166
Redcurrant 1.0000 0.9939 0.9969 164
Salak 0.9759 1.0000 0.9878 162
Strawberry 0.9576 0.8805 0.9174 410
Tamarillo 0.9881 1.0000 0.9940 166
Tangelo 0.9651 1.0000 0.9822 166
Tomato 0.9761 0.9555 0.9657 1707
Walnut 0.9765 1.0000 0.9881 249
Watermelon 0.9931 0.9108 0.9502 157
Zucchini 0.9639 1.0000 0.9816 160

accuracy 0.9417 23619
macro avg 0.9618 0.9436 0.9509 23619
weighted avg 0.9442 0.9417 0.9415 23619

Pentru Random Forest:
precision recall f1-score support
Apple 0.9011 0.9774 0.9377 2525
Apricot 0.9939 1.0000 0.9970 164
Avocado 0.9035 1.0000 0.9493 309
Banana 0.6423 0.8347 0.7260 484
Beetroot 0.8976 0.7600 0.8231 150
Blueberry 0.9935 1.0000 0.9968 154
Cabbage 0.9388 0.9787 0.9583 47
Cactus 0.9855 0.8193 0.8947 166
Cantaloupe 0.9909 0.9970 0.9939 328
Carambula 0.9750 0.9398 0.9571 166
Carrot 1.0000 1.0000 1.0000 50
Cauliflower 0.9331 0.9530 0.9429 234

Cherry 0.9665 0.9556 0.9610 1148
Chestnut 0.9481 0.8366 0.8889 153
Clementine 1.0000 1.0000 1.0000 166
Cocos 0.9866 0.8855 0.9333 166
Corn 0.8708 0.5987 0.7096 304
Cucumber 0.9713 0.8921 0.9300 417
Dates 0.9934 0.9096 0.9497 166
Eggplant 0.9029 0.7881 0.8416 236
Fig 1.0000 0.9957 0.9979 234
Ginger 0.9103 0.7172 0.8023 99
Granadilla 1.0000 1.0000 1.0000 166
Grape 0.9617 0.9860 0.9737 1146
Grapefruit 1.0000 1.0000 1.0000 330
Guava 1.0000 1.0000 1.0000 166
Hazelnut 0.9474 0.9172 0.9320 157
Huckleberry 1.0000 1.0000 1.0000 166
Kaki 1.0000 1.0000 1.0000 166
Kiwi 1.0000 1.0000 1.0000 156
Kohlrabi 1.0000 0.9682 0.9838 157
Kumquats 1.0000 1.0000 1.0000 166
Lemon 1.0000 1.0000 1.0000 330
Limes 0.9708 1.0000 0.9852 166
Lychee 1.0000 0.9699 0.9847 166
Mandarine 1.0000 0.9337 0.9657 166
Mango 0.9965 0.9351 0.9648 308
Mangostan 0.9515 0.9608 0.9561 102
Maracuja 0.9345 0.9458 0.9401 166
Melon 0.9959 0.9878 0.9918 246
Mulberry 1.0000 0.9939 0.9969 164
Nectarine 0.8586 0.7870 0.8213 324
Nut 0.9915 0.8838 0.9346 396
Onion 0.8756 0.8736 0.8746 451
Orange 0.9302 1.0000 0.9639 160
Papaya 1.0000 1.0000 1.0000 164
Passion 1.0000 1.0000 1.0000 166
Peach 0.9450 0.9878 0.9659 574
Pear 0.9091 0.8802 0.8944 1761

Pepino 0.9920 0.7470 0.8522 166
Pepper 0.8894 0.9540 0.9206 826
Physalis 0.8814 0.9970 0.9356 328
Pineapple 0.9792 1.0000 0.9895 329
Pitahaya 0.9708 1.0000 0.9852 166
Plum 0.9241 0.9581 0.9408 597
Pomegranate 0.9671 0.8963 0.9304 164
Pomelo 1.0000 1.0000 1.0000 153
Potato 0.9564 0.9118 0.9336 601
Quince 1.0000 1.0000 1.0000 166
Rambutan 0.9939 1.0000 0.9970 164
Raspberry 1.0000 0.9940 0.9970 166
Redcurrant 1.0000 0.9939 0.9969 164
Salak 0.9759 1.0000 0.9878 162
Strawberry 0.9576 0.8805 0.9174 410
Tamarillo 0.9881 1.0000 0.9940 166
Tangelo 0.9651 1.0000 0.9822 166
Tomato 0.9761 0.9555 0.9657 1707
Walnut 0.9765 1.0000 0.9881 249
Watermelon 0.9931 0.9108 0.9502 157
Zucchini 0.9639 1.0000 0.9816 160

accuracy 0.9417 23619
macro avg 0.9618 0.9436 0.9509 23619
weighted avg 0.9442 0.9417 0.9415 23619

Pentru xgboost:
precision recall f1-score support
0 0.7333 0.9723 0.8360 2525
1 0.6614 0.5122 0.5773 164
2 0.8812 0.9126 0.8967 309
3 0.7875 0.7273 0.7562 484
4 0.7128 0.4467 0.5492 150
5 1.0000 0.7338 0.8464 154
6 0.7143 0.9574 0.8182 47
7 0.7582 0.8313 0.7931 166
8 1.0000 0.9756 0.9877 328

9 0.8288 0.7289 0.7756 166

10 0.7833 0.9400 0.8545 50

11 0.9831 0.7479 0.8495 234

12 0.9179 0.9155 0.9167 1148

13 0.7890 0.5621 0.6565 153

14 0.9688 0.9337 0.9509 166

15 0.9392 0.8373 0.8854 166

16 0.7310 0.4737 0.5749 304

17 0.9531 0.7794 0.8575 417

18 0.9875 0.9518 0.9693 166

19 0.8786 0.6441 0.7433 236

20 0.9936 0.6667 0.7980 234

21 0.8085 0.3838 0.5205 99

22 0.9702 0.9819 0.9760 166

23 0.9184 0.9721 0.9445 1146

24 1.0000 0.9909 0.9954 330

25 1.0000 1.0000 1.0000 166

26 0.8690 0.9299 0.8985 157

27 1.0000 1.0000 1.0000 166

28 0.8557 1.0000 0.9222 166

29 0.9500 0.8526 0.8986 156

30 0.7794 0.3376 0.4711 157

31 0.9624 0.7711 0.8562 166

32 0.9414 0.8758 0.9074 330

33 0.8314 0.8614 0.8462 166

34 0.9931 0.8675 0.9260 166

35 1.0000 0.8614 0.9256 166

36 0.9043 0.6136 0.7311 308

37 0.9863 0.7059 0.8229 102

38 1.0000 0.7771 0.8746 166

39 0.9205 0.9878 0.9529 246

40 1.0000 0.9146 0.9554 164

41 0.9205 0.5000 0.6480 324

42 0.8393 0.7652 0.8005 396

43 0.7495 0.8293 0.7874 451

44 0.9353 0.9938 0.9636 160

45 0.9509 0.9451 0.9480 164

46 0.9811 0.9398 0.9600 166
47 0.8317 0.7317 0.7785 574
48 0.7046 0.7626 0.7325 1761
49 0.6919 0.7711 0.7293 166
50 0.9291 0.8414 0.8831 826
51 0.8511 0.9756 0.9091 328
52 0.7718 0.9970 0.8700 329
53 0.9200 0.6928 0.7904 166
54 0.8382 0.9112 0.8732 597
55 0.8418 0.8110 0.8261 164
56 1.0000 0.9477 0.9732 153
57 0.7256 0.7038 0.7145 601
58 0.9595 1.0000 0.9794 166
59 0.8256 0.8659 0.8452 164
60 1.0000 0.9940 0.9970 166
61 0.9669 0.8902 0.9270 164
62 0.8265 1.0000 0.9050 162
63 0.9344 0.8341 0.8814 410
64 0.9881 1.0000 0.9940 166
65 0.8728 0.9096 0.8909 166
66 0.8947 0.9602 0.9263 1707
67 0.8526 0.9759 0.9101 249
68 0.8515 0.5478 0.6667 157
69 1.0000 0.8375 0.9116 160

accuracy 0.8515 23619
macro avg 0.8878 0.8267 0.8477 23619
weighted avg 0.8591 0.8515 0.8478 23619

Algoritmii disting foarte bine features bazate pe contur, respectiv pe componente principale. Modelele reusesc sa invete fructe cu forme neobisnuite si culori neobisnuite, chiar daca sunt putine imagini in setul de date.

0.9 Prediction, Recall, F1-score MNIST

Pentru Random Forest:
precision recall f1-score support

0 0.7992 0.8480 0.8229 1000
1 0.9867 0.9670 0.9768 1000
2 0.7923 0.7860 0.7892 1000
3 0.8783 0.8950 0.8866 1000
4 0.7768 0.8320 0.8035 1000
5 0.9173 0.8990 0.9081 1000
6 0.7033 0.5880 0.6405 1000
7 0.8846 0.8740 0.8793 1000
8 0.9420 0.9740 0.9577 1000
9 0.8912 0.9260 0.9083 1000

accuracy 0.8589 10000
macro avg 0.8572 0.8589 0.8573 10000
weighted avg 0.8572 0.8589 0.8573 10000

Pentru xgboost:
precision recall f1-score support
0 0.8062 0.8570 0.8308 1000
1 0.9797 0.9650 0.9723 1000
2 0.8043 0.7850 0.7945 1000
3 0.8804 0.8910 0.8857 1000
4 0.7950 0.8340 0.8141 1000
5 0.9265 0.9080 0.9172 1000
6 0.7055 0.6300 0.6656 1000
7 0.9012 0.8940 0.8976 1000
8 0.9578 0.9750 0.9663 1000
9 0.9050 0.9340 0.9193 1000

accuracy 0.8673 10000
macro avg 0.8662 0.8673 0.8663 10000
weighted avg 0.8662 0.8673 0.8663 10000

Pentru Logistic Regression:
precision recall f1-score support
0 0.7755 0.8120 0.7934 1000
1 0.9504 0.9580 0.9542 1000
2 0.7543 0.6630 0.7057 1000

3 0.7953 0.8510 0.8222 1000
4 0.7059 0.7800 0.7411 1000
5 0.8871 0.8880 0.8876 1000
6 0.5562 0.4850 0.5182 1000
7 0.8659 0.8590 0.8624 1000
8 0.9336 0.9420 0.9378 1000
9 0.8928 0.9080 0.9003 1000

accuracy 0.8146 10000
macro avg 0.8117 0.8146 0.8123 10000
weighted avg 0.8117 0.8146 0.8123 10000

Pentru SVM:
precision recall f1-score support
0 0.8195 0.8400 0.8296 1000
1 0.9898 0.9740 0.9819 1000
2 0.8302 0.8070 0.8185 1000
3 0.8838 0.9050 0.8943 1000
4 0.8164 0.8450 0.8305 1000
5 0.9556 0.9260 0.9406 1000
6 0.7232 0.6870 0.7046 1000
7 0.9050 0.9340 0.9193 1000
8 0.9702 0.9780 0.9741 1000
9 0.9421 0.9430 0.9425 1000

accuracy 0.8839 10000
macro avg 0.8836 0.8839 0.8836 10000
weighted avg 0.8836 0.8839 0.8836 10000

Toate modelele, indiferent de felul lor, reusesc sa invete mai bine anumite clase: clasele 1, 3, 5, 8 si 9. Fiind grayscale, acesta ingreuneaza procesul modelului de a invata. De asemena, bazat si pe vizualizarea imaginilor folosind PCA, unele clase pot fi asemanatoare estetic, diferenta facandu-se la hu moments. De exemplu, coat si pullover seamana foarte mult atat la contur, cat si la pca, modelele deci pot sa le confunde.

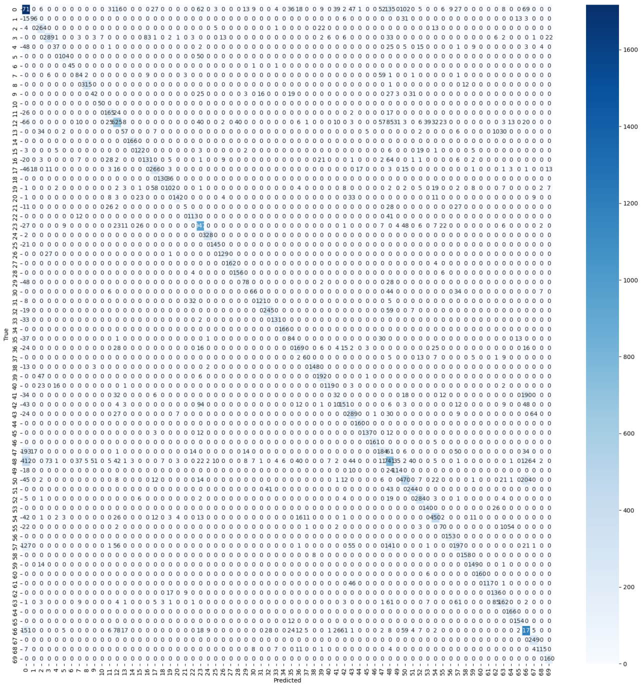

Abbildung 34: Confusion Matrix Logistic Regression Fruits

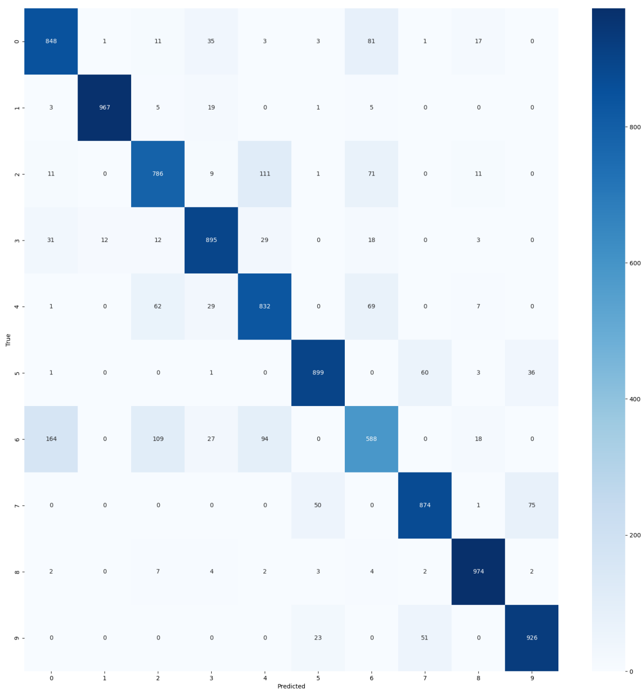

Abbildung 35: Confusion Matrix Logistic Regression MNIST

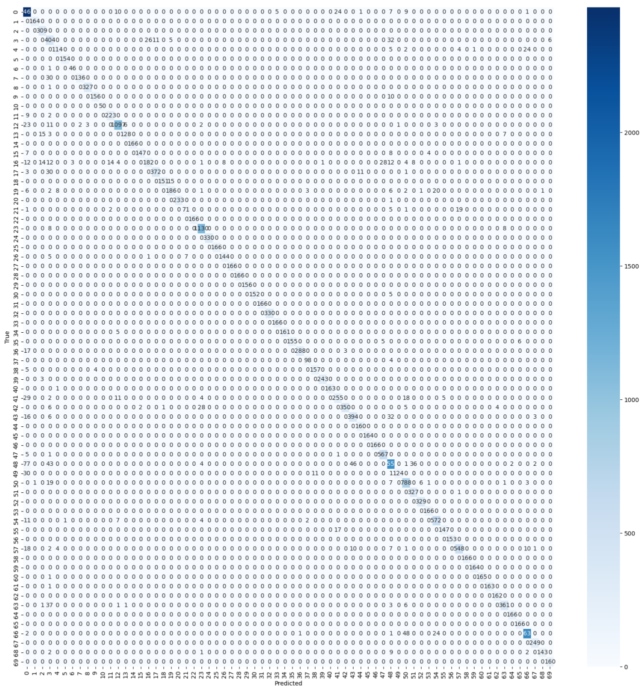

Abbildung 36: Confusion Matrix Random Forest Fruits

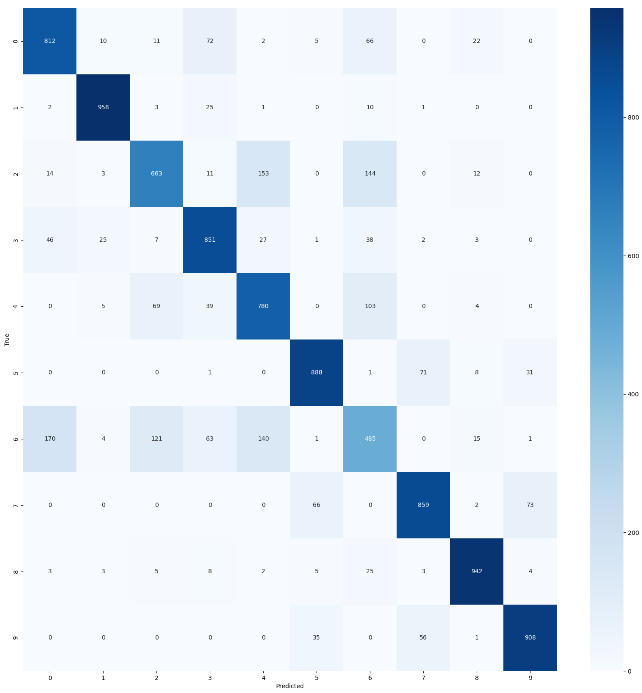

Abbildung 37: Confusion Matrix Random Forest MNIST

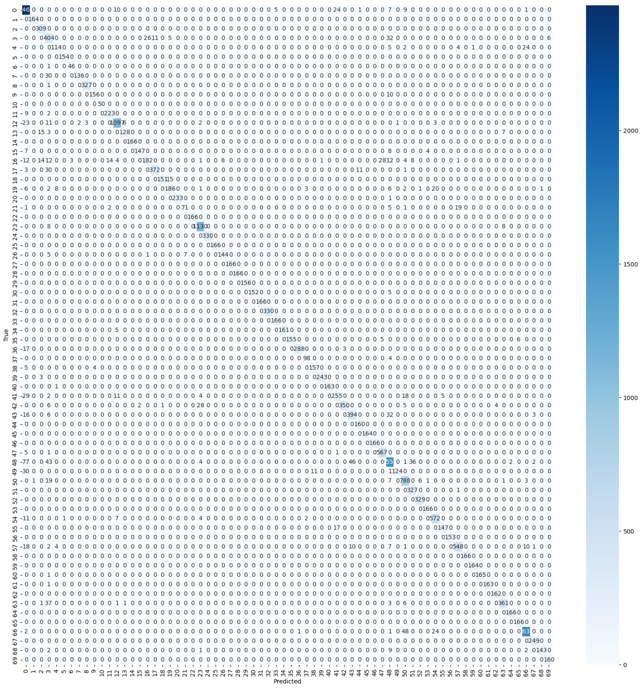

Abbildung 38: Confusion Matrix SVC Fruits

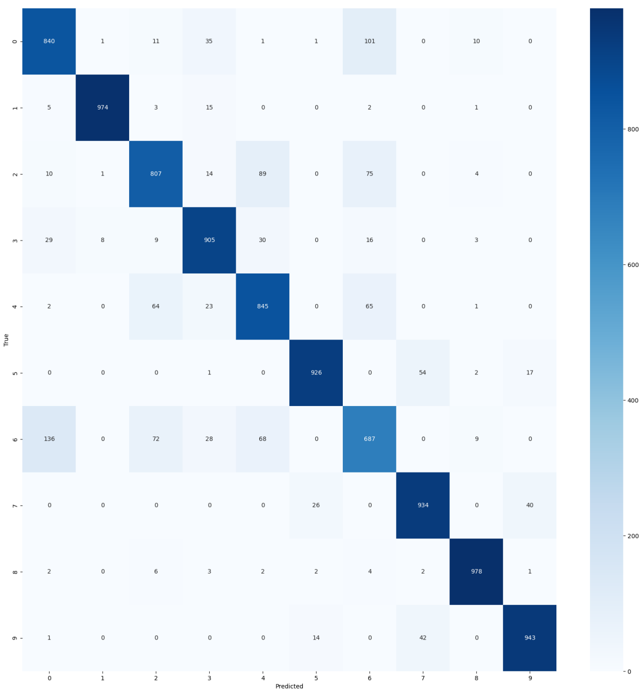

Abbildung 39: Confusion Matrix SVC MNIST

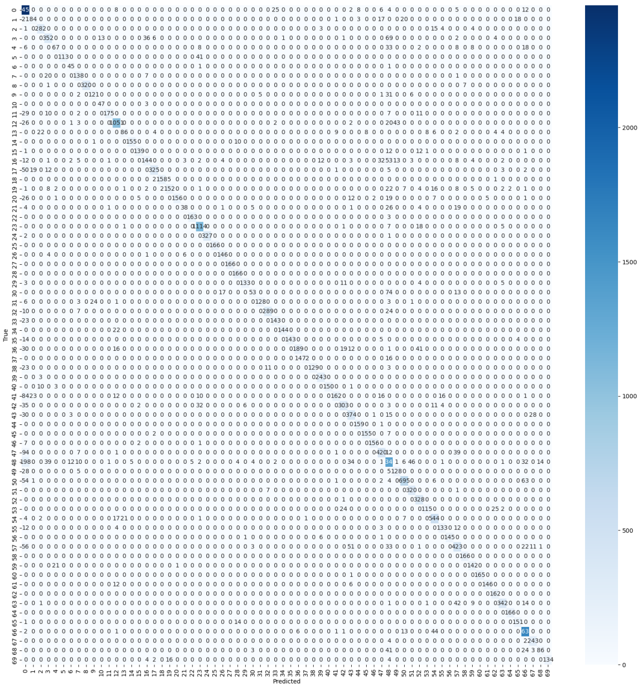

Abbildung 40: Confusion Matrix xgboost Fruits

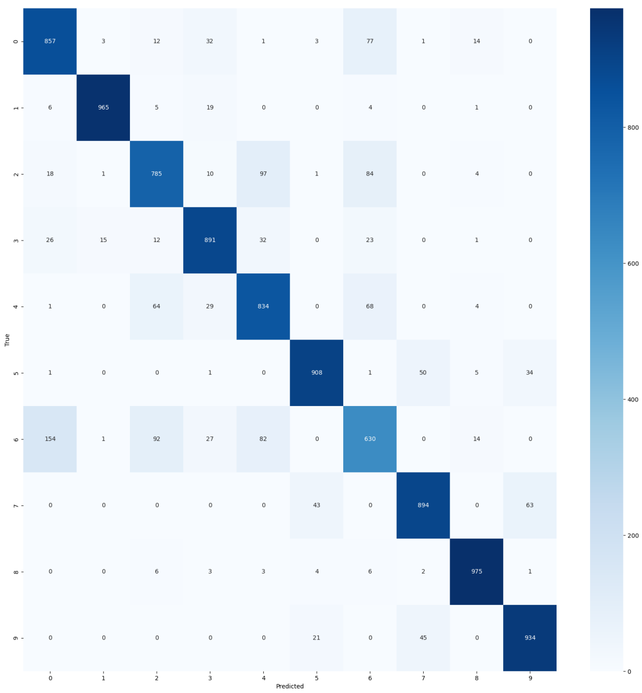

Abbildung 41: Confusion Matrix xgboost MNIST