

## Άσκηση 4

### Εφαρμογές της Ψηφιακής Επεξεργασίας Σημάτων

#### Περιεχόμενα

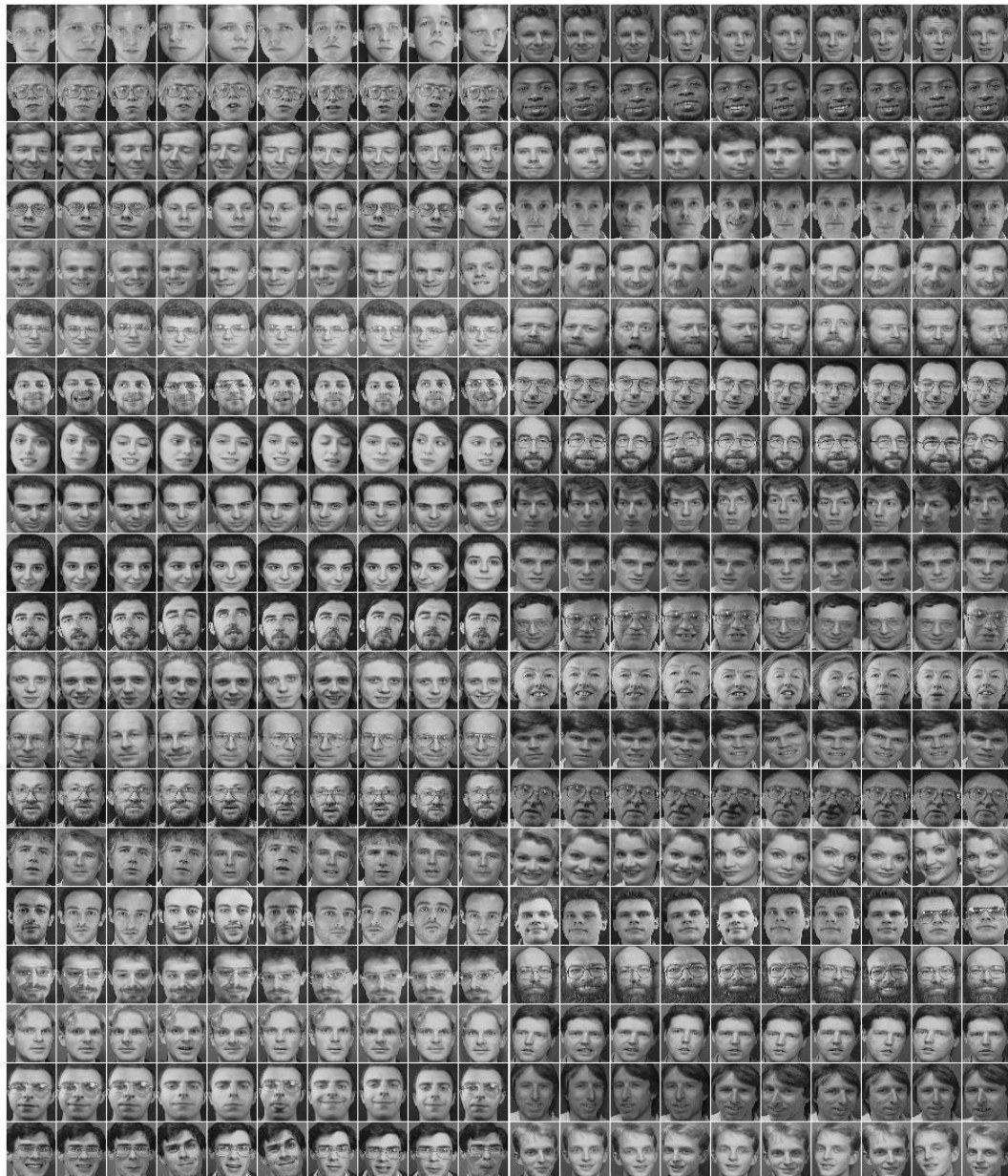
Εισαγωγή στο Dataset.....	2
Κατηγοριοποίηση δεδομένων.....	3
Διαχωρισμός δεδομένων.....	4
Ανάλυση τεχνικών.....	5
LDA.....	5
PCA.....	6
K-Means.....	8
CCA.....	12
Σύγκριση τεχνικών.....	13
Σενάριο 1:.....	13
Σενάριο 2:.....	20

Ο κώδικας μπορεί να βρεθεί εδώ:

<https://github.com/GrigorisTzortzakis/Applications-of-Digital-Signal-Processing/tree/main/Exercise%204>

## Εισαγωγή στο Dataset

Αρχικά, οφείλουμε να κατανοήσουμε την δομή του dataset που μας δίνεται. Με μια γρήγορη αναζήτηση βρίσκουμε ότι είναι το AT&T Dataset of faces. Το συγκεκριμένο αποτελείτε από 40 άτομα και το καθένα έχει από 10 φωτογραφίες. Συνεπώς, έχουμε 40 φακέλους, όπου ο καθένας είναι ένα διαφορετικό άτομο. Μέσα σε αυτούς, έχουμε 10 φωτογραφίες από το άτομο, καθεμιά διαφορετική από της υπόλοιπες. Συνολικά έχουμε 400 φωτογραφίες.



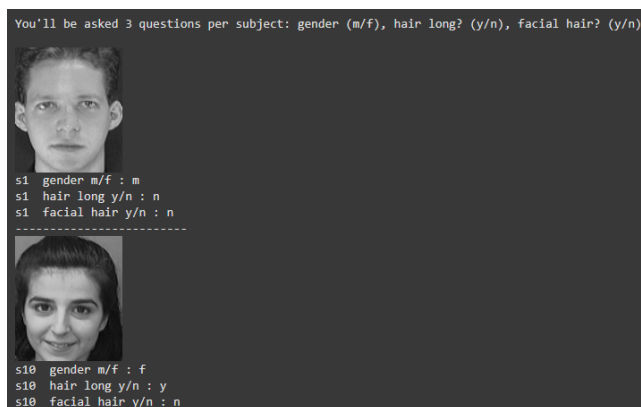
<https://www.kaggle.com/datasets/kasikrit/att-database-of-faces>

[https://git-disl.github.io/GTDLBench/datasets/att\\_face\\_dataset/](https://git-disl.github.io/GTDLBench/datasets/att_face_dataset/)

## Κατηγοριοποίηση δεδομένων

Από την παραπάνω φωτογραφία, μπορούμε να δούμε ότι υπάρχουν πολλά χαρακτηρίστηκα με τα οποία μπορούμε να δημιουργήσουμε κλασεις. Το δύσκολο τμήμα είναι να αποφασίσουμε πως θα τα αναθέσουμε σε κάθε εικόνα.

Συγκεκριμένα, τα χαρακτηριστικά που διαλέγουμε είναι το gender, facial hair, glasses, long/short hair, smile και pose turned. Μερικά από αυτά είναι σταθερά σε όλο τον φάκελο, άρα μπορούμε απευθείας να πούμε ότι π.χ φάκελος s1=male. Συγκεκριμένα, αυτό ισχύει για το gender, facial hair και long/short hair. Επιλέγουμε για αυτά να αναθέσουμε τις κλασεις με το χέρι αντί για κάποιο αυτόματο library της rython, καθώς παρατηρήθηκε ότι υπάρχει μεγάλη ανακρίβεια για το συγκεκριμένο βήμα.



Συνεχίζοντας, έως τώρα έχουμε αναθέσει 3 κλασεις σε όλους τους φακέλους και συνεπώς σε όλες τις εικόνες μέσα σε αυτούς. Το επόμενο βήμα είναι πιο δύσκολο, καθώς εκείνα τα χαρακτηριστικά δεν είναι σταθερά για όλες τις εικόνες μέσα στον φάκελο π.χ βλέπουμε ότι το άτομο δεν φοράει γυαλιά σε όλες τις φωτογραφίες.



Τότε, πρέπει να αναθέσουμε αυτά τα χαρακτηριστικά σε κάθε φωτογραφία αντί για όλο τον φάκελο. Επιλέγουμε πάλι αυτό να γίνει χειροκίνητα για τα γυαλιά και το χαμόγελο, ενώ χρησιμοποιούμε αυτόματο detector για την πόζα. Αξίζει να σημειωθεί ότι ο αυτόματος τρόπος είχε αρκετά καλή ακρίβεια, αλλά για να έχουμε όσο καλύτερα αποτελέσματα γίνεται επιλέξαμε την χειροκίνητη επιλογή.

Συνεπώς, έχουμε 6 κλασεις ανά εικόνα. Πρακτικά, έχουμε 6 binary επιλογές.

	image	folder	gender	hair_long	facial_hair	pose_turned	smile	glasses
0	/content/drive/MyDrive/faces_dataset/s1/1.pgm	s1	1	0	0	0	0	0
1	/content/drive/MyDrive/faces_dataset/s1/10.pgm	s1	1	0	0	1	0	0
2	/content/drive/MyDrive/faces_dataset/s1/2.pgm	s1	1	0	0	1	0	0
3	/content/drive/MyDrive/faces_dataset/s1/3.pgm	s1	1	0	0	0	0	0
4	/content/drive/MyDrive/faces_dataset/s1/4.pgm	s1	1	0	0	1	0	0
5	/content/drive/MyDrive/faces_dataset/s1/5.pgm	s1	1	0	0	1	0	0
6	/content/drive/MyDrive/faces_dataset/s1/6.pgm	s1	1	0	0	1	0	0
7	/content/drive/MyDrive/faces_dataset/s1/7.pgm	s1	1	0	0	0	0	0
8	/content/drive/MyDrive/faces_dataset/s1/8.pgm	s1	1	0	0	0	0	0
9	/content/drive/MyDrive/faces_dataset/s1/9.pgm	s1	1	0	0	1	0	0
10	/content/drive/MyDrive/faces_dataset/s10/1.pgm	s10	0	1	0	1	1	0
11	/content/drive/MyDrive/faces_dataset/s10/10.pgm	s10	0	1	0	0	0	0

Τέλος, οι κλασεις που έχουμε επιλέξει είναι μια καλή δοκιμή για όλα τα συστήματα που θα κάνουμε παρακάτω. Έχουμε επιλογές με τεράστιο bias (έχουμε 360 φωτογραφίες με άντρες και 40 με γυναίκες) αλλά και πιο ισορροπημένες (218 φωτογραφίες με κανονική πρόσοψη, 182 με μετασχηματισμένη). Αναλυτικά, οι κλασεις είναι:

- Gender= 40 γυναίκες/360 άντρες
- Hair= 260 κοντά μαλλιά/140 μακριά μαλλιά
- Facial hair= 310 χωρίς μούσι/90 με μούσι
- Pose= 218 με κανονική πόζα/182 με αλλοιωμένη
- Smile= 315 χωρίς χαμόγελο/ 85 με χαμόγελο
- Glasses= 280 χωρίς γυαλιά/ 120 με γυαλιά.

## Διαχωρισμός δεδομένων

Επιλέγουμε 2 διαφορετικούς τρόπους για την δημιουργία training και testing set. Ο πιο κλασσικός τρόπος είναι να κάνουμε 80/20 split σε κάθε φάκελο, δηλαδή να έχουμε 8 εικόνες από τον κάθε φάκελο για εκπαίδευση και 2 για δοκιμές. Τότε, έχουμε 320 εικόνες για train και 80 για test.

Ο δεύτερος τρόπος είναι πιο δύσκολος για την δοκιμή των συστημάτων. Δεν επιτρέπουμε εικόνες από τον ίδιο φάκελο, αλλά μόνο από διαφορετικούς. Δηλαδή, επιλέγουμε 32 φακέλους για train και 8 για test. Με αυτόν τον τρόπο, έχουμε τελείως άγνωστα πρόσωπα για τις δοκιμές, ένα πρόβλημα το οποίο είναι σαφώς πολύ πιο δύσκολο.

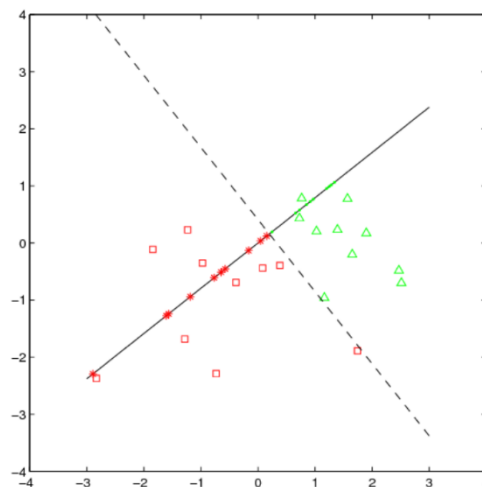
# Ανάλυση τεχνικών

## LDA

Ο FLDA επιδιώκει να βρει έναν μόνο γραμμικό άξονα στον πολυδιάστατο χώρο των χαρακτηριστικών. Η βασική ιδέα είναι ότι, για κάθε δυαδική ιδιότητα (π.χ. «φοράει γυαλιά»/«δεν φοράει γυαλιά»), πρέπει να ορίσουμε ένα διαχωριστικό επίπεδο που αποτελείτε από ένα διάνυσμα βάρους  $w$  και ένα κατώφλι  $b$ , έτσι ώστε οι προβολές των δειγμάτων της μίας τάξης να βρίσκονται μακριά από τις προβολές των δειγμάτων της άλλης.

Για κάθε ιδιότητα, αρχικά υπολογίζουμε το κεντρικό διάνυσμα (centroid) κάθε τάξης, δηλαδή το μέσο όλων των χαρακτηριστικών των δειγμάτων που ανήκουν σε αυτήν την τάξη. Στη συνέχεια, υπολογίζουμε την inclass διασπορά, η οποία μετράει πόσο πολύ «σκορπίζονται» τα δείγματα γύρω από το δικό τους μέσο. Προσθέτοντας, η κύρια πράξη του είναι να χρησιμοποιούμε αυτή την πληροφορία για να φτιάξουμε το διάνυσμα βάρους  $w$ . Με άλλα λόγια, ορίζουμε την κατεύθυνση στο χώρο των χαρακτηριστικών η οποία μεγιστοποιεί την απόσταση ανάμεσα στα δύο μέσα (outclass διασπορά) και ταυτόχρονα ελαχιστοποιεί την inclass διασπορά. Το αποτέλεσμα είναι μία μονάχα ευθεία (άξονας) επί του οποίου η προβολή των δειγμάτων δίνει όσο το δυνατόν καλύτερες τιμές για τις δύο κλάσεις.

Τέλος, για να πάρουμε την τελική απόφαση, χρειαζόμαστε ένα σημείο κατωφλίου πάνω στον άξονα αυτό, ώστε κάθε νέα παρατήρηση που θα προβάλετε στον άξονα να μπορεί να χαρακτηρίζεται είτε ως τάξη «0» είτε ως τάξη «1», ανάλογα με το αν η προβολή της βρίσκεται χαμηλότερα ή υψηλότερα από το κατώφλι. Άρα, τοποθετούμε το κατώφλι ακριβώς στο κέντρο των δύο μέσων, έτσι ώστε να έχουμε ένα συμμετρικό διαχωρισμό.



[https://www.researchgate.net/figure/Example-of-Fishers-Linear-Discriminant\\_fig1\\_22110901](https://www.researchgate.net/figure/Example-of-Fishers-Linear-Discriminant_fig1_22110901)



FLDA (manual) accuracies	
gender	: 37.50 %
hair_long	: 65.00 %
facial_hair	: 50.00 %
glasses	: 47.50 %
smile	: 53.75 %
pose_turned	: 86.25 %

FLDA (manual) accuracies	
gender	: 76.25 %
hair_long	: 88.75 %
facial_hair	: 72.50 %
glasses	: 87.50 %
smile	: 61.25 %
pose_turned	: 47.50 %

Στην αριστερή εικόνα βλέπουμε τα αποτελέσματα για το «δύσκολο» σενάριο δεδομένων (τα εντελώς διαφορετικά πρόσωπα), ενώ η δεξιά εικόνα δείχνει το σενάριο με τα ίδια πρόσωπα.

Αναλυτικότερα, στο πρώτο σενάριο, όπου κανένα πρόσωπο δεν εμφανίζεται και στους δύο φακέλους εκπαίδευσης και αξιολόγησης, οι σχετικά χαμηλές επιδόσεις αντικατοπτρίζουν την δυσκολία του να ξεχωρίσει κανείς υποσύνολα διακριτών χαρακτηριστικών σε άγνωστα πρόσωπα με μία μόνο γραμμική προβολή. Για το gender (37,5 %) η ποικιλία στις ανδρικές και γυναικείες όψεις καθιστά τον γραμμικό διαχωρισμό τυχαίο. Τα γυαλιά (47,5 %) επίσης δεν χωρίζονται, καθώς το σχήμα των σκελετών και οι γωνίες του προσώπου ποικίλλουν. Το χαμόγελο (53,75 %) αποδίδει ελαφρώς καλύτερα (και πάλι τυχαίο) χάρη στη συνέπεια που έχουν τα δόντια και τα χείλη. Το μήκος μαλλιών (65 %) είναι το ευκολότερο, αφού η διαφορά μεταξύ κοντών και μακριών μαλλιών δημιουργεί εκτεταμένες αλλαγές στην ένταση των pixel. Τα facial hair (50 %) είναι τυχαία, καθώς εισάγουν υφές που ο FLDA δεν μπορεί να συλλάβει. Τέλος, η πόζα (86,25 %) είναι η υψηλότερη, επειδή οι στροφές αριστερά/δεξιά προβάλλονται πολύ καθαρά σε έναν μεμονωμένο άξονα.

Αντίθετα, στο δεύτερο σενάριο, ο διαχωριστής εκμεταλλεύεται τις ιδιαιτερότητες. Όταν τα ίδια πρόσωπα υπάρχουν και στα δύο σύνολα, ο FLDA «απομνημονεύει» τα χαρακτηριστικά τους, με αποτέλεσμα όλες οι ακρίβειες να βρίσκονται σε αποδεκτά επίπεδα. Το gender ανεβαίνει στο 76,25 % γιατί το μοντέλο αναγνωρίζει άτομα και συνδέει συγκεκριμένα πρόσωπα με το φύλο τους. Το μήκος μαλλιών φτάνει στο 88,75 % και τα γυαλιά στο 87,5 %, διότι ήδη γνωρίζει ποιοι φορούν γυαλιά ή έχουν τα μαλλιά τους μακριά. Ακόμη και το χαμόγελο και τα μούσια έχουν σημαντική βελτίωση (61,25 % και 72,5 % αντίστοιχα), καθώς το σύστημα μαθαίνει το κάθε πρόσωπο. Το μόνο που χειροτερεύει είναι η πόζα, πιθανόν επειδή σε κάθε άτομο δεν υπάρχει ισορροπημένη αναπαράσταση κλίσεων, με αποτέλεσμα ο ταξινομητής να μην μπορεί να γενικεύσει ούτε σε γνώριμα πρόσωπα.

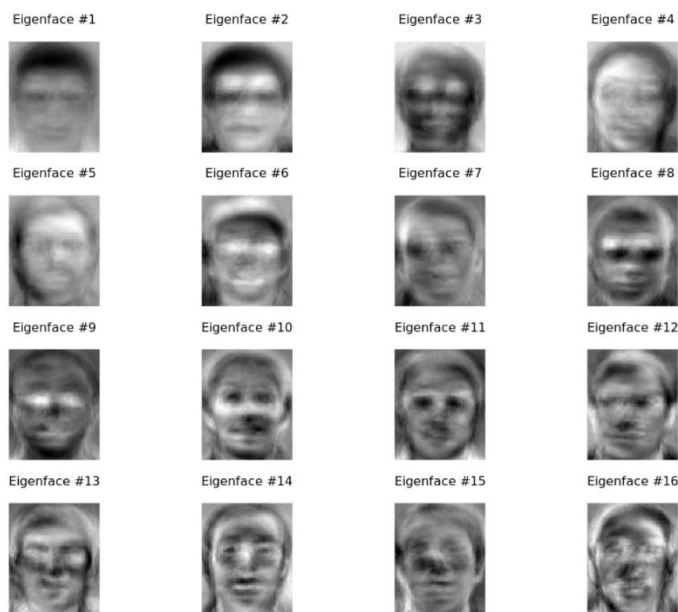
## PCA

Πρώτα, συλλέγουμε όλες τις εικόνες και χωρίς να λάβουμε υπόψιν τις ετικέτες τους, υπολογίζουμε τις κύριες συνιστώσες (principal components) οι οποίες

εξηγούν το 95 % της συνολικής διακύμανσης των pixels. Με αυτόν τον τρόπο παράγεται ένας νέος, συμπιεσμένος χώρος χαρακτηριστικών, όπου κάθε εικόνα αναπαρίσταται από ένα πολύ μικρότερο διάνυσμα συντελεστών.

Πριν από τον υπολογισμό των κύριων συνιστωσών, πρέπει να υπολογίσουμε το μέσο πρόσωπο (average face) από τις training εικόνες, δηλαδή το διάνυσμα με τον μέσο όρο όλων των pixels. Έπειτα, αφαιρούμε αυτό το μέσο πρόσωπο από κάθε εικόνα (training και test), ώστε τα δεδομένα που τροφοδοτούν την PCA να είναι zero-mean.

Αφού αποκτήσουμε τον χώρο Eigenfaces, προβάλλουμε όλες τις training εικόνες σε αυτό το υπο-διάστημα, δηλαδή υπολογίζουμε για κάθε εικόνα τους συντελεστές της ως γραμμικούς συνδυασμούς των κορυφαίων eigenvectors. Το ίδιο κάνουμε και για το test set. Για την ταξινόμηση κάθε ετικέτας, υπολογίζουμε δύο κέντρα, το μέσο διάνυσμα των προβαλλόμενων χαρακτηριστικών στο train set για την κλάση 0 και αντίστοιχα για την κλάση 1. Στη συνέχεια, για κάθε νέα εικόνα του test set, μετράμε την Ευκλείδεια απόστασή της από αυτά τα δύο κέντρα. Αν η εικόνα είναι πιο κοντά στο centroid της κλάσης 1, την ταξινομούμε ως «1», διαφορετικά ως «0».



<https://www.bytefish.de/blog/eigenfaces.html>

PCA (Eigenfaces) accuracies	
gender	: 72.50 %
hair_long	: 87.50 %
facial_hair	: 35.00 %
glasses	: 73.75 %
smile	: 67.50 %
pose_turned	: 70.00 %

PCA (Eigenfaces) accuracies	
gender	: 83.75 %
hair_long	: 73.75 %
facial_hair	: 67.50 %
glasses	: 75.00 %
smile	: 65.00 %
pose_turned	: 73.75 %

Έχουμε τα ίδια σενάρια ακριβώς με παραπάνω. Όταν οι εικόνες εκπαίδευσης και αξιολόγησης είναι από εντελώς διαφορετικά άτομα, η PCA επιλέγει κύριες συνιστώσες που αντιπροσωπεύουν τις πιο έντονες μεταβολές. Έτσι, χαρακτηριστικά όπως το μήκος των μαλλιών ξεχωρίζουν εύκολα. Αντίθετα, λεπτομέρειες όπως τα γένια ή το χαμόγελο ποικίλλουν πολύ περισσότερο μεταξύ προσώπων και δεν ευθυγραμμίζονται πάντα με τους πρώτους άξονες, για αυτό η ακρίβεια είναι χαμηλότερη.

Όταν όμως ελέγχουμε τα ίδια άτομα στο δεύτερο σενάριο, η PCA ουσιαστικά μαθαίνει τα μοναδικά χαρακτηριστικά κάθε προσώπου. Η εσωτερική διαφοροποίηση ενός ατόμου είναι μικρή σε σύγκριση με τη διαφοροποίηση μεταξύ ατόμων, οπότε αυτό ανεβάζει την ακρίβεια σε ιδιότητες όπου διατηρούνται σταθερά χαρακτηρίστηκα και άρα το μοντέλο «θυμάται».

Παραδόξως, συγκεκριμένα χαρακτηριστικά επιδεινώνονται. Το μήκος μαλλιών πέφτει από 87,5 % σε 73,75 % γιατί τα μακριά ή κοντά μαλλιά είναι σχεδόν πάντα σταθερά, άρα η PCA δεν βρίσκει διακύμανση σε αυτές τις εικόνες και δεν μαθαίνει καλά την συγκεκριμένη διάσταση.

## K-Means

Ο αλγόριθμος στοχεύει στη διαίρεση ενός συνόλου δεδομένων σε K ομάδες (clusters), έτσι ώστε τα δείγματα εντός της ίδιας ομάδας να είναι όσο το δυνατόν πιο όμοια μεταξύ τους, ενώ ταυτόχρονα να διαφοροποιούνται όσο το δυνατόν περισσότερο από τα δείγματα άλλων ομάδων. Ως κριτήριο ομοιότητας χρησιμοποιείται συνήθως η Ευκλείδεια απόσταση ( $l_2$  νόρμα), δηλαδή κάθε δείγμα αντιστοιχεί σε σημείο στον πολυδιάστατο χώρο των χαρακτηριστικών του.

Η διαδικασία εκκίνησης του K-Means περιλαμβάνει την αρχική τοποθέτηση K τυχαίων κέντρων στον ίδιο χώρο. Έπειτα, η μέθοδος ακολουθεί έναν επαναληπτικό κύκλο δύο φάσεων. Στην πρώτη φάση, κάθε δείγμα ανατίθεται στο πλησιέστερο κέντρο (assignment step), σχηματίζοντας προσωρινές ομάδες. Είναι μια διαδικασία που μοιάζει αρκετά στην λογική βασικών αλγορίθμων



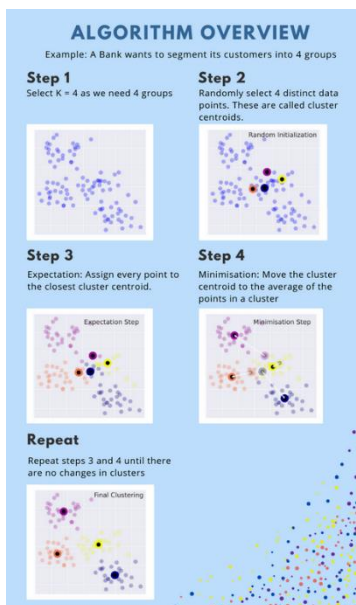
κβαντοποίησης, όπως ο PCM όπου μετράμε την απόσταση του σημείου από το κάθε προκαθορισμένο επίπεδο ώστε να το αναθέσουμε στην κατάλληλη ομάδα. Στη δεύτερη φάση, υπολογίζεται ένα νέο κέντρο για κάθε cluster ως ο μέσος όρος όλων των δειγμάτων που του ανατέθηκαν (update step). Η επανάληψη συνεχίζεται μέχρι οι θέσεις των κέντρων να σταθεροποιηθούν, δηλαδή μέχρι οι αλλαγές να γίνουν ασήμαντες.

Η μαθηματική έκφραση του στόχου που ελαχιστοποιεί ο K-Means είναι το άθροισμα των τετραγώνων των αποστάσεων κάθε δείγματος από το κέντρο, δηλαδή:

$$\sum_{k=1}^K \sum_{x_i \in S_k} \|x_i - \mu_k\|^2$$

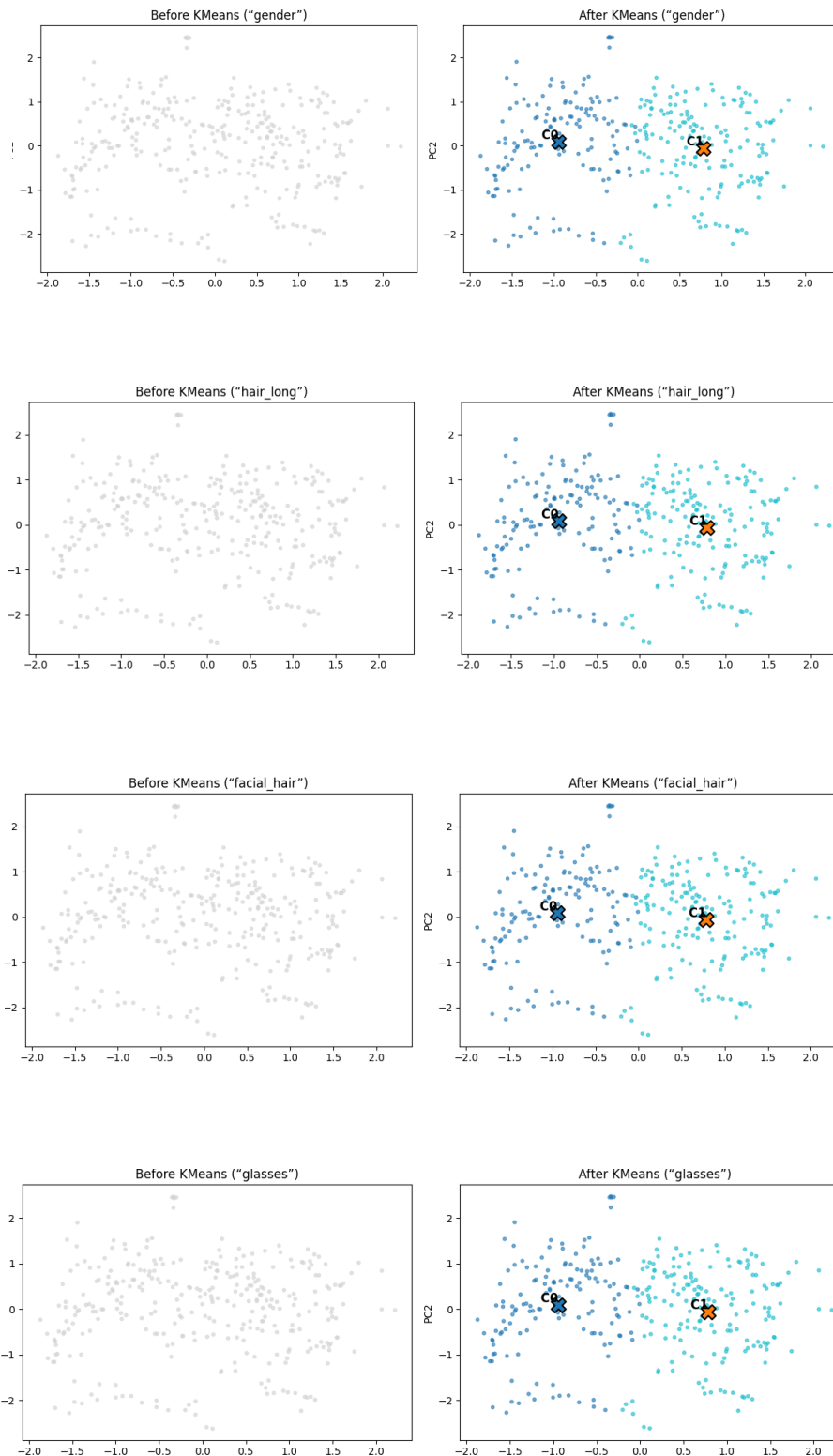
όπου  $S_k$  είναι το σύνολο δειγμάτων του k-οστού cluster και  $\mu_k$  το κέντρο του. Με την ανανεωμένη ανάθεση δειγμάτων και τον επανυπολογισμό των κέντρων, επιτυγχάνεται σταδιακή μείωση του παραπάνω κόστους, οδηγώντας σε τοπικό ελάχιστο.

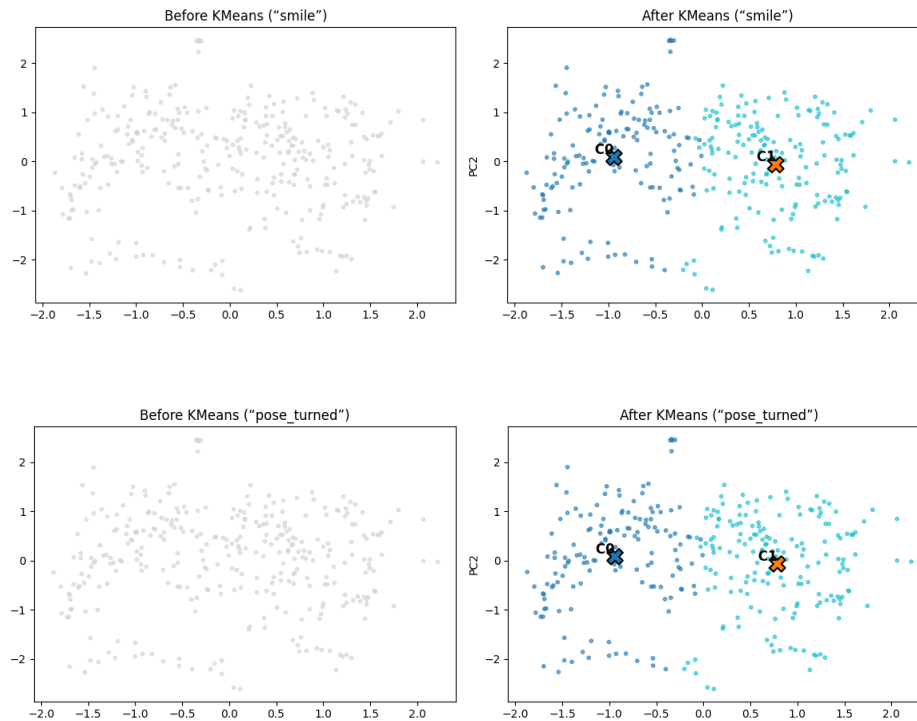
Στην περίπτωση μας, όπου έχουμε πρόβλημά ταξινόμησης με δυαδικές ετικέτες, όπου το κάθε χαρακτηριστικό παίρνει τιμή 0 ή 1, χρησιμοποιούμε  $k=2$  χωρίς να έχει το σύστημα γνώση των ετικετών. Ειδικότερα, δεν έχει νόημα να χρησιμοποιήσουμε μεγαλύτερο  $k$  για το συγκεκριμένο πρόβλημα που έχουμε ορίσει, καθώς όπως θα δούμε παρακάτω η βελτίωση είναι είτε ελάχιστη είτε μηδενική.



[https://www.reddit.com/r/learnmachinelearning/comments/kipra3/i\\_made\\_an\\_infographic\\_to\\_summarise\\_kmeans/](https://www.reddit.com/r/learnmachinelearning/comments/kipra3/i_made_an_infographic_to_summarise_kmeans/)

Για λόγους απλότητας της αναφοράς, θα παρουσιάσουμε τα αποτελέσματα του δεύτερου σεναρίου, αφού έχουμε εξηγήσει παραπάνω λεπτομερώς την διαφορά στα αποτελέσματα μεταξύ τους.



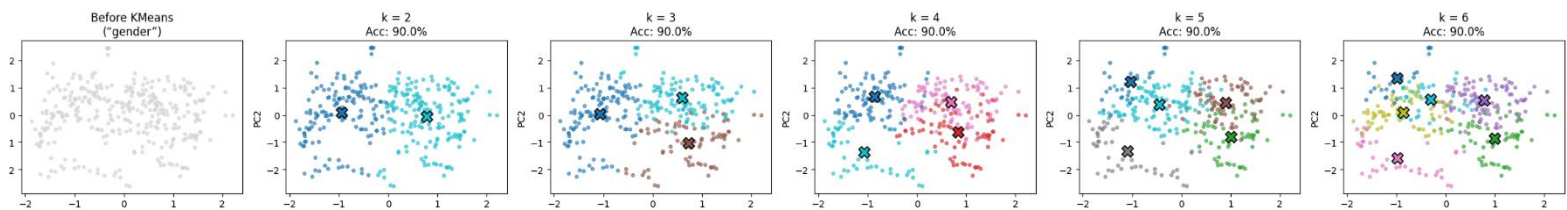


```

gender      : 90.00 %
hair_long   : 67.50 %
facial_hair : 77.50 %
glasses     : 68.75 %
smile       : 83.75 %
pose_turned : 52.50 %

```

Βλέπουμε ότι τα αποτελέσματα είναι ικανοποιητικά, ειδικά για το gender. Βέβαια, παρακάτω θα δούμε ότι το συγκεκριμένο αποτέλεσμα είναι παραπλανητικό.



gender	:	k=2: 90.0%		k=3: 90.0%		k=4: 90.0%		k=5: 90.0%		k=6: 90.0%
hair_long	:	k=2: 67.5%		k=3: 67.5%		k=4: 71.2%		k=5: 70.0%		k=6: 71.2%
facial_hair	:	k=2: 77.5%		k=3: 77.5%		k=4: 77.5%		k=5: 77.5%		k=6: 77.5%
glasses	:	k=2: 68.8%		k=3: 76.2%		k=4: 70.0%		k=5: 73.8%		k=6: 73.8%
smile	:	k=2: 83.8%		k=3: 83.8%		k=4: 83.8%		k=5: 83.8%		k=6: 83.8%
pose_turned	:	k=2: 52.5%		k=3: 62.5%		k=4: 66.2%		k=5: 67.5%		k=6: 68.8%

Εδώ πέρα επιβεβαιώνουμε το συμπέρασμα μας ότι δεν υπάρχει νόημα να χρησιμοποιήσουμε  $k$  μεγαλύτερο από το 2.

## CCA

Ο αλγόριθμος Canonical Correlation Analysis χρησιμοποιείται για να βρει δύο γραμμικούς μετασχηματισμούς, έναν των εικόνων και έναν των αντίστοιχων ετικετών σε δυαδική μορφή (one-hot), ώστε οι προβολές τους στον ίδιο μονοδιάστατο χώρο να παρουσιάζουν τη μέγιστη δυνατή συσχέτιση. Πιο συγκεκριμένα, για τις εικόνες  $X$  και τις ετικέτες  $Y$  (μετασχηματισμένες σε διάνυσμα  $[1,0]$  ή  $[0,1]$  για τις δύο κλάσεις) αναζητούμε διανύσματα βαρών  $a$  (για το  $X$ ) και  $b$  (για το  $Y$ ), τέτοια ώστε ο συσχετισμός μεταξύ των μερικών προβολών  $u=Xa$  και  $u=Yb$  να είναι μέγιστος.

Συνεχίζοντας, ο CCA επιστρέφει ένα ζεύγος μετασχηματισμών, αλλά εφόσον οι ετικέτες είναι διμερές, ενδιαφερόμαστε μόνο για το πρώτο ( $n\_components=1$ ). Αυτό σημαίνει ότι προβάλλουμε κάθε εικόνα σε μία μονή τιμή  $u$ , η οποία έχει τη μεγαλύτερη δυνατή γραμμική συσχέτιση με την αντίστοιχη «μορφή» της ετικέτας. Στην φάση της εκπαίδευσης, υπολογίζουμε τις μέσες τιμές αυτών των  $u$  προβολών ξεχωριστά για τις εικόνες με ετικέτα 0 και για αυτές με ετικέτα 1 (τα «κέντρα»  $c0$  και  $c1$  αντίστοιχα).

Τέλος, κατά την αξιολόγηση των νέων εικόνων, υπολογίζουμε την προβολή τους  $u_{test}$  με τον ίδιο γραμμικό μετασχηματισμό. Κάθε πρόβλεψη γίνεται συγκρίνοντας την απόσταση της τιμής  $u_{test}$  από τα δύο κέντρα  $c0$  και  $c1$ . Αν η τιμή είναι πιο κοντά στο  $c0$  προβλέπουμε «0», αλλιώς βάζουμε «1».

Για λόγους απλότητας της αναφοράς, θα παρουσιάσουμε τα αποτελέσματα του δεύτερου σεναρίου, αφού έχουμε εξηγήσει παραπάνω λεπτομερώς την διαφορά στα αποτελέσματα μεταξύ τους.

```
CCA accuracies - 80/20
gender      : 97.50 %
hair_long   : 93.75 %
facial_hair : 92.50 %
glasses     : 93.75 %
smile       : 81.25 %
pose_turned : 90.00 %
```

Βλέπουμε ότι οι ακρίβειες που πετυχαίνουμε είναι σε άριστο επίπεδο, καθώς από ότι φαίνεται υπάρχει υψηλή συσχέτιση αναμεσά στα pixel.

## Σύγκριση τεχνικών

Αρχικά, ας ξαναδούμε την απόδοση όλων των τεχνικών για το πρώτο σενάριο (εντελώς διαφορετικοί φάκελοι για εκπαίδευση και δοκιμές).

### Σενάριο 1:

```
FLDA (manual) accuracies
gender      : 37.50 %
hair_long   : 65.00 %
facial_hair : 50.00 %
glasses     : 47.50 %
smile       : 53.75 %
pose_turned : 86.25 %
```

```
PCA (Eigenfaces) accuracies
gender      : 72.50 %
hair_long   : 87.50 %
facial_hair : 35.00 %
glasses     : 73.75 %
smile       : 67.50 %
pose_turned : 70.00 %
```

```
CCA accuracies
gender      : 75.00 %
hair_long   : 66.25 %
facial_hair : 73.75 %
glasses     : 65.00 %
smile       : 77.50 %
pose_turned : 86.25 %
```

```
k-means accuracies
gender      : 75.00 %
hair_long   : 75.00 %
facial_hair : 87.50 %
glasses     : 58.75 %
smile       : 72.50 %
pose_turned : 43.75 %
```

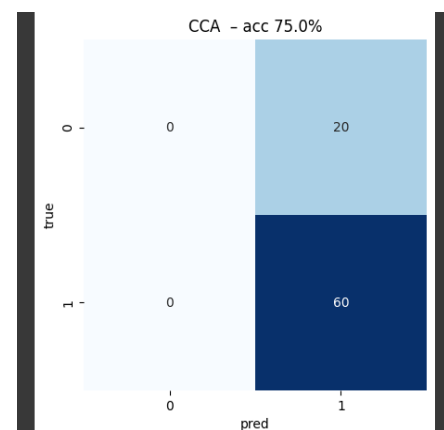
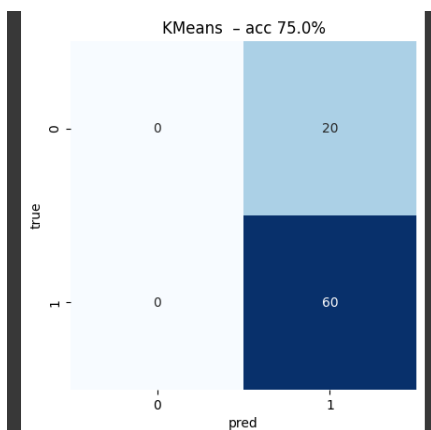
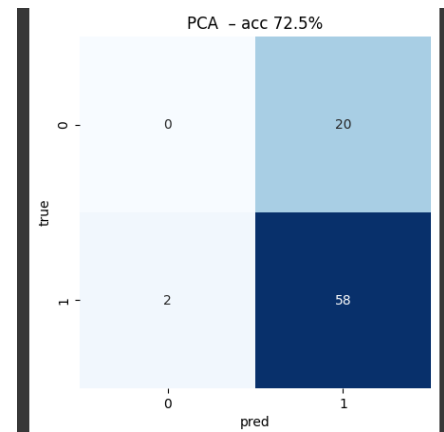
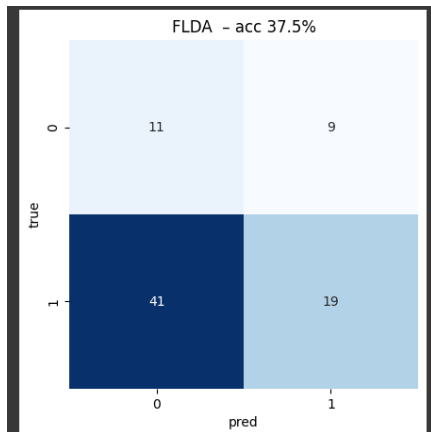
- Gender: Όταν εξετάζουμε το «φύλο», η FLDA καταρρέει, αφού επιτυγχάνει μόλις 37,5 % γιατί οι διαφορές ανδρικών-γυναικείων χαρακτηριστικών, όπως το περίγραμμα του μετώπου, διαφέρουν τόσο πολύ σε νέα άτομα που καμία σταθερή περιοχή δεν μπορεί να τις συλλάβει με αξιοπιστία. Η pca ανεβαίνει στο 72,5 %, καθώς οι κύριες συνιστώσες κωδικοποιούν τις συνολικές διαφορές σχήματος προσώπου. Ενδιαφέρον είναι ότι τόσο ο K-Means όσο και η CCA φτάνουν στο 75 %.

- Hair long: Στο «μήκος μαλλιών», η PCA νικάει (87,5 %) γιατί η αλλαγή αυτή αποτελεί μια υψηλής έντασης μεταβολή, άρα τα κορυφαία ιδιοδιανυσματα ανιχνεύουν τη μετατόπιση. Ο K-Means επίσης τα καταφέρνει καλά (75 %), διαχωρίζοντας περιοχές με «πυκνά» ρixel μαλλιών. Η CCA, αντιθέτως, πέφτει στο 66,25 %, επειδή οι ποικιλίες είναι υπερβολικά ετερογενείς για να περιγραφούν πλήρως από έναν γραμμικό άξονα. Η FLDA βρίσκεται στη μέση (65 %), καθώς έχει παρόμοιο θέμα με την CCA.
- Facial hair: Στην περίπτωση των «γενειάδων», ο K-Means κυριαρχεί (87,5 %), επειδή τα γένια εισάγουν έντονα σκοτεινά ρixel γύρω από το πηγούνι και τα χείλη, σχηματίζοντας δύο διακριτά cluster. Η CCA, εξοπλισμένη με ετικέτες, ακολουθεί στο 73,75 % αλλά εξακολουθεί να «αραιώνει» τον θόρυβο. Η PCA και η FLDA αποτυγχάνουν πλήρως, καθώς αδυνατούν να προσαρμοστούν.
- Glasses: Στον εντοπισμό των «γυαλιών», η PCA ξαναπρωτεύει (73,75 %), αφού τα συνεπή μοτίβα σκιάς και φωτεινότητας γύρω από τα μάτια αποτελούν διακυμάνσεις που συλλαμβάνονται από τις κορυφαίες συνιστώσες. Η CCA δίνει 65 %, καλύτερα από τυχαία επιλογή, αλλά υποφέρει από τα λεπτά μεταλλικά πλαίσια και τις αντανάκλασεις που παραβιάζουν την αυστηρή γραμμικότητά της. Ο K-Means φτάνει και ο FLDA πρακτικά αποτυγχάνουν.
- Smile: Στην περίπτωση του «χαμόγελου», η CCA πρωτοπορεί με 77,5 %, επειδή ευθυγραμμίζει γραμμικά τις αλλαγές στις εντάσεις των ρixel της περιοχής των χειλιών με τις αντίστοιχες ετικέτες, εξάγοντας τον άξονα που μεγιστοποιεί αυτή τη συσχέτιση. Ο K-Means ακολουθεί στενά στο 72,5 %, καθώς οι διαφορετικές καμπυλότητες και οι αντιθέσεις των ρixel μεταξύ χαμογελαστών και μη σχηματίζουν δύο σχετικά διακριτά cluster. Η PCA φτάνει στο 67,5 %, καθώς οι κορυφαίες συνιστώσες αιχμαλωτίζουν μερικά γενικά μοτίβα έκφρασης, αλλά χάνουν σε ακρίβεια. Η FLDA πρακτικά αποτυγχάνει.
- Pose: Τέλος, για τη «πόζα», τόσο η FLDA όσο και η CCA φτάνουν στο 86,25 %. Τα γεωμετρικά φίλτρα της FLDA, που μετρούν την αριστερή-δεξιά ασυμμετρία στα μάτια και τα ζυγωματικά, γενικεύονται καλά σε νέα άτομα για τον διαχωρισμό. Η CCA συγκλίνει στον ίδιο ισχυρό γραμμικό διαχωριστικό άξονα. Η PCA πέφτει στο 70 %, επειδή αναμειγνύει τη μεταβλητότητα της στάσης με την έκφραση και το φωτισμό. Ο K-Means πρακτικά αποτυγχάνει.



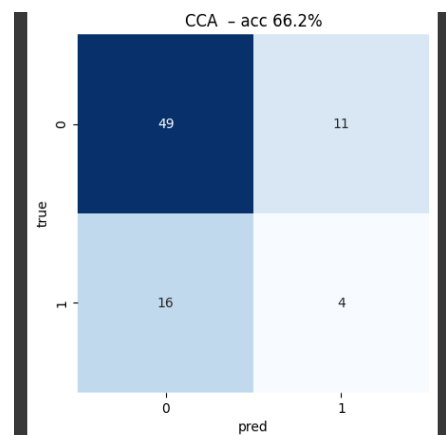
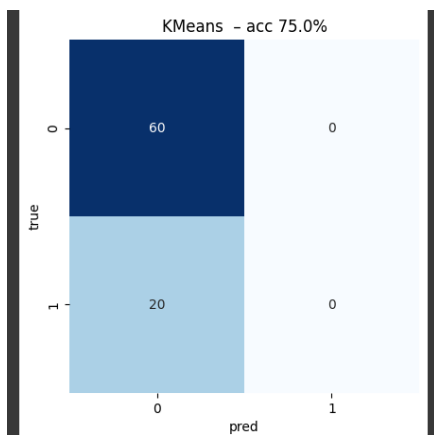
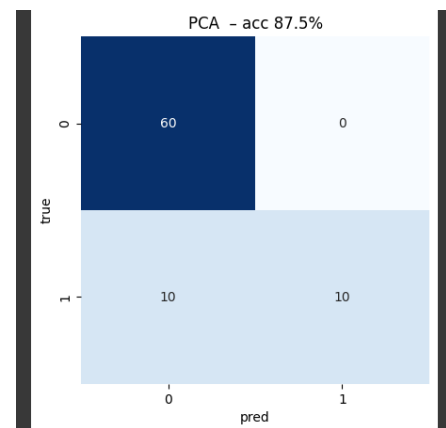
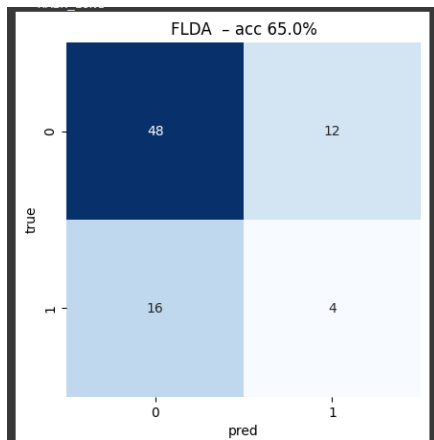
Ας δούμε τώρα το confusion matrix:

Gender:



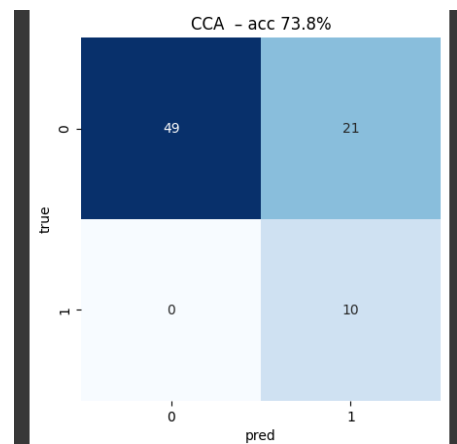
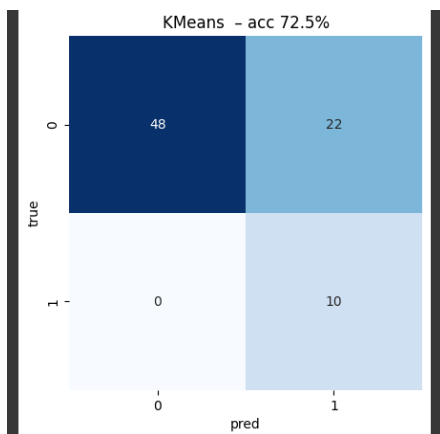
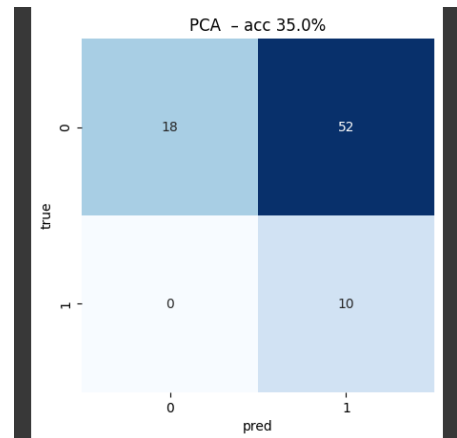
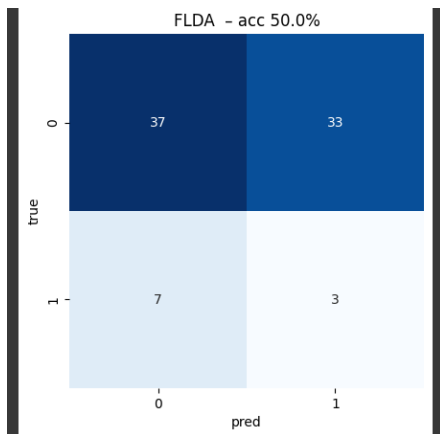
Τώρα πλέον μπορούμε να δούμε την σοβαρότητα της κατάστασης. Ενώ φαινομενικά έχουμε μεγάλη ακρίβεια, αυτό γίνεται επειδή εντοπίζουμε μεγάλο ποσοστό από τους άντρες σωστά (από τους 60 βρίσκουμε και τους 60). Όμως, ένα παρατηρήσουμε τις γυναίκες, τότε μπορούμε να δούμε ότι το σύστημα τις τοποθετεί εντελώς λάθος ως άντρες (και τις 20 γυναίκες τις εντοπίζει ως άντρες). Για τον FLDA πρακτικά αποδείξαμε ότι έχει εντελώς λάθος προβλέψεις.

Hair long:



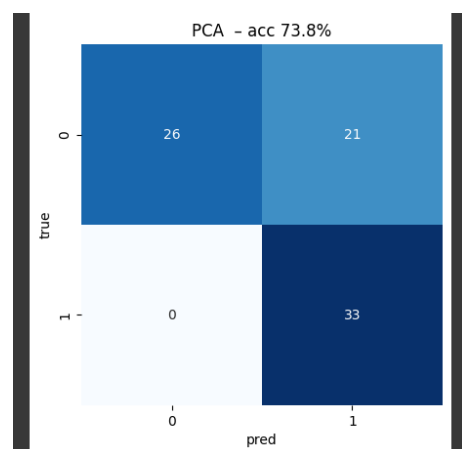
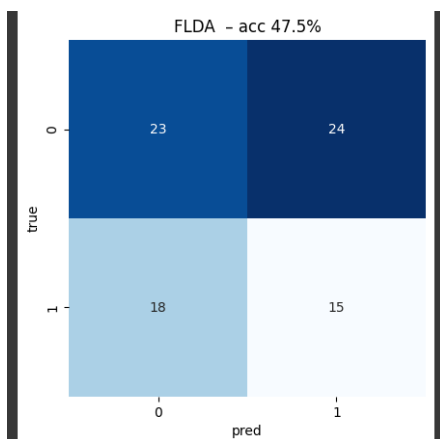
Τα αποτελέσματα και πάλι δείχνουν ότι είδαμε παραπάνω.

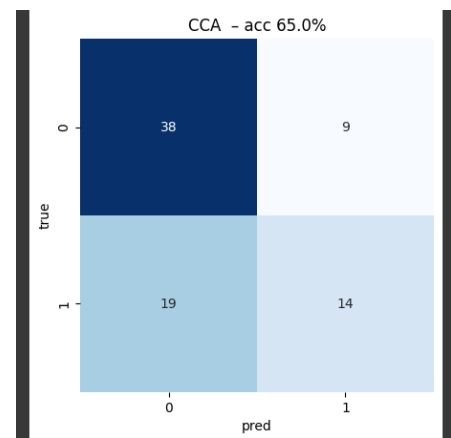
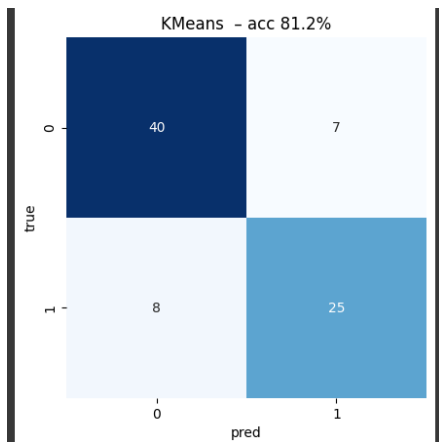
## Facial hair:



Σε αυτή την κατηγορία, τα πράγματα είναι κάπως καλύτερα. Η ακρίβεια πράγματι συμβαδίζει με το τι θα περιμέναμε.

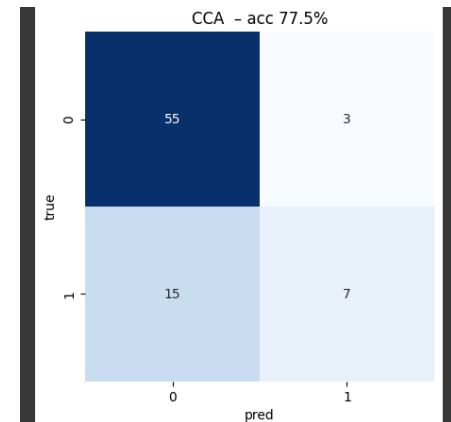
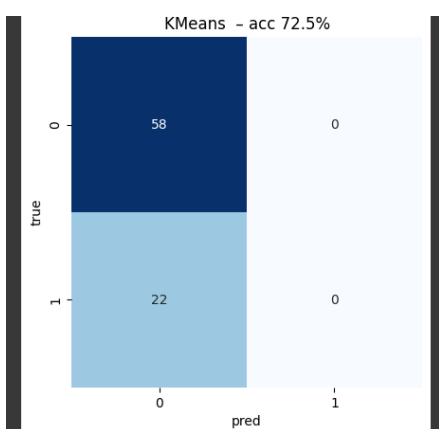
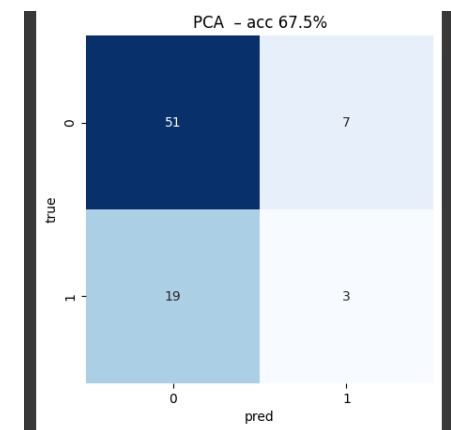
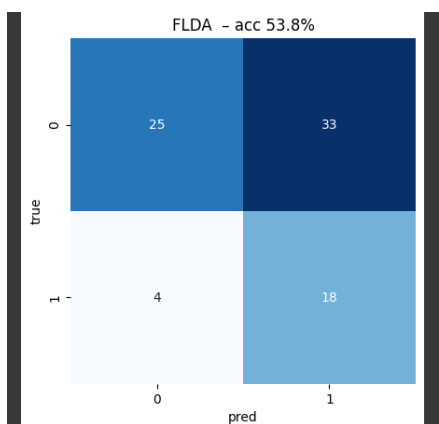
## Glasses:





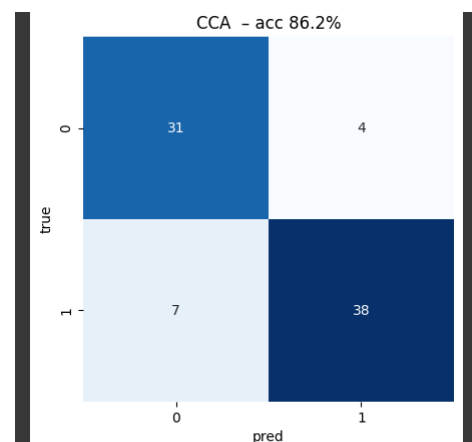
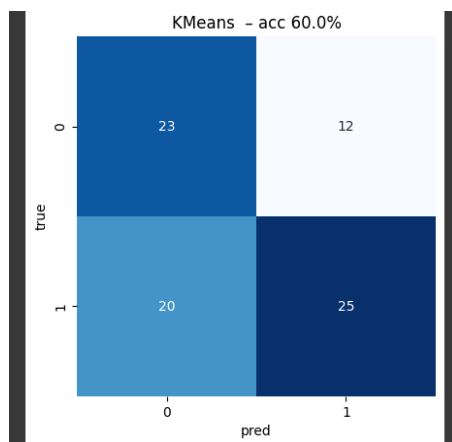
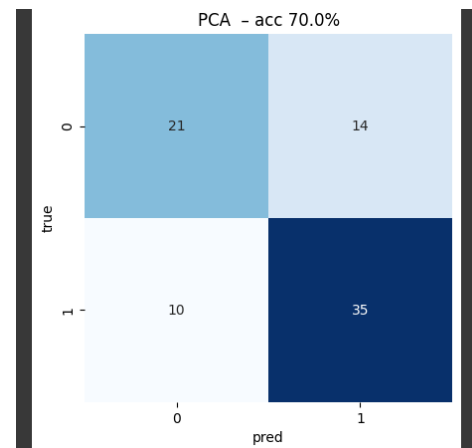
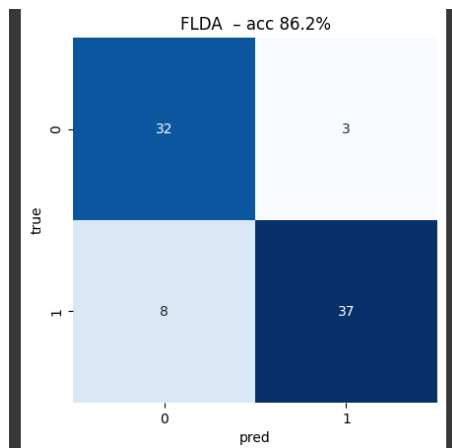
Σε αυτή την κατηγορία, ο μονός αξιολογος αλγόριθμος είναι ο Kmeans, αφού η ρα πρακτικά για την ετικέτα 0 είναι 50/50 τυχαία επιλογή, μόνο την ετικέτα 1 εντοπίζει με επιτυχία.

Smile:



Τα αποτελέσματα και πάλι είναι πρακτικά λάθος.

Pose:



Σε αυτή την κατηγορία, ο FLDA και ο CCA έχουν απόδοση ανάλογη της ακρίβειας που πετυχαίνουν.

Συμπερασματικά, βλέπουμε ότι για *biased labels* όλοι οι αλγόριθμοι αποτυγχάνουν, εφόσον εντοπίζουν σωστά μόνο τους άντρες και καθόλου τις γυναίκες. Είναι λογικό να έχουμε τέτοια αποτελέσματα, εφόσον το συγκεκριμένο σενάριο που έχουμε χωρίσει το *dataset* είναι ένα αρκετά πολύπλοκο πρόβλημα, το οποίο σιγουρά απαιτεί άλλες μεθόδους για την αποτελεσματική αντιμετώπιση του. Σε ελάχιστες κατηγορίες έχουμε σωστές προβλέψεις και έχουμε αναλύσει παραπάνω γιατί συμβαίνει αυτό για τον κάθε αλγόριθμο.

## Σενάριο 2:

```
FLDA (manual) accuracies
gender      : 76.25 %
hair_long   : 88.75 %
facial_hair : 72.50 %
glasses     : 87.50 %
smile       : 61.25 %
pose_turned : 47.50 %
```

```
PCA (Eigenfaces) accuracies
gender      : 83.75 %
hair_long   : 73.75 %
facial_hair : 67.50 %
glasses     : 75.00 %
smile       : 65.00 %
pose_turned : 73.75 %
```

```
k-means accuracies -
gender      : 90.00 %
hair_long   : 67.50 %
facial_hair : 77.50 %
glasses     : 68.75 %
smile       : 83.75 %
pose_turned : 52.50 %
```

```
CCA accuracies - 80/20
gender      : 97.50 %
hair_long   : 93.75 %
facial_hair : 92.50 %
glasses     : 93.75 %
smile       : 81.25 %
pose_turned : 90.00 %
```

Αρχικά, βλέπουμε ένα παράδοξο. Σε σχέση με το προηγούμενο σενάριο, ενώ το σύνολο των αποτελεσμάτων είναι καλύτερο, υπάρχουν μερικά τα οποία είναι χειρότερα. Ένα από αυτά είναι το *pose* στον FLDA. Αυτό συμβαίνει επειδή όταν την εκπαιδεύουμε σε εντελώς διαφορετικά άτομα, κάθε πρόσωπο συμβάλλει με μια μικρή αλλά σταθερά ευθυγραμμισμένη “μετατόπιση” μεταξύ πλάγιας και μετωπικής όψης. Σε δεκάδες πρόσωπα, αυτές οι μετατοπίσεις αθροίζονται σε μια ξεκάθαρη, κυρίαρχη διεύθυνση που η LDA αναγνωρίζει και γενικεύει καλά σε νέα άτομα ( $\approx 86\%$  ακρίβεια).

Αντίθετα, όταν εκπαιδεύουμε σε άτομα που επίσης θα δοκιμάσουμε, οι μεγάλες διαφορές ταυτότητας καταπνίγουν τις πολύ πιο διακριτές αλλαγές της στάσης κεφαλιού. Η LDA τότε κλίνει την προβολή της προς το να ξεχωρίζει άτομα παρά τη διάκριση μεταξύ “πλάγιας” και “μετωπικής” όψης, με αποτέλεσμα η ικανότητά της να διακρίνει την στάση να καταρρέει ( $\approx 47\%$  ακρίβεια).

Συνεχίζοντας, μια ακόμη κατηγορία όπου παρατηρούμε επιδείνωση είναι η *hair long* στην *pca*. Αυτό γίνεται επειδή όταν υπολογίζουμε πρόσωπα εντελώς διαφορετικών ατόμων, οι *global variation* όπως το μήκος μαλλιών κυριαρχούν στις κύριες συνιστώσες. Η προβολή άγνωστων προσώπων σε αυτές τις συνιστώσες διαχωρίζει καθαρά τα μακριά από τα κοντά μαλλιά ( $\approx 87\%$  ακρίβεια).



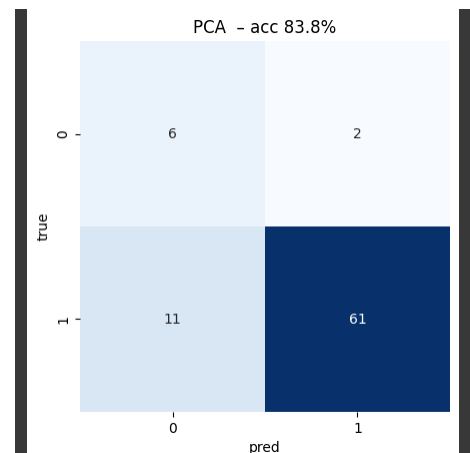
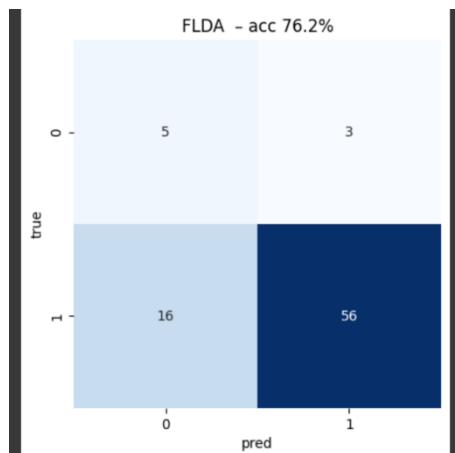
Αν όμως υπολογίζουμε τους «άξονες» από το ίδιο μικρό σύνολο ταυτοτήτων, οι κύριες συνιστώσες «καταλαμβάνονται» από χαρακτηριστικά της ταυτότητας (δομή προσώπου). Το μήκος μαλλιών γίνεται πρακτικά ασήμαντη παραλλαγή, οπότε η προβολή δεν διαχωρίζει εξίσου καλά το μήκος μαλλιών ( $\approx 74\%$  ακρίβεια).

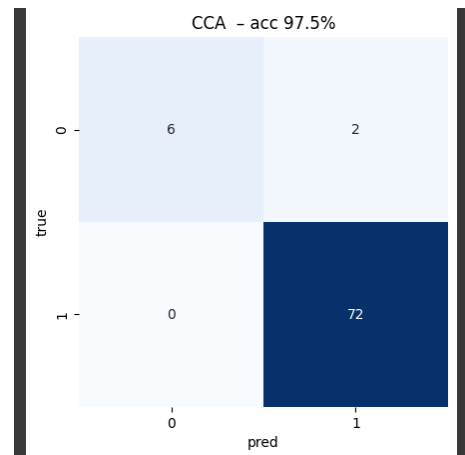
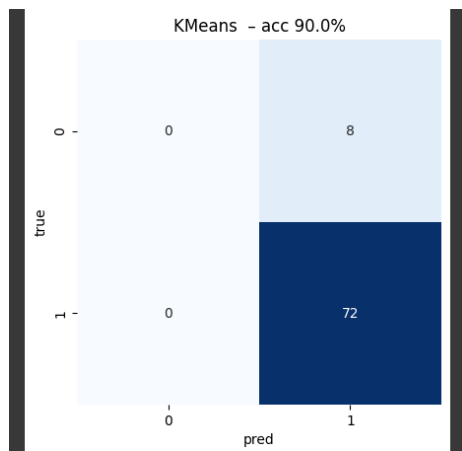
Τέλος, αυτό επίσης παρατηρείτε στο facial hair με k means. Όταν ομαδοποιούμε πρόσωπα από διαφορετικά άτομα, καταφέρνει να χωρίσει τη γενική παραλλαγή σε ξεχωριστά κέντρα και οι ομάδες αυτές γενικεύονται καλά ( $\approx 87\%$  ακρίβεια). Όμως, όταν εκτελούμε k-means στο ίδιο περιορισμένο σύνολο ταυτοτήτων, ο αλγόριθμος εστιάζει σε ιδιόμορφα χαρακτηριστικά, με αποτέλεσμα οι ομάδες να γίνονται «θολές» και η ακρίβεια να μειώνεται ( $\approx 77\%$ ).

Για όλα τα υπόλοιπα, υπάρχει σαφής βελτίωση και έχουμε αναλύσει παραπάνω για ποιον λόγο η κάθε μέθοδος είναι καλύτερη.

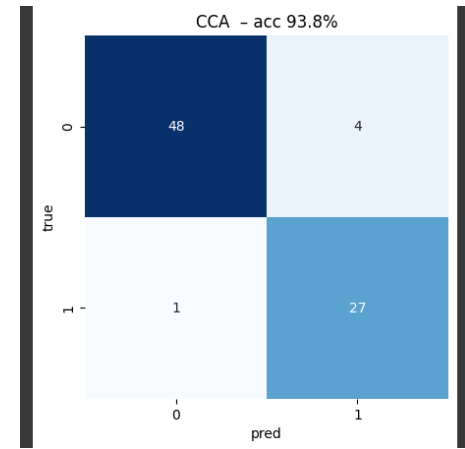
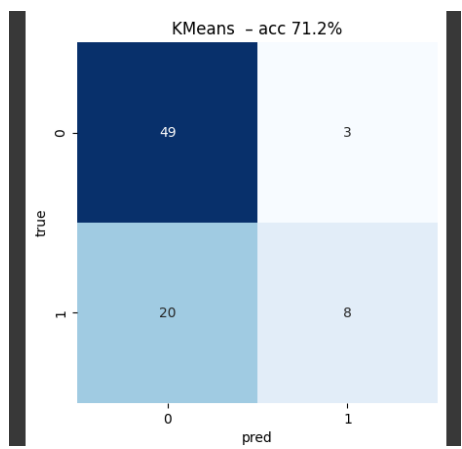
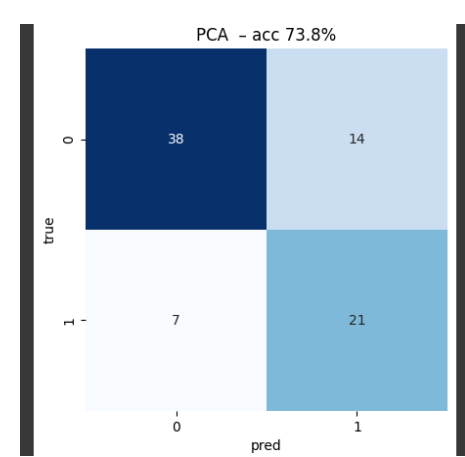
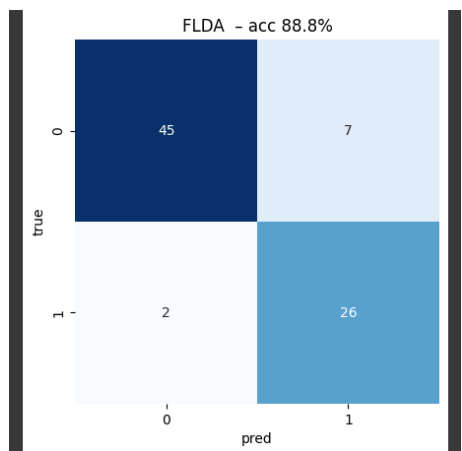
Ας δούμε τώρα το confusion matrix:

Gender:

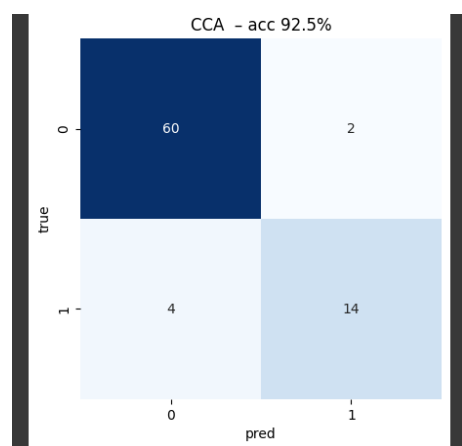
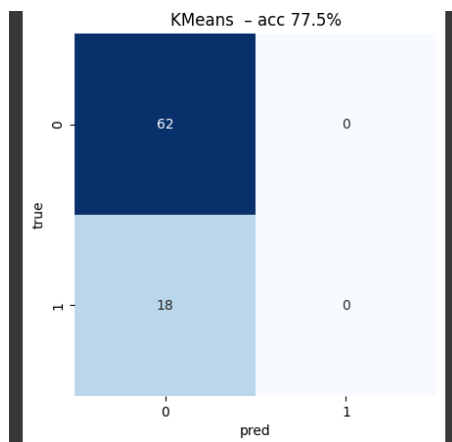
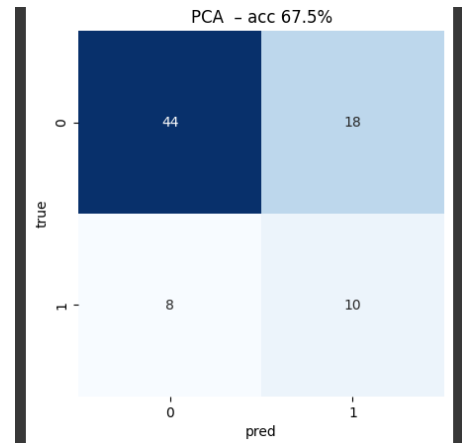
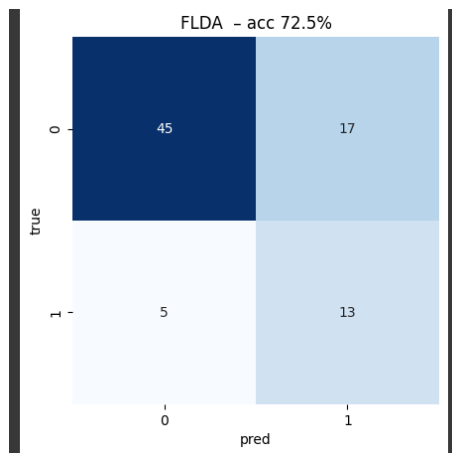




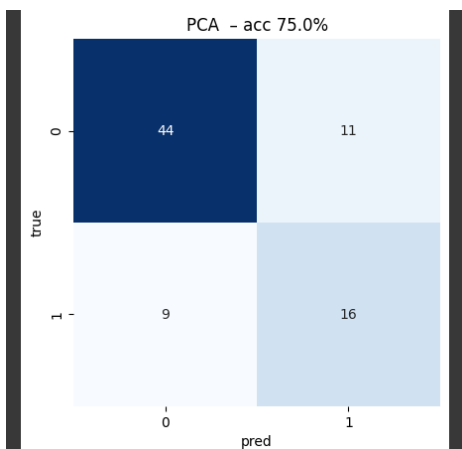
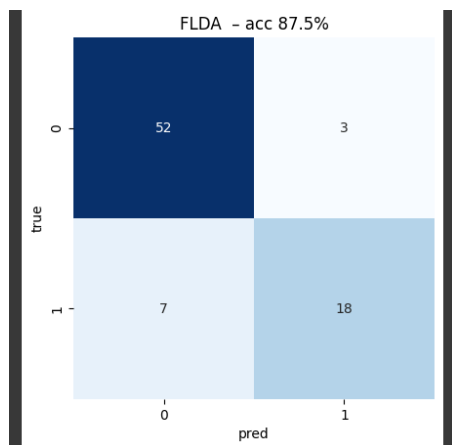
Hair long:

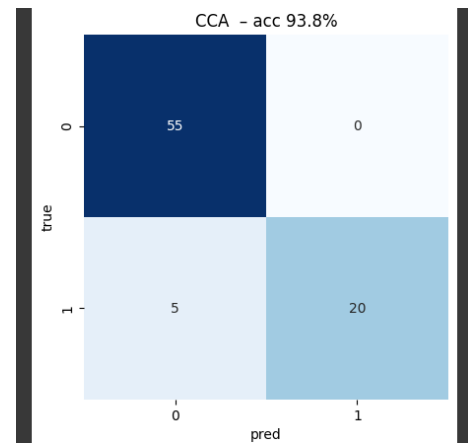
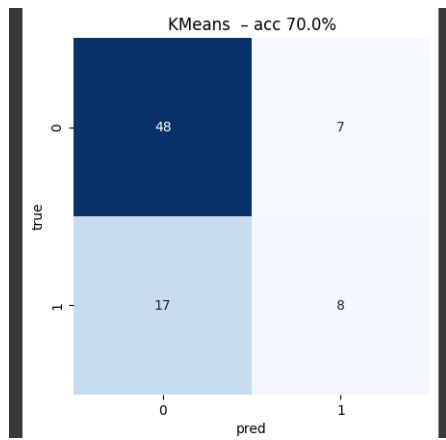


## Facial hair:

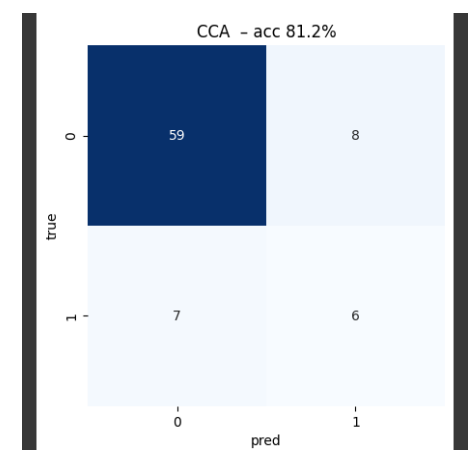
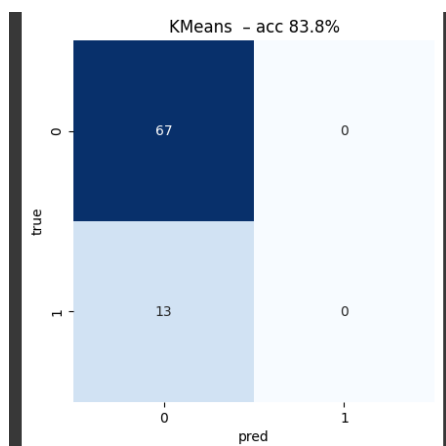
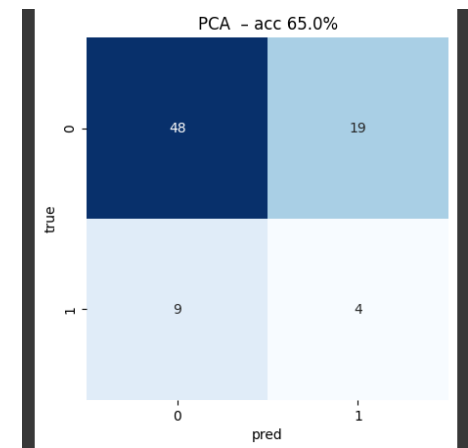
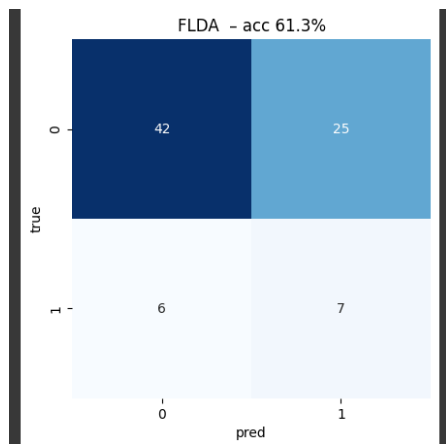


## Glasses:

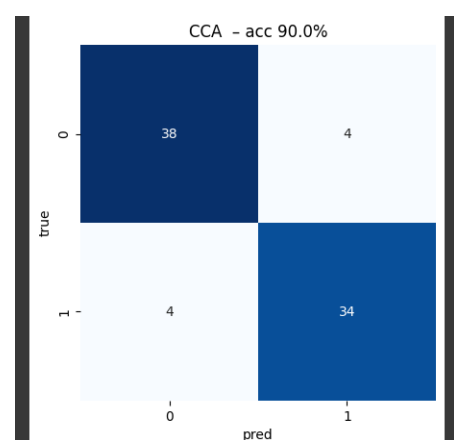
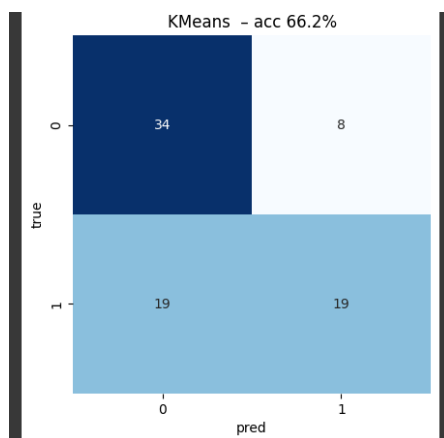
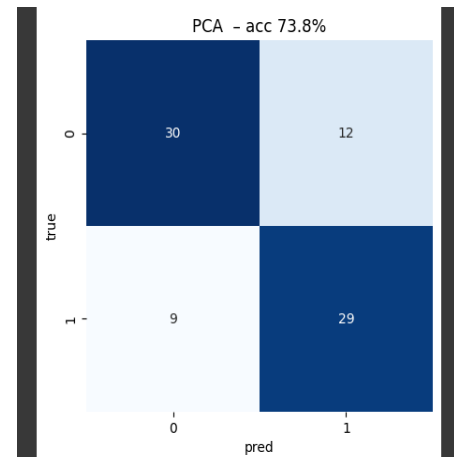
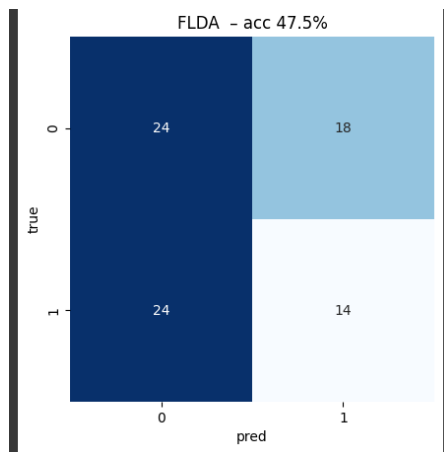




Smile:



Pose turned:



Συμπερασματικά, βλέπουμε ότι σε αυτό το σενάριο η ακρίβεια πράγματι αντικατοπτρίζεται στα αποτελέσματα που παίρνουμε. Επιπλέον, ο cca είναι ξεκάθαρα ο νικητής, αφού έχει την μεγαλύτερη ακρίβεια τις περισσότερες φορές. Όσον αφορά τις υπόλοιπες μεθόδους, έχουμε αναλύσει παραπάνω που υπερिशύει η καθεμιά.