



UNIVERSITÀ
DI TRENTO

IDENTIFICAZIONE DELLA PROVENIENZA DI IMMAGINI CONDIVISE SU SOCIAL NETWORK

ANDREA TOMASONI

Corso di Laurea in Informatica
Anno accademico 2021/2022

Supervisore
GIULIA BOATO

Correlatore
SEBASTIANO VERDE

SOCIAL MEDIA FORENSICS

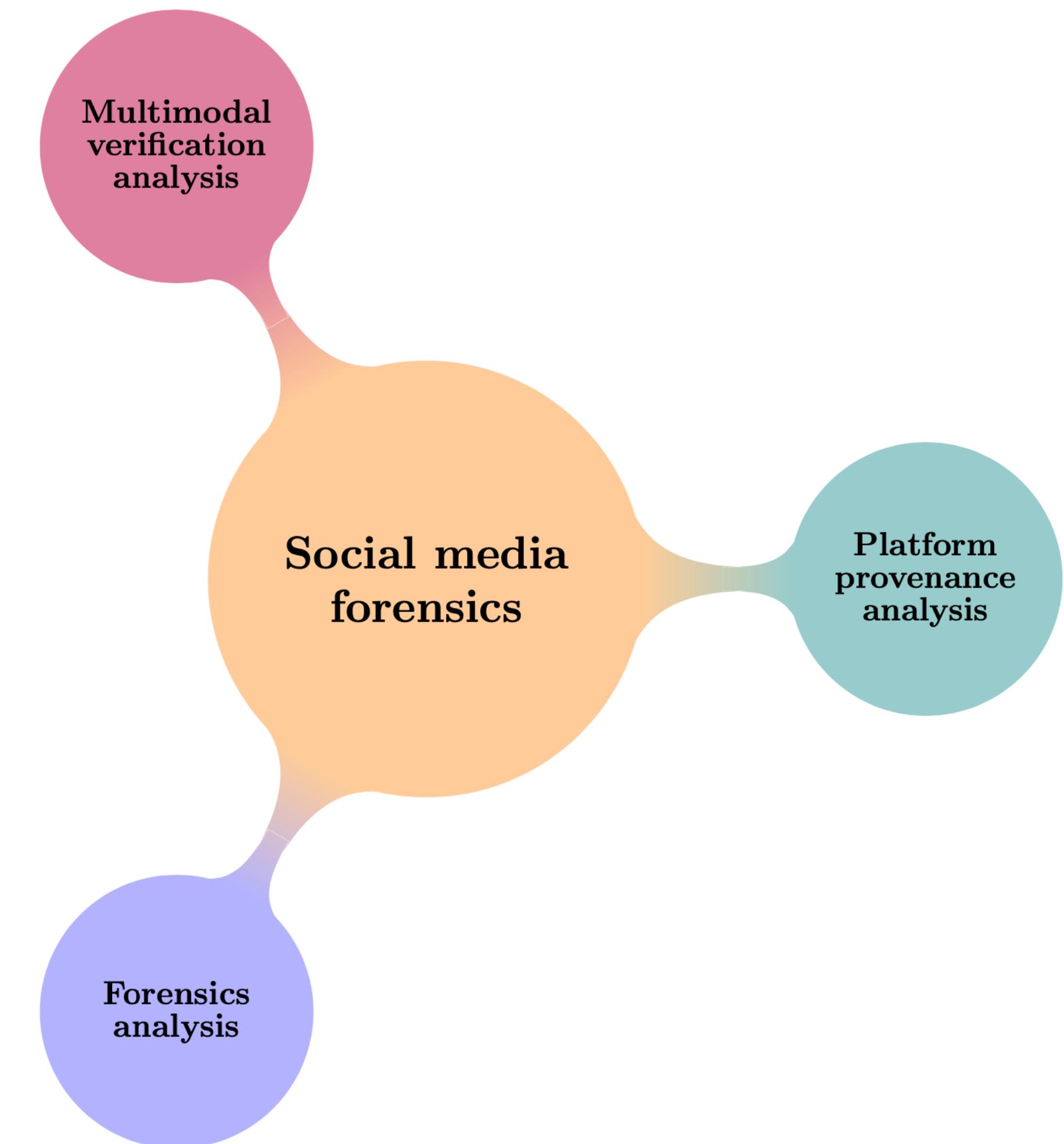
STUDIO SCIENTIFICO SULLA RACCOLTA, L'ANALISI E
L'INTERPRETAZIONE DI AUDIO, VIDEO E IMMAGINI CONDIVISI SU
UNO O PIÙ DISPOSITIVI E UNO O PIÙ SOCIAL NETWORK

OBIETTIVI:

- MIGLIORARE LA COMPRENSIONE DEL PERCORSO SVOLTO NELLA RETE DA UN OGGETTO DIGITALE
- IDENTIFICARE E CONTRASTARE POTENZIALI MANOMISSIONI E MODIFICHE ILLICITE EFFETTUATE SUI CONTENUTI CONDIVISI

CAMPI DI APPLICAZIONE:

- FORENSICS ANALYSIS
- MULTIMODAL VERIFICATION ANALYSIS
- PLATFORM PROVENANCE ANALYSIS



PLATFORM PROVENANCE ANALYSIS

PUNTA AD IDENTIFICARE LA STORIA DIGITALE ASSOCIATA AD UN CONTENUTO PER:

- RISALIRE ALLE PIATTAFORME CHE L'HANNO PROCESSATO
- RICOSTRUIRE L'INTERA CATENA DI CONDIVISIONI EFFETTUATE
- ESTRARRE INFORMAZIONI RILEVANTI SUI SISTEMI UTILIZZATI DURANTE LA FASE DI CARICAMENTO

IL PROCESSO DI STUDIO È COMPOSTO DA TRE FASI ESEGUITE IN MODO SEQUENZIALE:



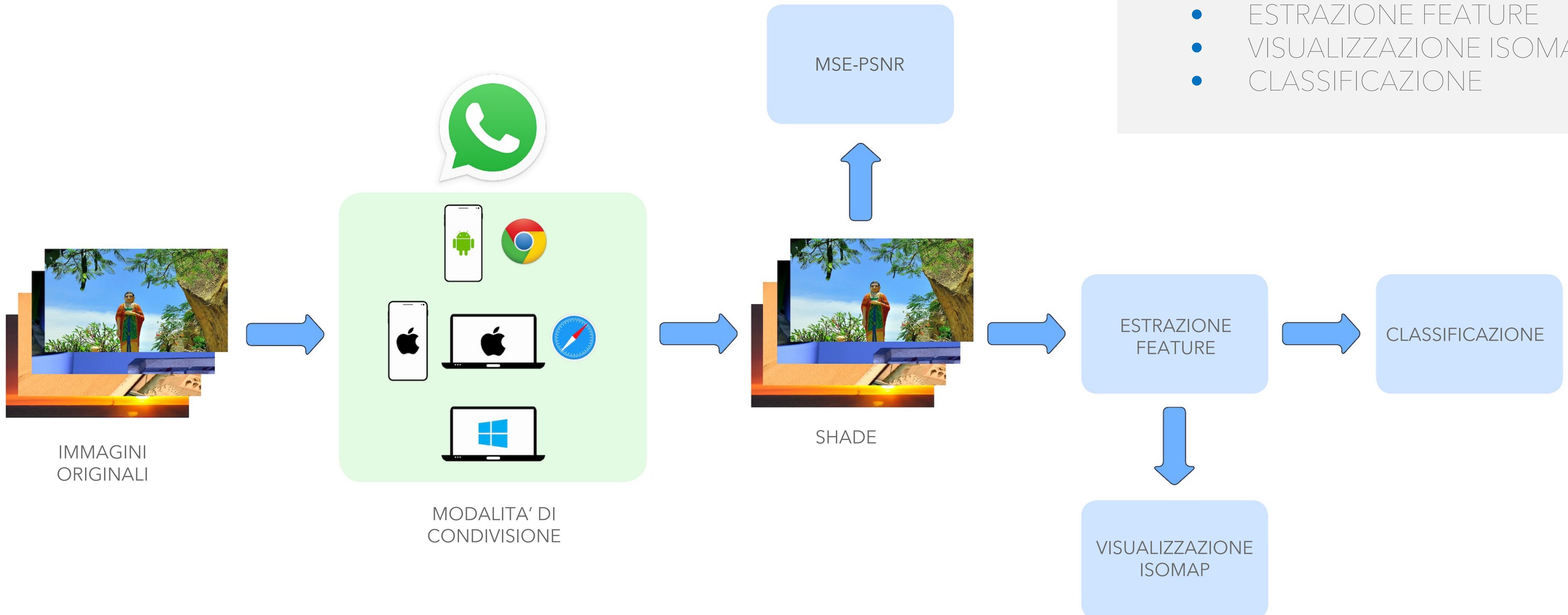
SCOPO TESI:

VERIFICARE ATTRAVERSO QUESTO PROCEDIMENTO SE È POSSIBILE IDENTIFICARE I DISPOSITIVI
E I SISTEMI OPERATIVI IMPIEGATI PER LA CONDIVISIONE DI IMMAGINI SU WHATSAPP

METODOLOGIA SEGUITA

5 PUNTI FONDAMENTALI:

- CREAZIONE SHADE
- CALCOLO MSE E PSNR
- ESTRAZIONE FEATURE
- VISUALIZZAZIONE ISOMAP
- CLASSIFICAZIONE



SHADE (SHARING DEVICE)

È COSTITUITO DA IMMAGINI CONDIVISE TRAMITE WHATSAPP
ATTRAVERSO MOLTEPLICI INTERFACCE SOFTWARE PER
DIVERSI SISTEMI OPERATIVI

LE 50 IMMAGINI DI PARTENZA OTTENUTE DAL DATASET RAISE
SONO STATE SOTTOPOSTE A DUE OPERAZIONI:

- RIDIMENSIONATE PER OTTENERE TRE COPIE CON
RISOLUZIONE PARI A:
337 x 600
1012 x 1800
1687 x 3000
CON UN ASPECT RATIO DI 9:16
- COMPRESSE IN FORMATO JPEG UTILIZZANDO 6 QUALITY
FACTOR DIFFERENTI (QF = 50, 60, 70, 80, 90, 100)

METODI DI CONDIVISIONE:

APPLICAZIONE MOBILE PER ANDROID
APPLICAZIONE DESKTOP PER MACOS
APPLICAZIONE DESKTOP PER WINDOWS 10
APPLICAZIONE MOBILE PER IOS
BROWSER (SAFARI) PER IPADOS
BROWSER (SAFARI) PER MACOS
BROWSER (CHROME) PER WINDOWS 10

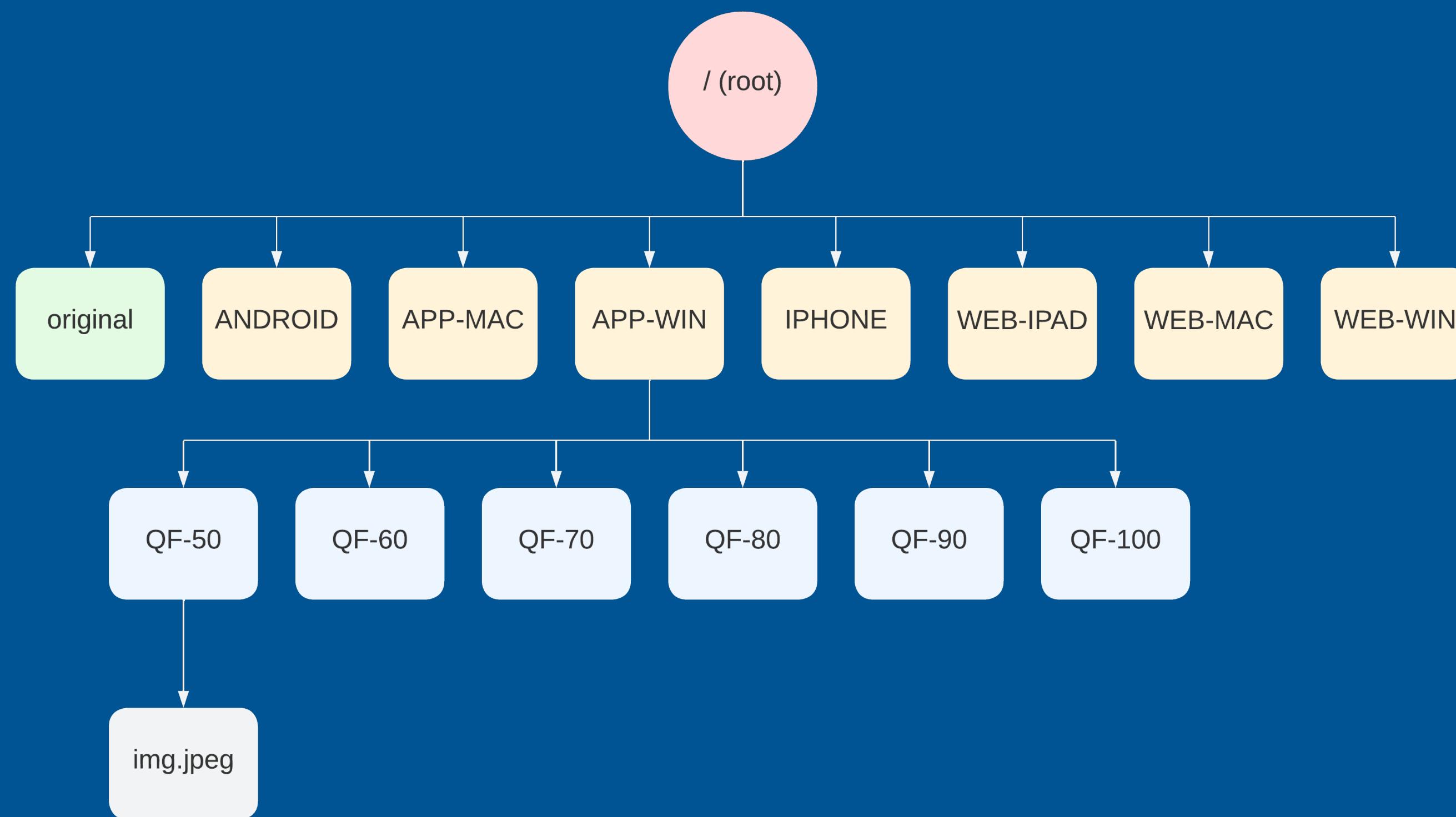


SHADE (SHARING DEVICE)

IN TOTALE SONO STATE OTTENUTE
 $50 \times 3 \times 6 = 900$ IMMAGINI PER LA
CONDIVISIONE

OGNI IMMAGINE CONDIVISA 7 VOLTE
ATTRaverso i METODI SCELTI PIÙ
ALTRE 6 VOLTE PER I DIVERSI LIVELLI DI
COMPRESsione

DATASET FINALE COSTITUITO DA
 $150 \times 7 \times 6 = 6300$ ELEMENTI A CUI
VENGONO AGGIUNTE LE 900
IMMAGINI ORIGINALI



RISULTATI

ALCUNE DELLE IMMAGINI CONDIVISE PRESENTANO DIFFERENZE NELLA RISOLUZIONE:

- GLI ELEMENTI CON RISOLUZIONE PARI A 1012×1800 E 1687×3000 SONO STATI RIDIMENSIONATI A 899×1600
- GLI ELEMENTI CON RISOLUZIONE PARI A 337×600 NON SONO STATI MODIFICATI

SI SONO VERIFICATI DUE CASI PARTICOLARI:

ANDROID
IPHONE

IMMAGINI RIDOTTE A 1151×2048
IMMAGINI RIDOTTE A 576×1024

- ASPECT RATIO INVARIATO A 9:16



CALCOLO MSE - PSNR

LE IMMAGINI CONTENUTE IN SHADE HANNO SUBITO UN'ANALISI PRELIMINARE ATTRAVERSO L'USO DI DUE METRICHE:

MEAN SQUARE ERROR (MSE)

PEAK SIGNAL-TO-NOISE RATIO (PSNR)

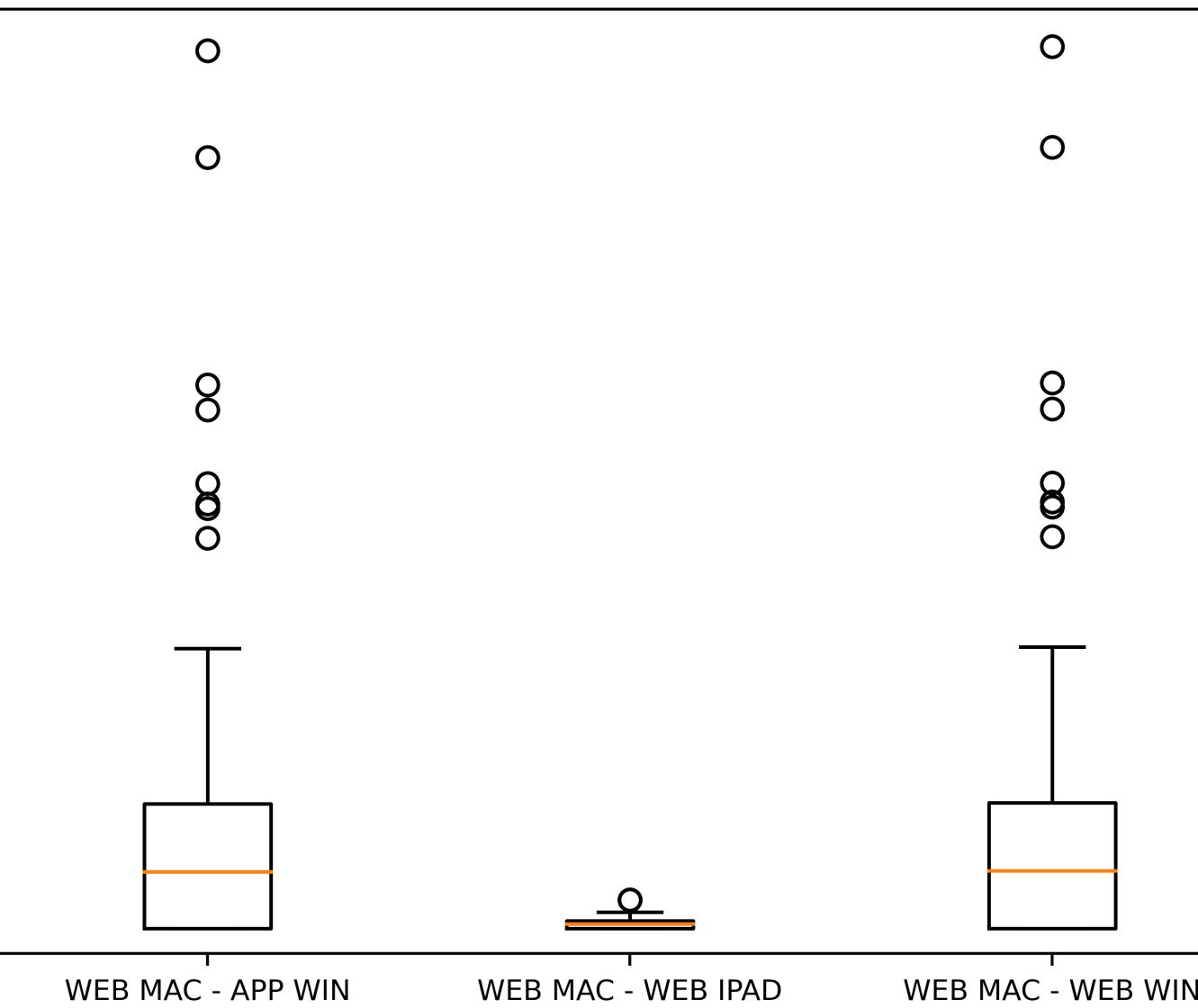
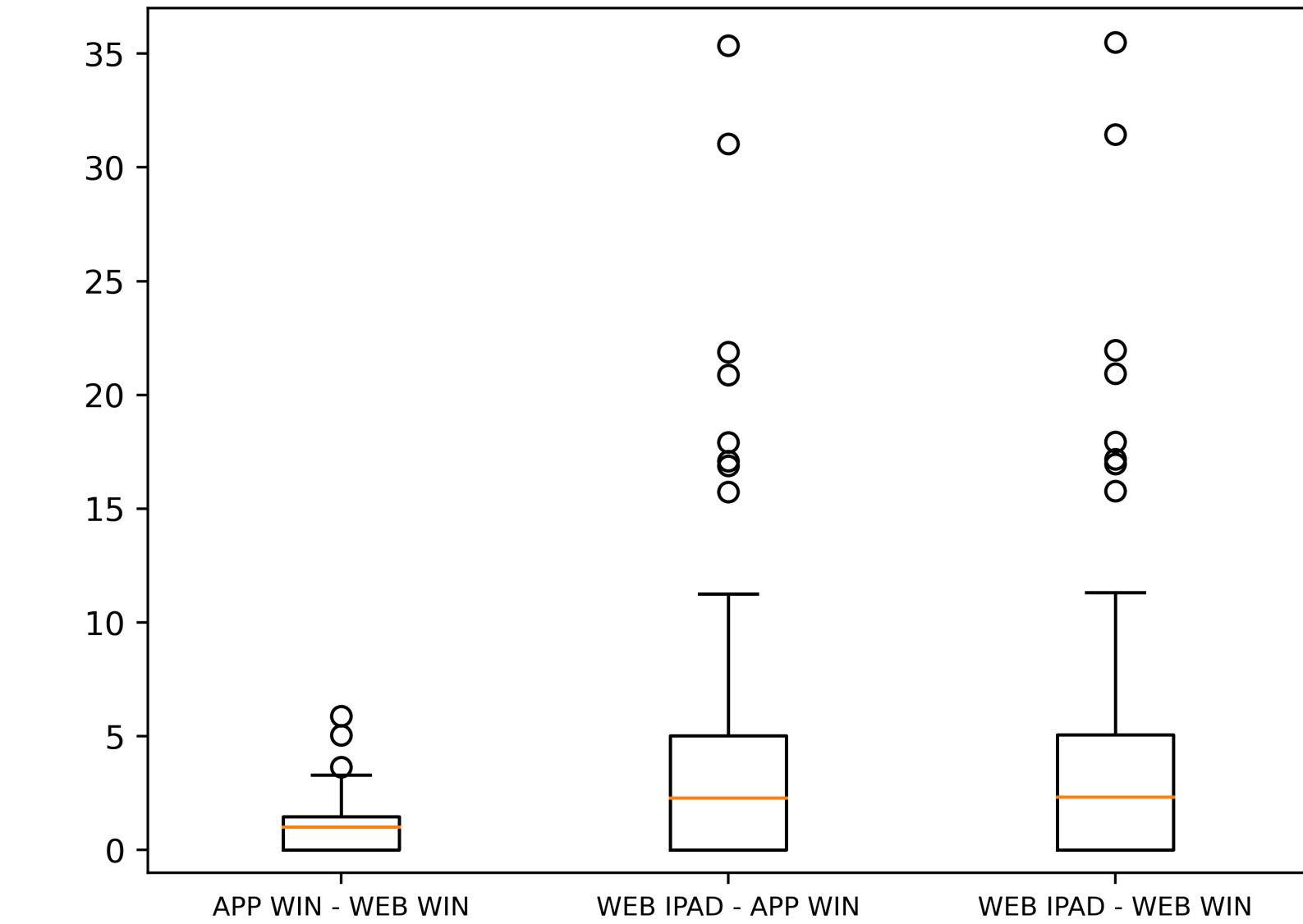
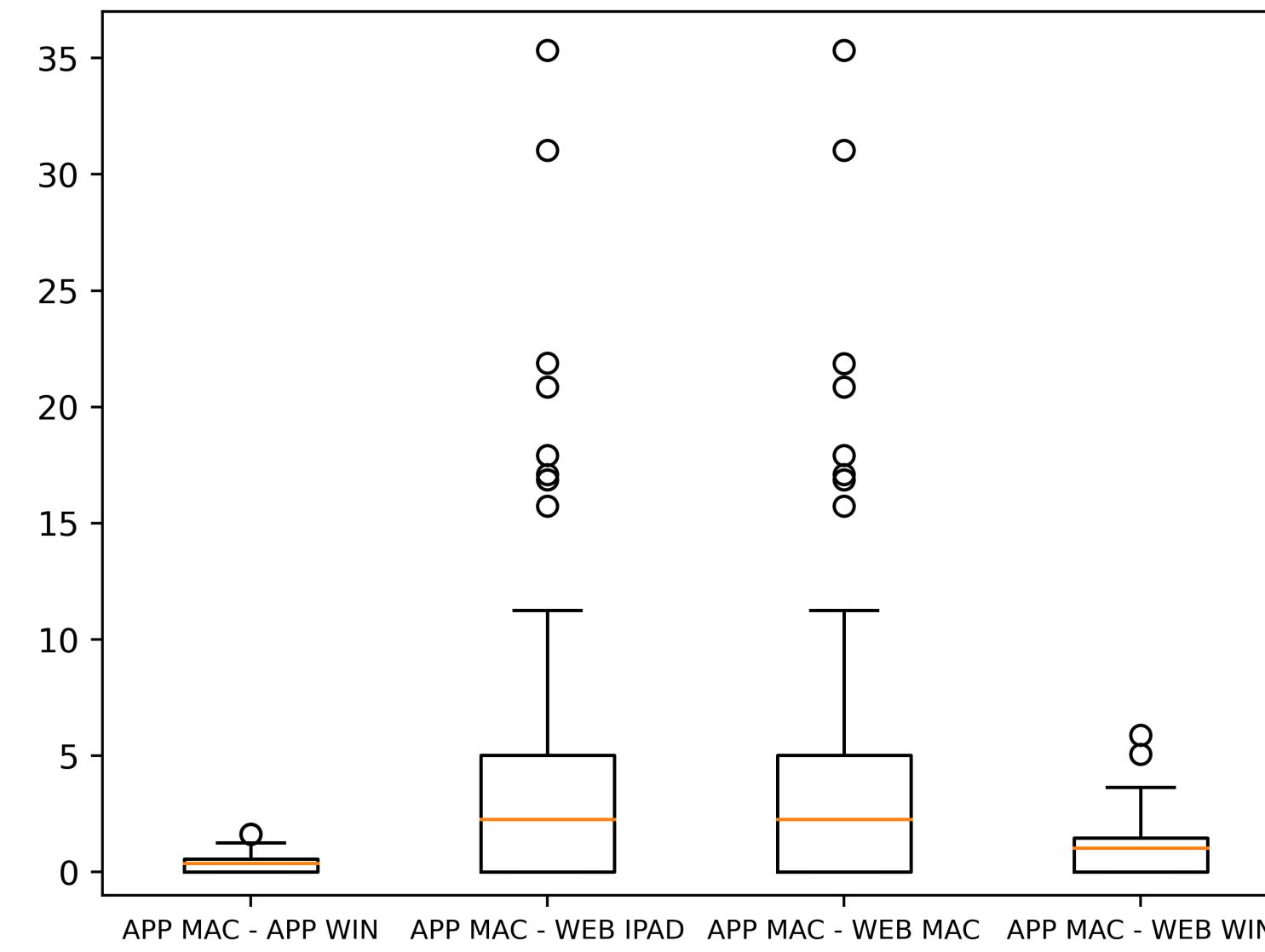
PER IL CALCOLO SONO STATE UTILIZZATE IMMAGINI CON IL LIVELLO DI COMPRESSIONE PIÙ ALTO (QF = 100)

COPPIE ANALIZZATE:

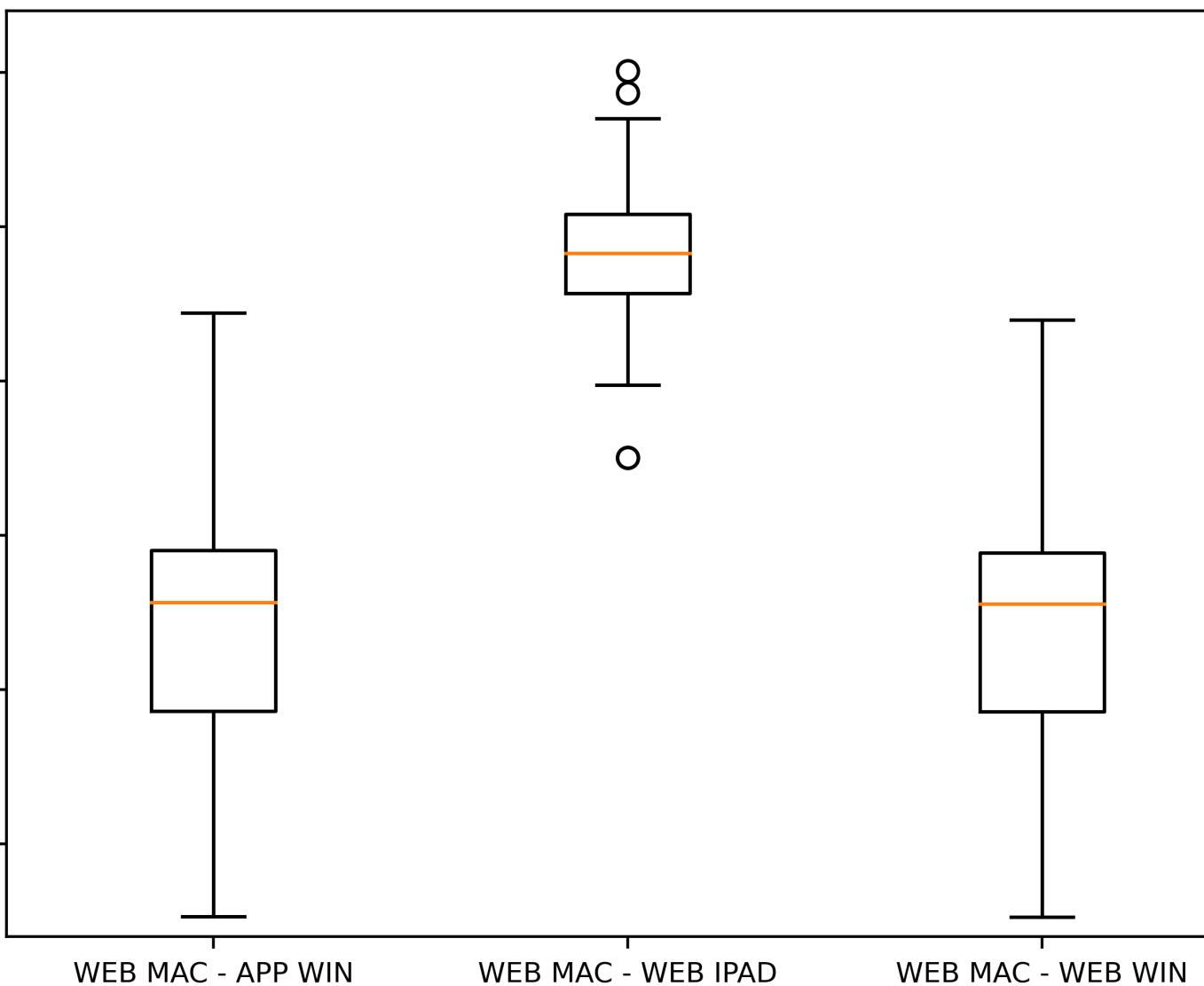
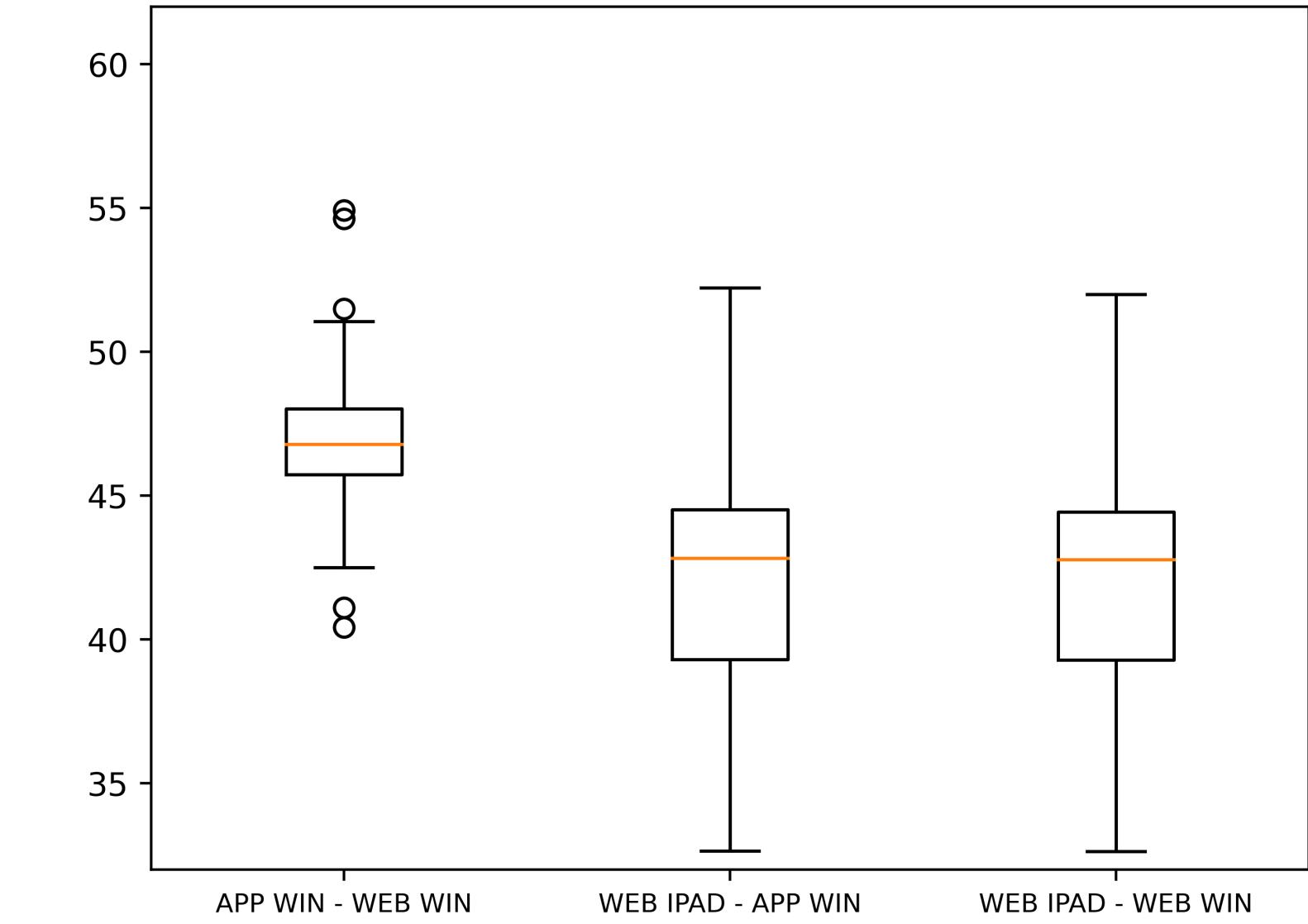
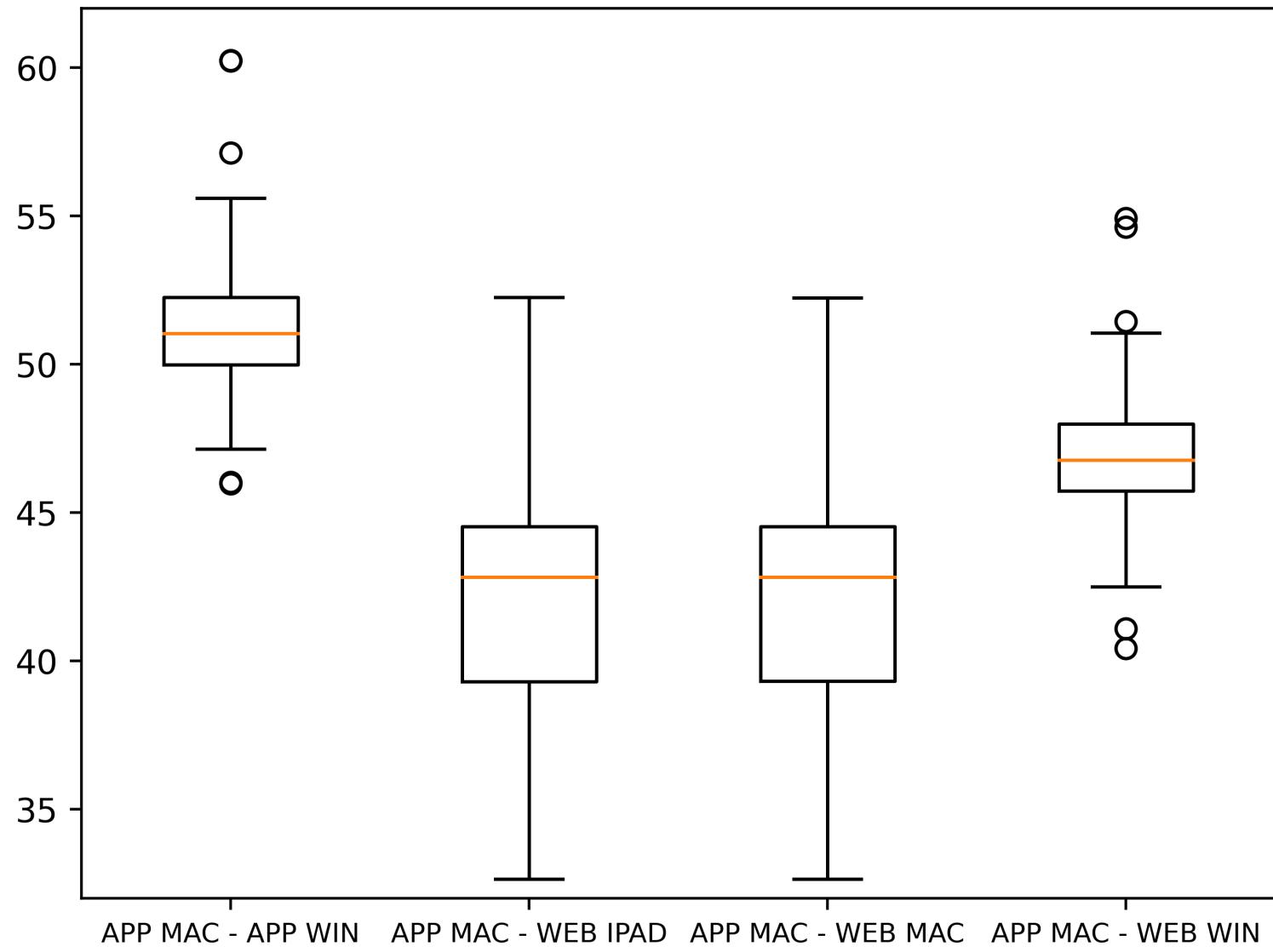
- (APP-MAC, APP-WIN)
- (APP-MAC, WEB-IPAD)
- (APP-MAC, WEB-MAC)
- (APP-MAC, WEB-WIN)
- (APP-WIN, WEB-WIN)
- (WEB-IPAD, APP-WIN)
- (WEB-IPAD, WEB-WIN)
- (WEB-MAC, APP-WIN)
- (WEB-MAC, WEB-IPAD)
- (WEB-MAC, WEB-WIN)



RISULTATI MSE



RISULTATI PSNR



ESTRAZIONE E VISUALIZZAZIONE FEATURE

PER POTER EFFETTUARE LA CLASSIFICAZIONE SONO STATE ESTRATTE PARTICOLARI INFORMAZIONI DALLE IMMAGINI, CHIAMATE FEATURE.

PER OGNI IMMAGINE È STATO CONSIDERATO UN INSIEME DI 3 DIVERSE TIPOLOGIE:

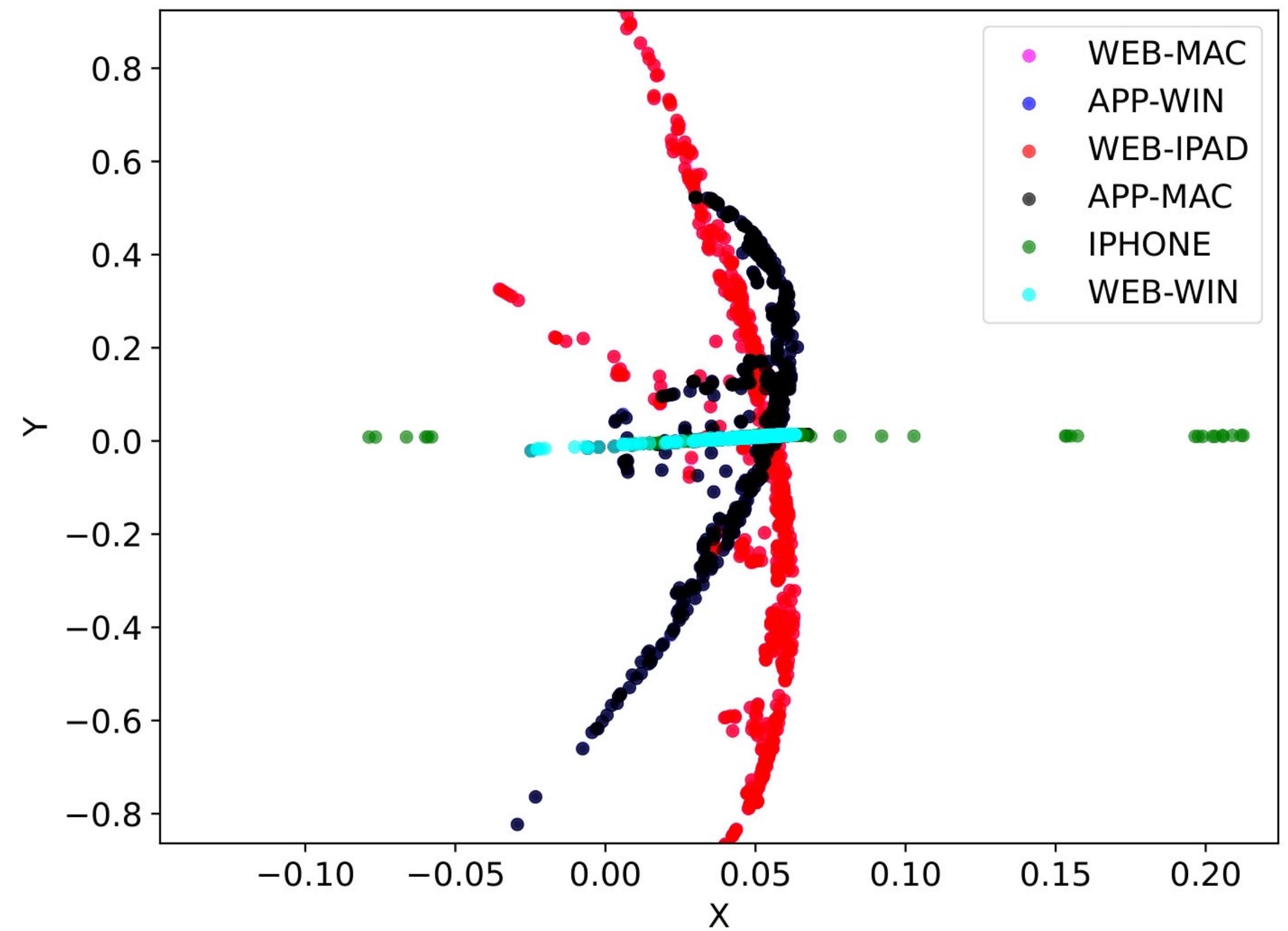
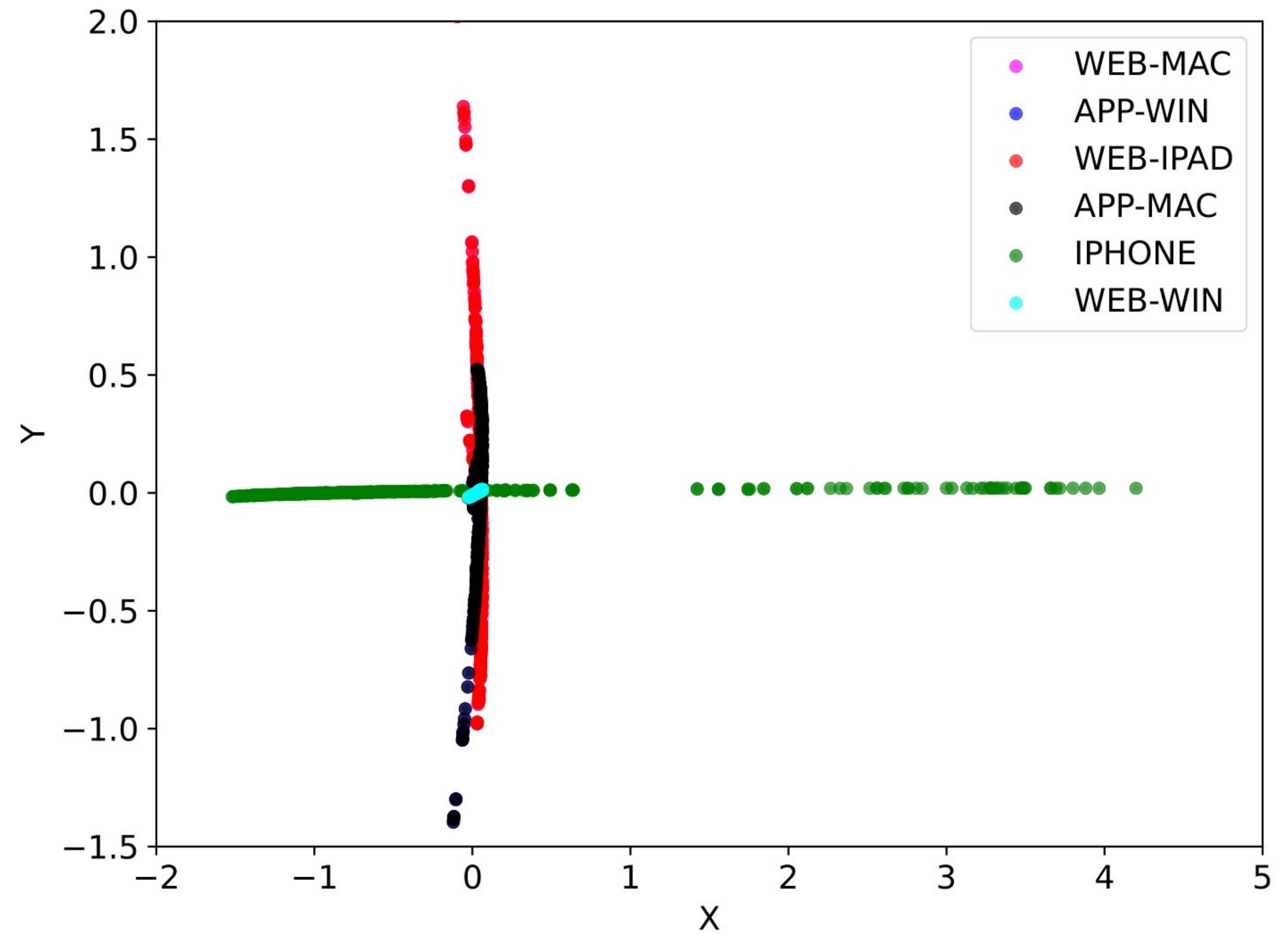
DCT (369)

HEADER (8)

META (152)

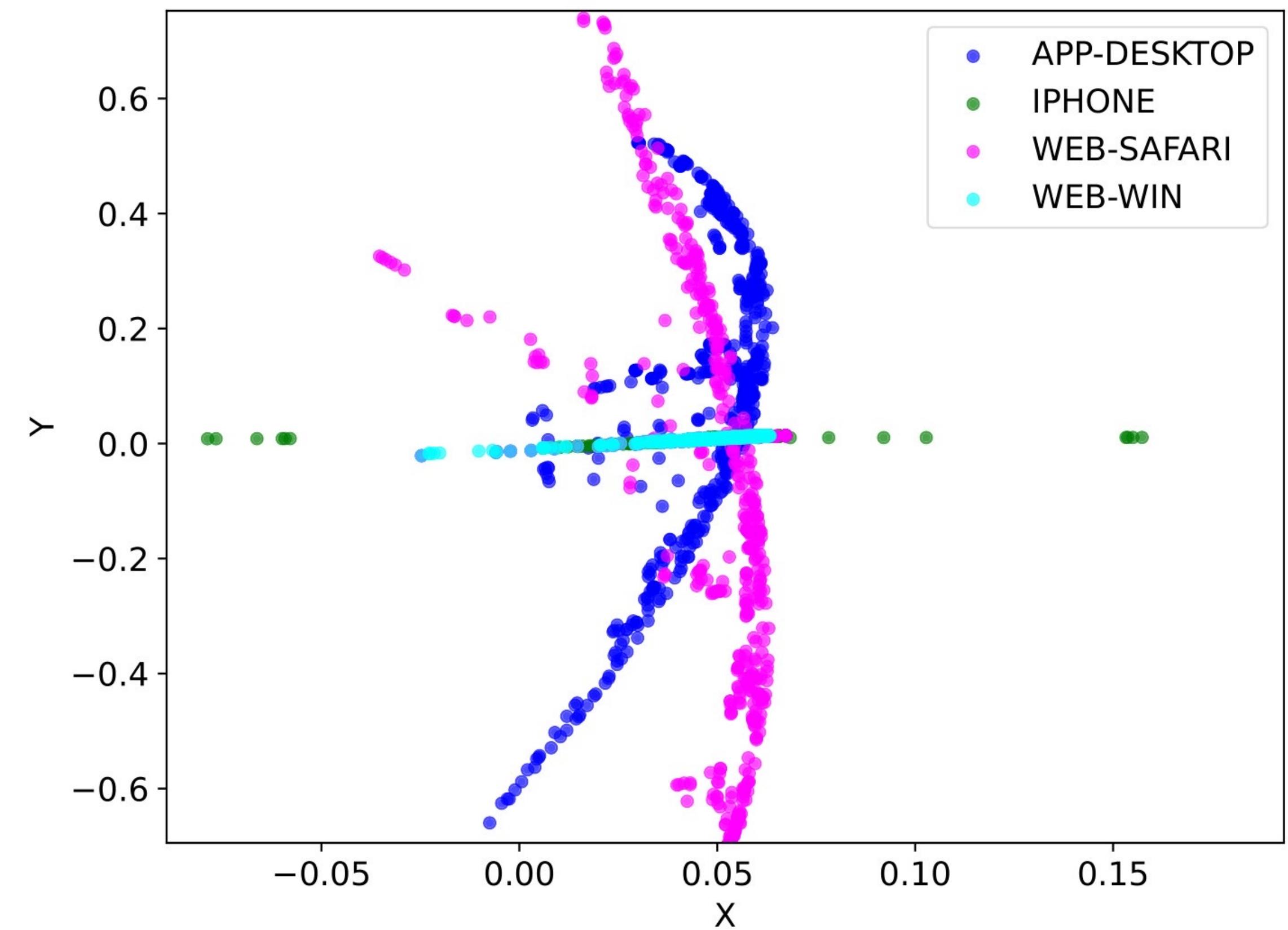
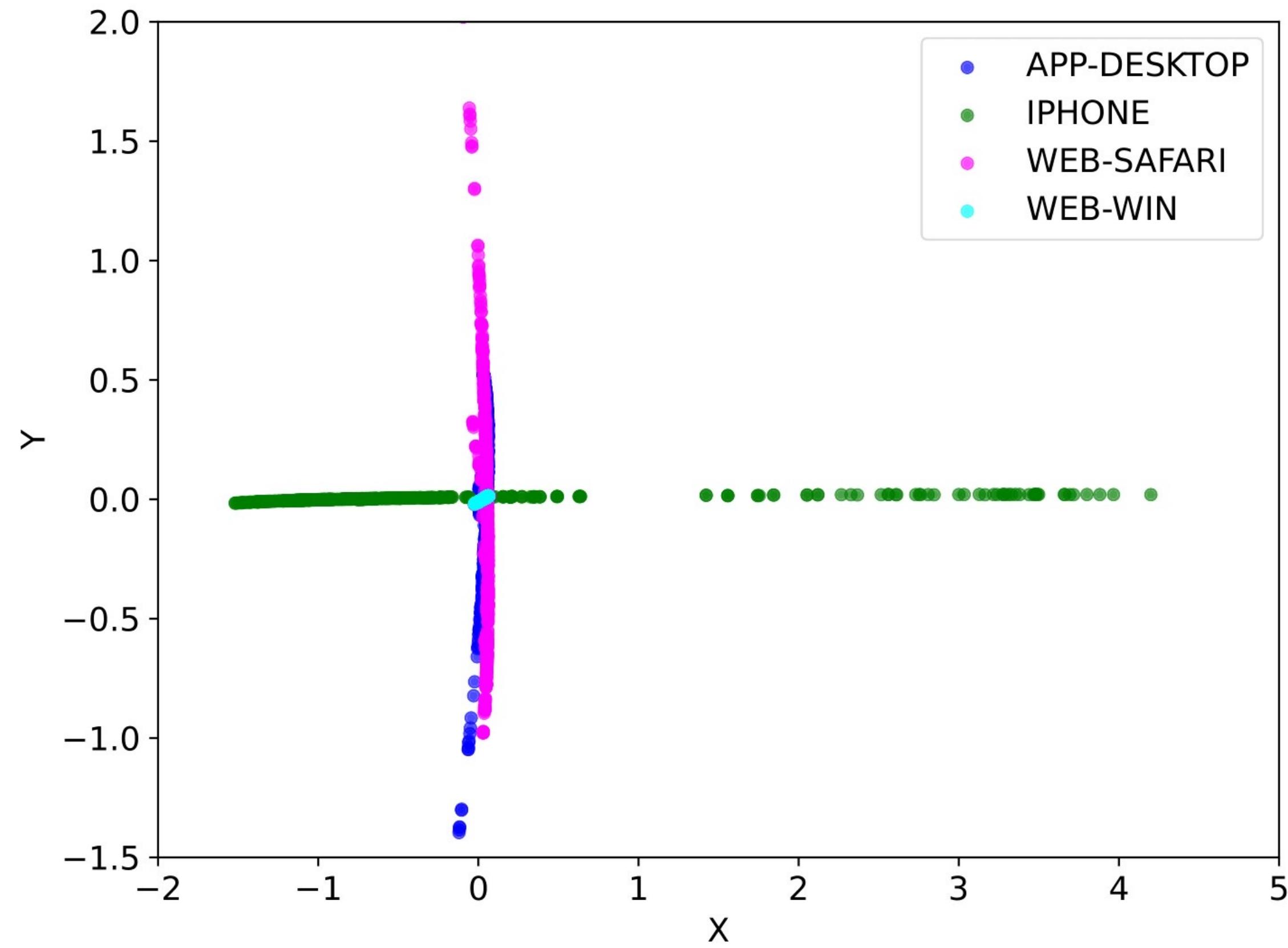
- LE FEATURE OTTENUTE SONO STATE RAGGRUPPATE IN UN UNICO VETTORE, SALVATE IN UN FILE E INSERITE IN SHADE ADOTTANDO LA STESSA STRUTTURA E NOMENCLATURA DELLE IMMAGINI
- COME ULTIMA OPERAZIONE LA **DIMENSIONALITY REDUCTION** DELLE FEATURE TRAMITE IL METODO ISOMAP E LA RELATIVA VISUALIZZAZIONE HANNO PERMESSO DI VERIFICARE IL GRADO DI SEPARAZIONE DEI METODI DI CONDIVISIONE CONSIDERATI

RISULTATI



RISULTATI CON FUSIONE DELLE CLASSI

APP-DESKTOP: APP-MAC + APP-WIN
WEB-SAFARI: WEB-IPAD + WEB-MAC



CLASSIFICAZIONE

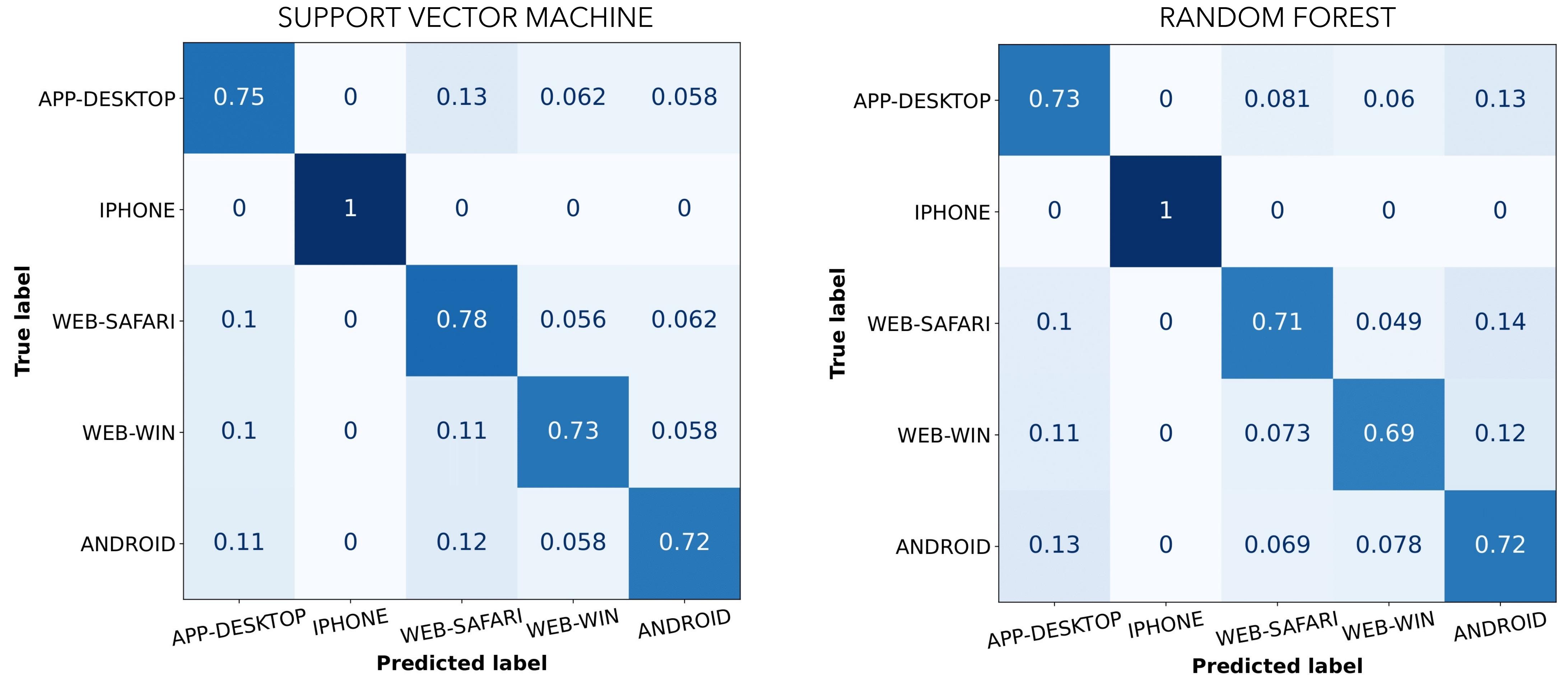
OGNI IMMAGINE È STATA CLASSIFICATA UTILIZZANDO DUE ALGORITMI DI MACHINE LEARNING:

SUPPORT VECTOR MACHINE (SVM)

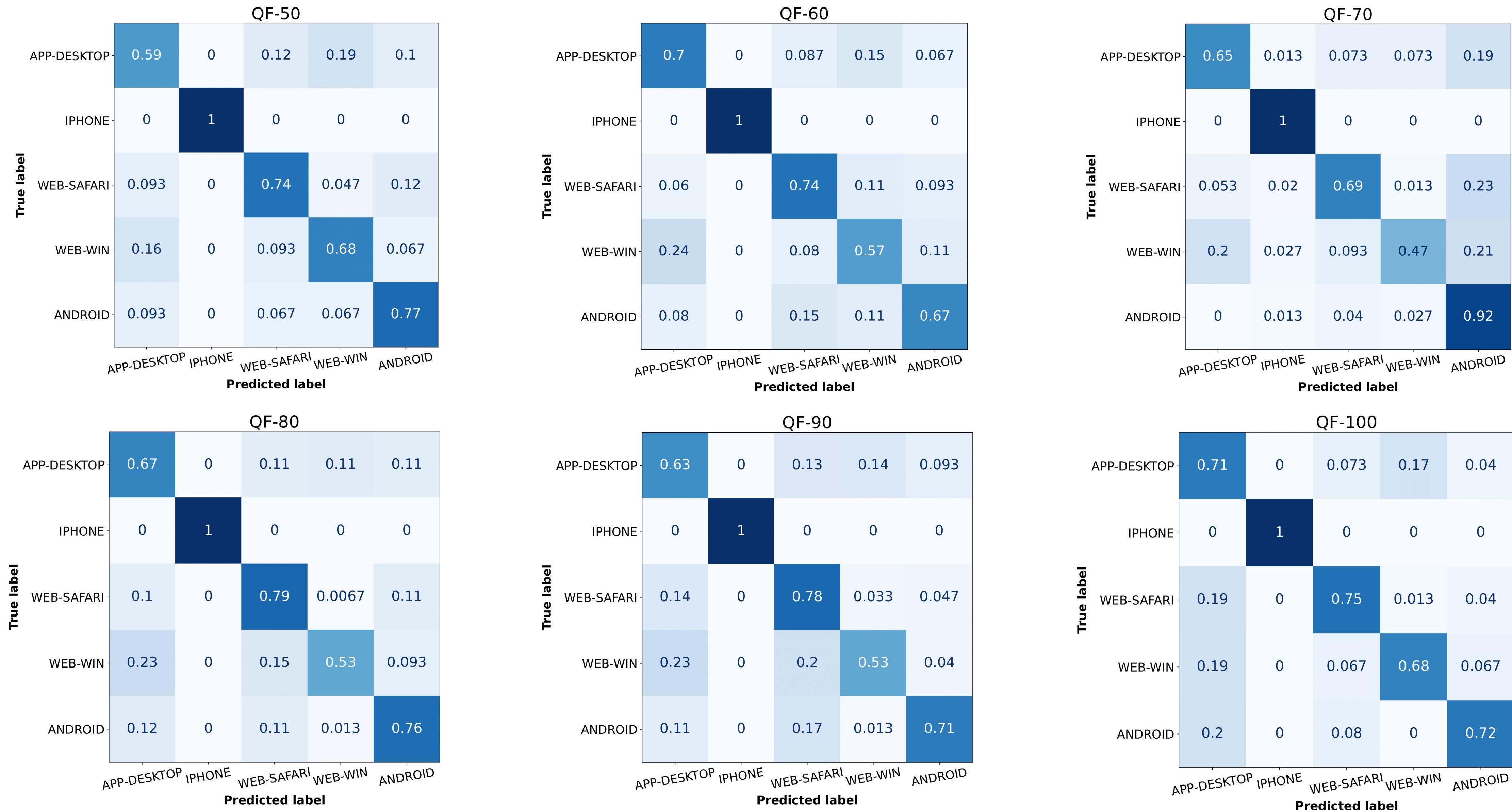
RANDOM FOREST (RF)

- I CLASSIFICATORI SONO STATI ALLENATI E TESTATI ADOTTANDO COME TRAINING SET E TEST SET LA SEGUENTE CONFIGURAZIONE:
 - ▶ 50% PER LA FASE DI TRAINING (3150 IMMAGINI)
 - ▶ 50% PER LA FASE DI TEST
- L'UTILIZZO DI SPECIALI TABELLE CHIAMATE CONFUSION MATRIX HA PERMESSO LA VISUALIZZAZIONE DELLE PERFORMANCE OTTENUTE PER I SINGOLI METODI DI CONDIVISIONE
- LA CLASSIFICAZIONE È STATA RIPETUTA CON LE STESSE MODALITÀ ANCHE PER I SINGOLI LIVELLI DI COMPRESIONE

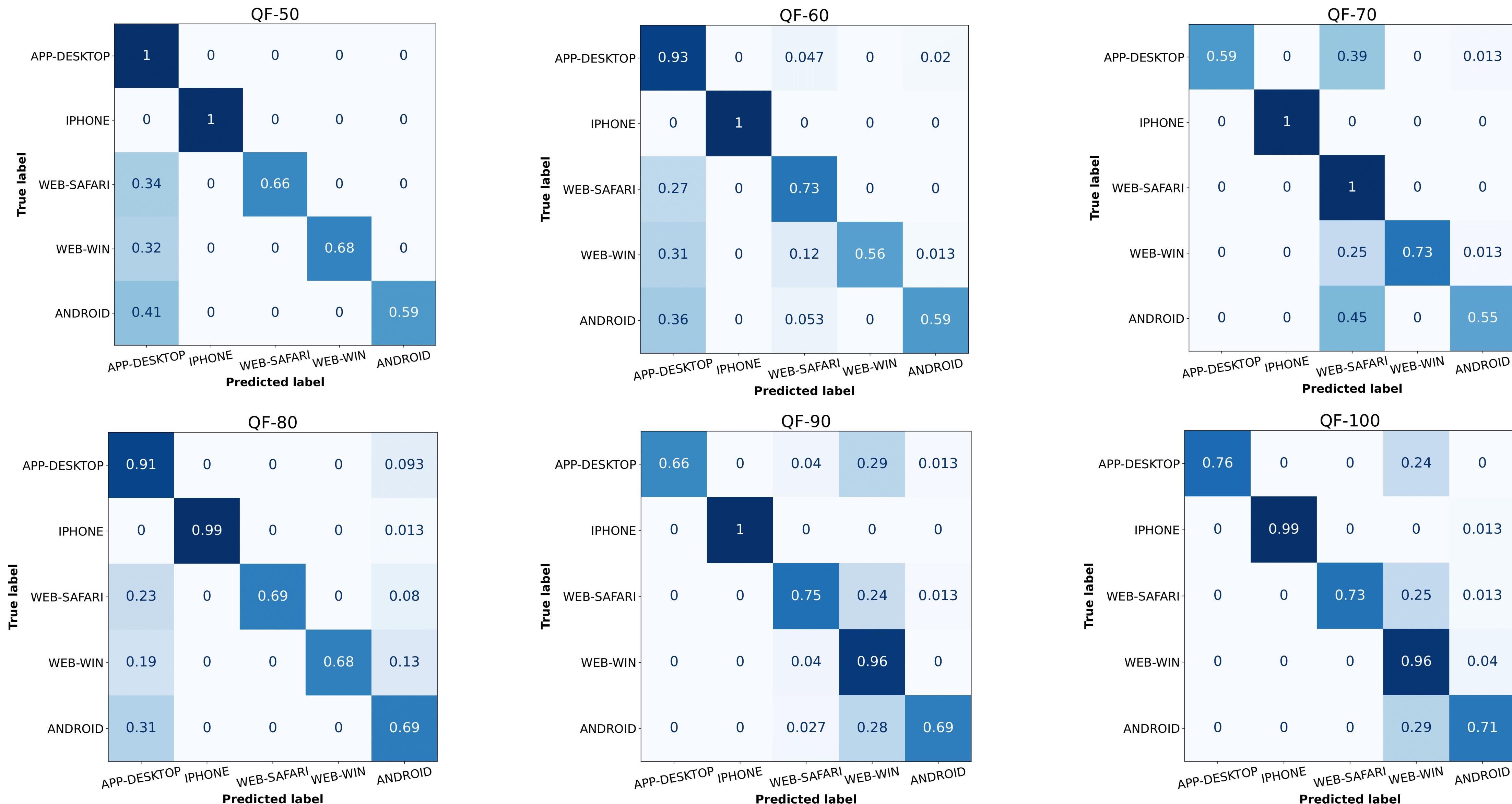
RISULTATI



CONFUSION MATRIX RANDOM FOREST



CONFUSION MATRIX SUPPORT VECTOR MACHINE



CONSIDERAZIONI FINALI

METTENDO INSIEME I RISULTATI OTTENUTI PER TUTTE LE OPERAZIONI SVOLTE (CALCOLO MSE-PSNR, ESTRAZIONE FEATURE, VISUALIZZAZIONE ISOMAP E CLASSIFICAZIONE) È POSSIBILE DISTINGUERE SE UN'IMMAGINE PROVIENE:

- DA UN DISPOSITIVO APPLE O ANDROID
- DA UN'APPLICAZIONE DESKTOP (PER MACOS o WINDOWS 10)
- DA UN BROWSER, DISTINGUENDO TRA SAFARI (PER DISPOSITIVI APPLE) E CHROME (PER WINDOWS 10)

IL LAVORO SVOLTO È STATO PUBBLICATO IN UN ARTICOLO CHE VERRÀ PRESENTATO ALLA CONFERENZA MULTIMEDIA SIGNAL PROCESSING 2022^[1]

POSSIBILI SVILUPPI FUTURI POTRANNO PRENDERE IN CONSIDERAZIONE DUE SCENARI PLAUSIBILI:

- MOLTEPLICI PIATTAFORME DI CONDIVISIONE
- IMMAGINI E VIDEO

^[1] Andrea Tomasoni, Sebastiano Verde, Giulia Boato. Sharing device identification on images from social media platforms. In 2022 IEEE 24th International Workshop on Multimedia Signal Processing (MMSP), 2022

GRAZIE DELL'ATTENZIONE

