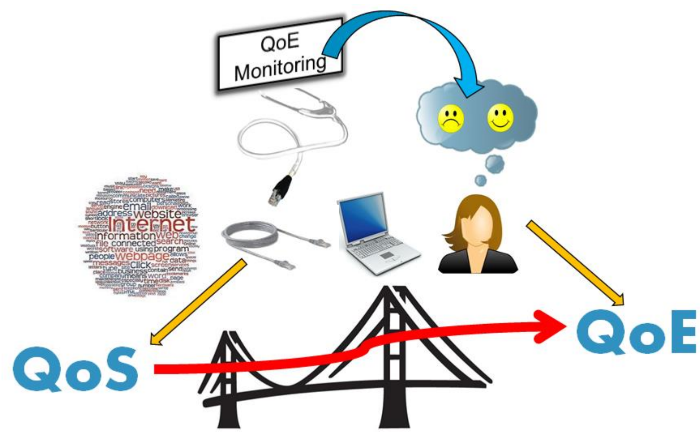
Predicting QoE in Cellular Networks using Machine Learning and in-Smartphone Measurements

P. Casas, A. D'Alconzo, F. Wamser, M. Seufert, B. Gardlo, A. Schwind,
P. Tran-Gia, R. Schatz

AIT Austrian Institute of Technology, Vienna
University of Würzburg, Institute of Computer Science, Würzburg



AGENDA

- ***Field Trial Description***
- *Monitoring Apps and Machine Learning Models*
- *QoE Assessment Results*
- *Relevant Features for QoE in Smartphones*
- *Concluding Remarks*

Field Trial Overview

- 30 users (Vienna, 2 weeks in 2015) equipped with their **own devices** & connected to their **own/preferred cellular ISPs.**
- 3 apps evaluated:
  - **YouTube →** task: watch short (~2 minutes) YouTube videos (free selection).
  - **Facebook →** task: browse fake user timeline + photo albums
  - **Google Maps →** task: browse pre-defined city maps in satellite mode
- **QoE feedback** reported through **customized QoE crowdsourcing app**
  - Overall experience (ACR MOS 1 – 5)
  - Acceptability (Yes/No)
- **Apps network traffic** monitored by **monitoring app installed in devices**

AGENDA

- *Field Trial Description*
- ***Monitoring Apps and Machine Learning Models***
- *QoE Assessment Results*
- *Relevant Features for QoE in Smartphones*
- *Concluding Remarks*

Network Traffic Monitoring & QoE Feedback Apps

- We implemented two Android apps to monitor participants' network traffic and to collect their QoE feedbacks
- **Passive flow-level traffic monitor (13 features)**
- **Web-based QoE feedback** (this paper: only overall **MOS** and **ACC**)
- Both data are merged at the **session-level** (time-synchronized)
- **The resulting 10 session-level features are used to train several ML models**

METRICS COLLECTED FOR EACH DATA FLOW.

Metric ID	Metric Name	Units	Example
1	device id (IMEI)	–	352668049725157
2	flow start time	s	1430825689
3	flow direction (up/down)	–	downlink
4	flow duration	s	10,24
5	flow size	KB	4041,00
6	avg. flow throughput	kbps	3157
7	max. flow throughput	kbps	4320,15
8	app (Android API package)	–	com.android.browser
9	signal strength	dBm	-71
10	operator (MCC.MNC)	–	295.4
11	cell id	–	16815
12	cell location (lat-lon)	deg (°)	{40,198-12,347}
13	RAT	–	LTE

flow level features

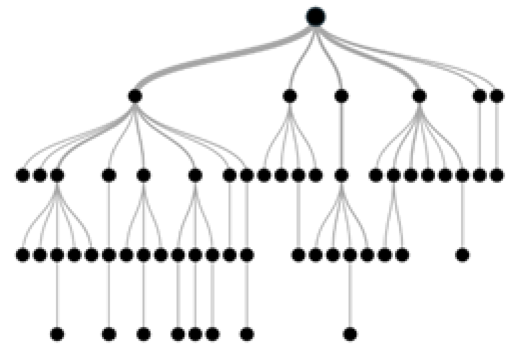
SESSION-BASED KPIS. (U) INDICATES USER-REPORTED.

KPI Name	KPI Description (U – reported by user)
MOS	overall user experience **(U)**
ACC	service acceptability **(U)**
ISP	cellular network operator
RAT	radio access technology
G	avg. signal strength
$_{max}$	max. session downlink flow throughput
TH_{avg}	avg. session downlink flow throughput
DUR	session duration
VOL	session volume
$FLOW_{ratio}$	ratio (# flows up)/(# flows down)
CELL	cell id
LOC	user location context **(U)**

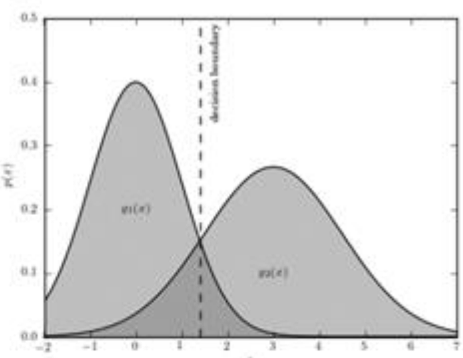
session level features

Machine Learning Models

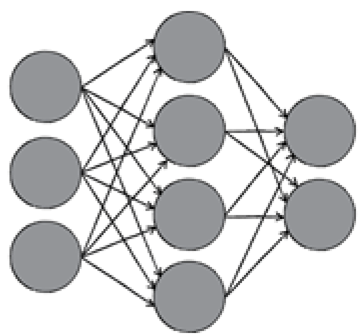
- **Supervised ML models (classification)** to predict MOS/ACC for each session
- Different algorithms **trained on field-trial labelled dataset** (10-fold cross validation)
- Benchmark **different learning models** (classifiers):
  - Bayesian Learning – Naïve Bayes (NB)
  - Neural Networks (MLP)
  - Support Vector Machines (SVM)
  - **Decision Trees (selected-model)**/Random Forest
- **WEKA used as ML library**, with extensive trial/error testing for calibration



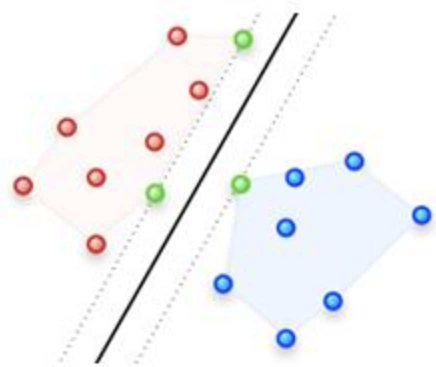
C4.5 model



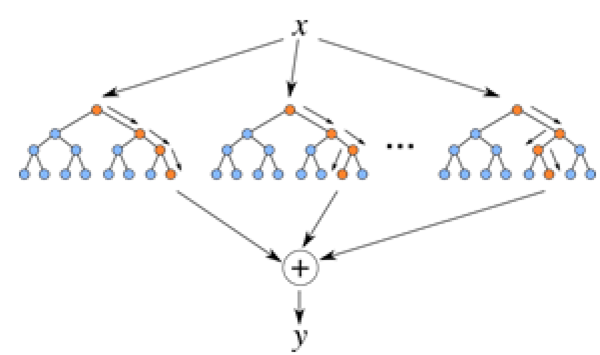
NB model



MLP model



SVM model

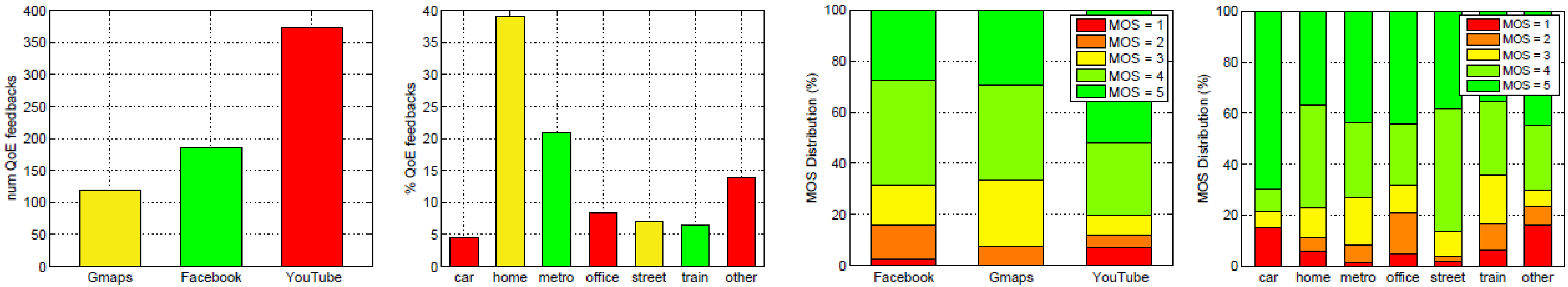


RF model

AGENDA

- *Field Trial Description*
- *Monitoring Apps and Machine Learning Models*
- ***QoE Assessment Results***
- *Relevant Features for QoE in Smartphones*
- *Concluding Remarks*

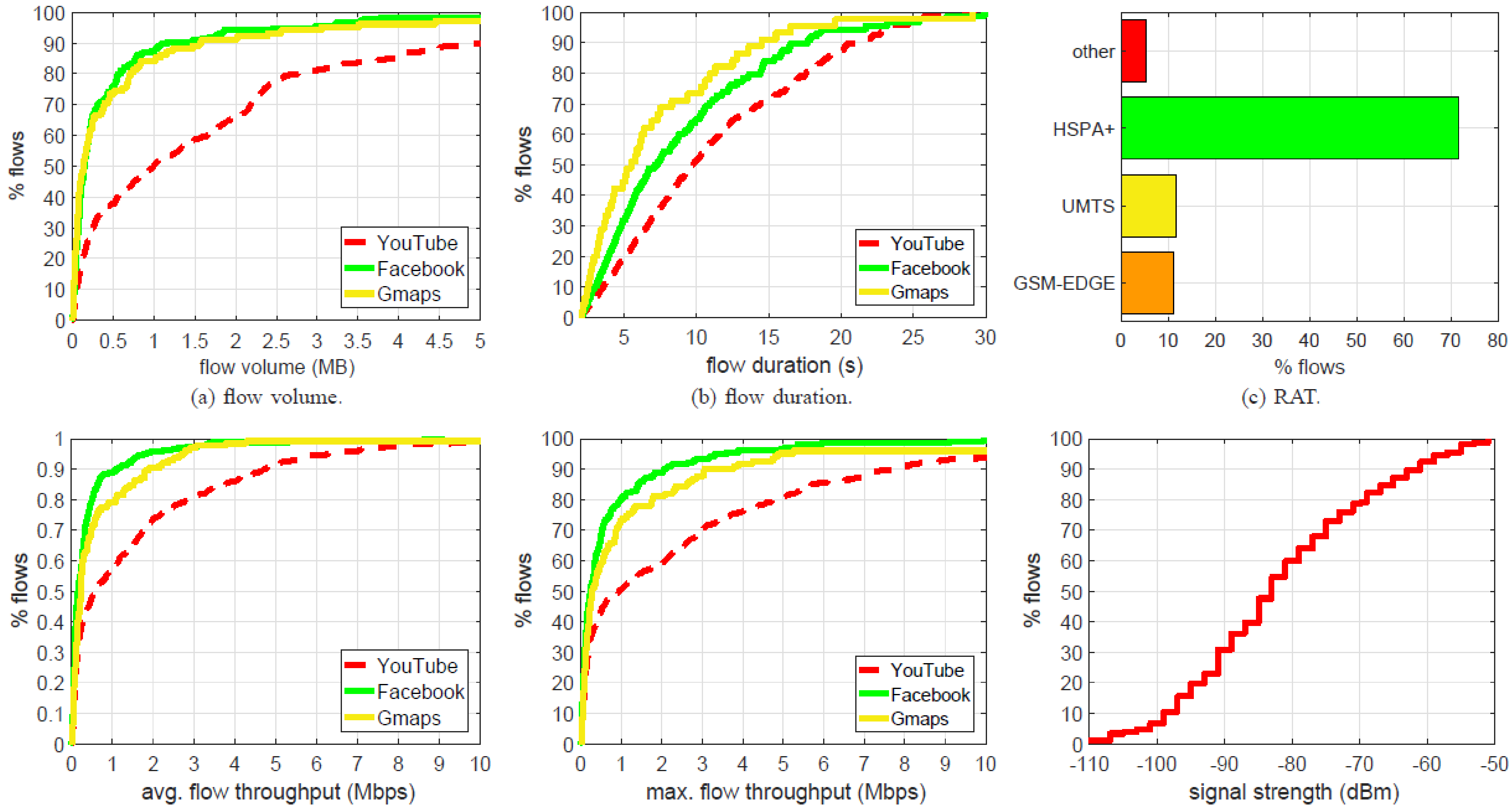
QoE Feedback Overview



(a) Ratings per App. (b) Ratings per location. (c) MOS dist. per App. (d) MOS dist. per location.

- The biggest share of ratings were done for YouTube.
- The **preferred location was home**, follow by **underground** (not surprising)
- **MOS distributions are rather similar** for the tested apps and selected locations…
- …suggesting that **network performance was rather stable** during the span of the study (**Vienna cellular connectivity is excellent**).

Collected Flow Measurements Overview

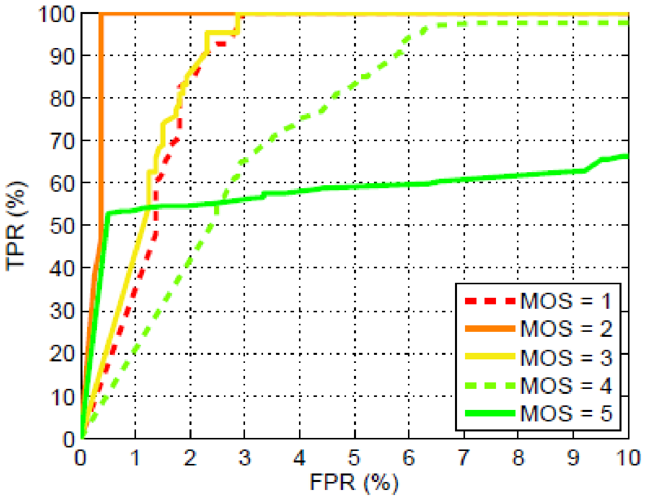


(a) flow volume. (b) flow duration. (c) RAT.

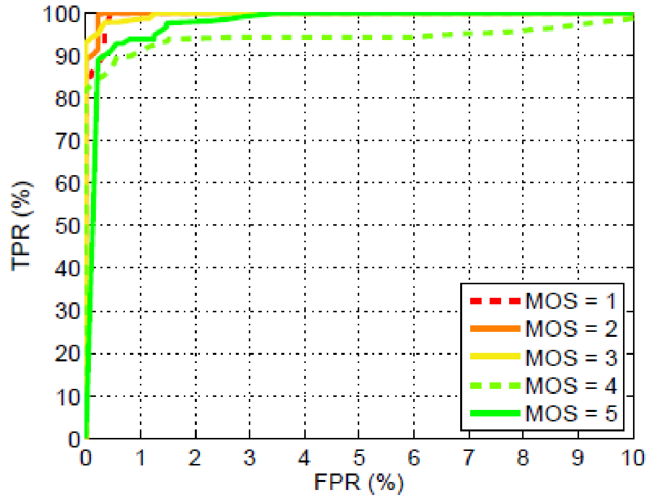
(d) avg. downlink flow throughput. (e) max. downlink flow throughput. (f) signal strength.

- Distributions of multiple flow-level measurements
- Most of the flows were transmitted with high signal strength on HSPA+

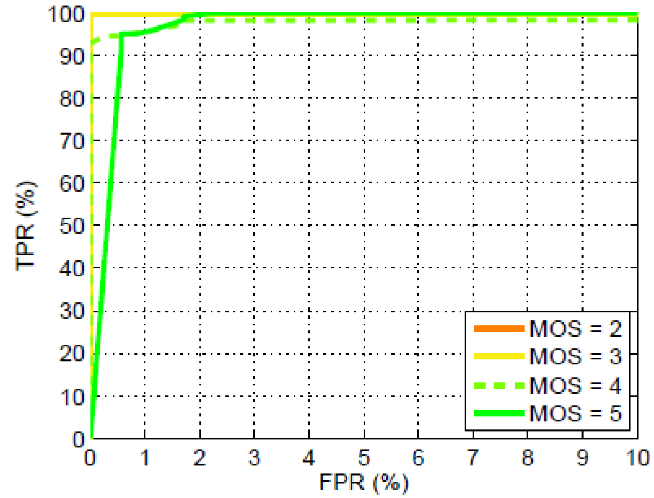
MOS Prediction – TPR vs. FPR



(a) YT MOS (C4.5).



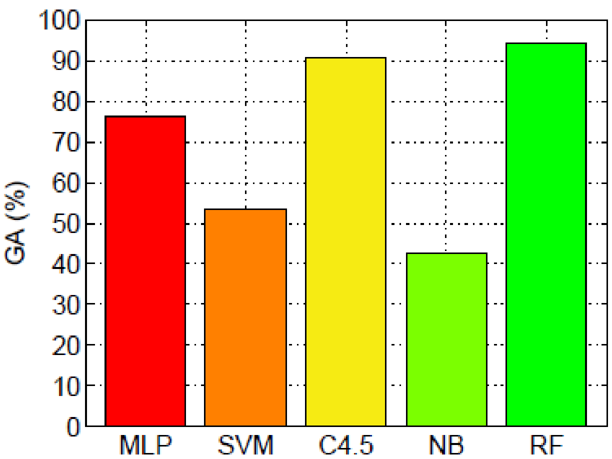
(b) FB MOS (C4.5).



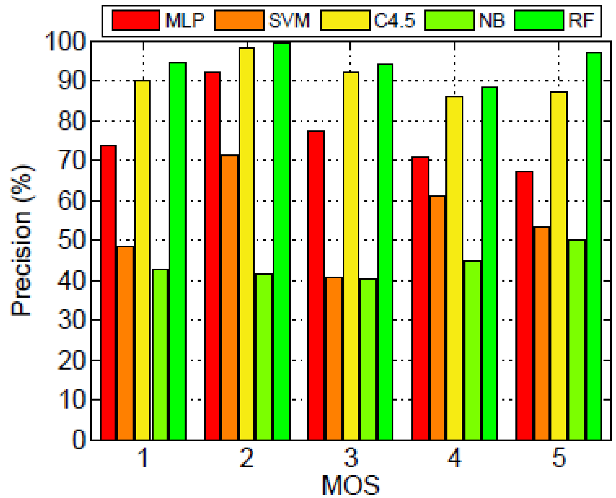
(c) Gmaps MOS (C4.5).

- The ROC curves present the **TPR vs. FPR trade-offs for the MOS prediction** on the three apps
- Prediction **results are excellent for both Facebook and Gmaps** → easier to predict QoE for these apps (at least for the considered tasks)
- YouTube results are less promising, which is expected as **network-level measurements are not enough to fully predict YouTube QoE** (e.g., buffering)
- Confusion Matrix → **MOS = 5 is mainly misclassified as MOS = 4** (but also MOS = 1 and MOS = 3)

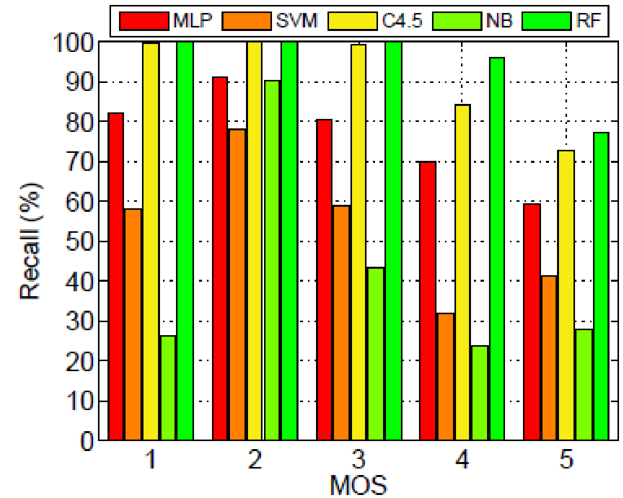
Benchmarking Different ML Models: MOS



(a) Global Accuracy.



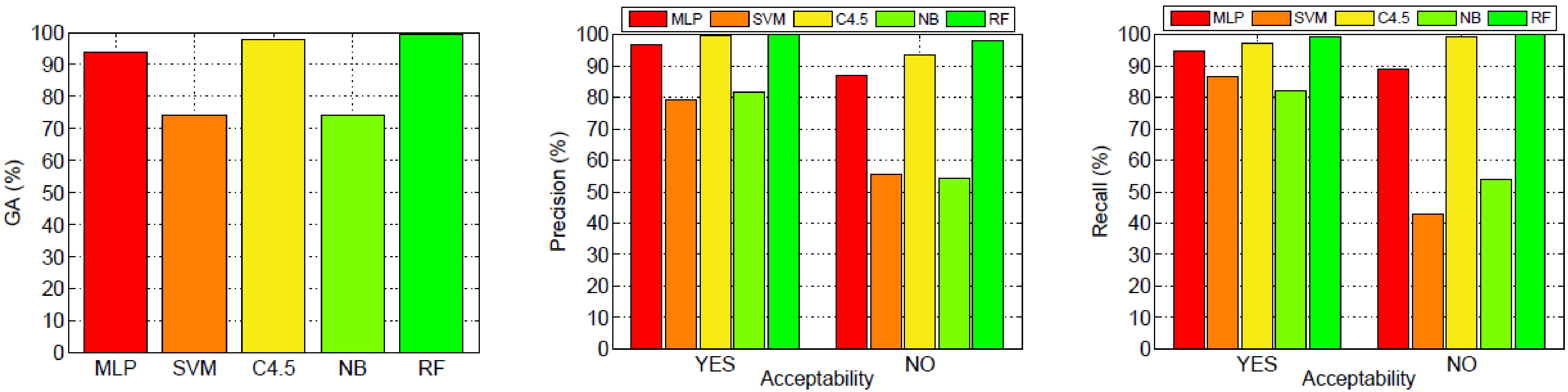
(b) Precision.



(c) Recall.

- **Accuracy, Precision and Recall for the three apps merged** (classes are balanced by bootstrapping to avoid biased evaluations)
- **Decision Tree (DT)-based models achieve the best results**
- **DTs are fast** and **provide clear information** on the inputs leading to a particular output
- **DTs are more robust than other models to noisy inputs** (e.g., much better than MLPs), as they perform embedded feature selection
- **A single DT** can correctly **predict more than 90%** of the session **QoE MOS** scores

Benchmarking Different ML Models: ACC



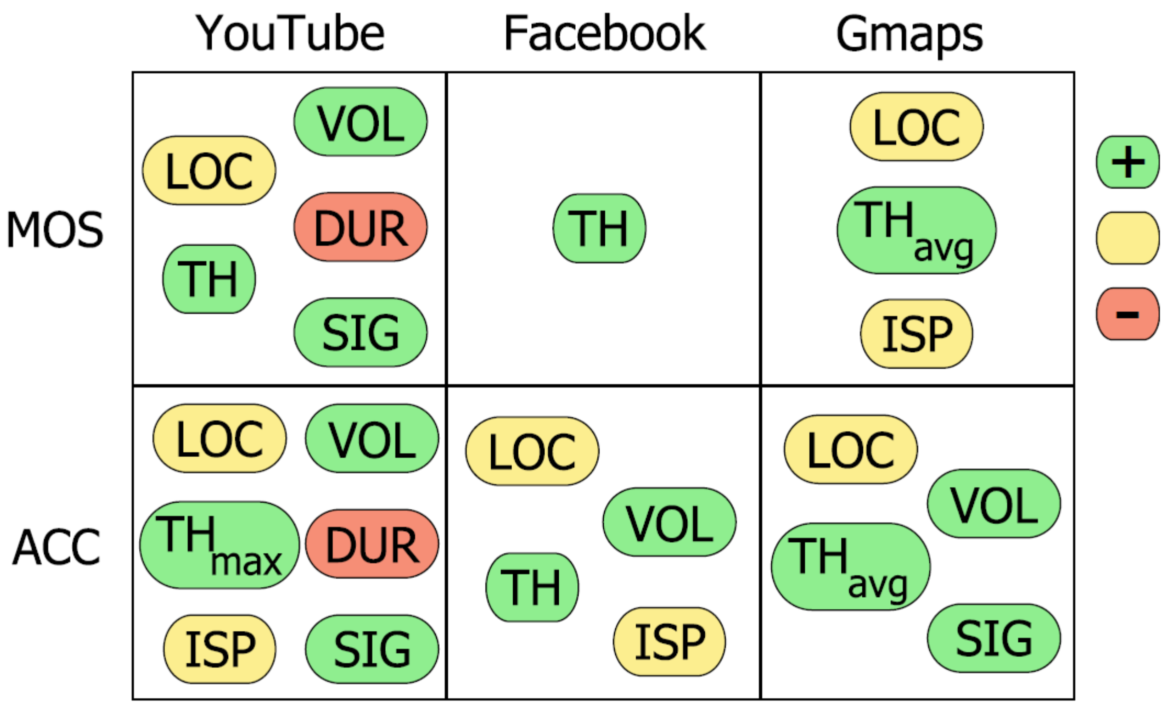
(a) Global Accuracy. (b) Precision. (c) Recall.

- Similar results are obtained for the prediction of service acceptability
- For this metric, **the model can correctly predict more than 98% of the ACC feedbacks**

AGENDA

- *Field Trial Description*
- *Monitoring Apps and Machine Learning Models*
- *QoE Assessment Results*
- ***Relevant Features for QoE in Smartphones***
- *Concluding Remarks*

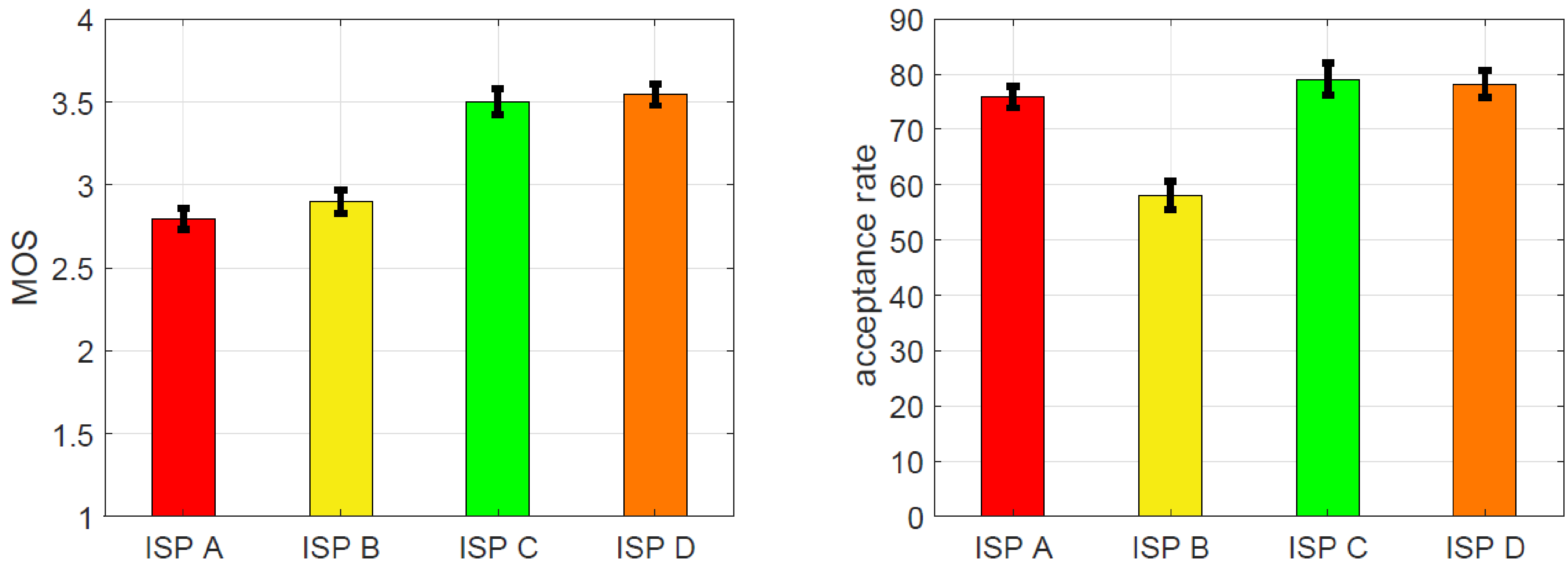
Contribution of Relevant Features to QoE



- We apply **feature selection** to understand the **contribution of the most relevant features**
- Correlation-based group testing, using Best First search

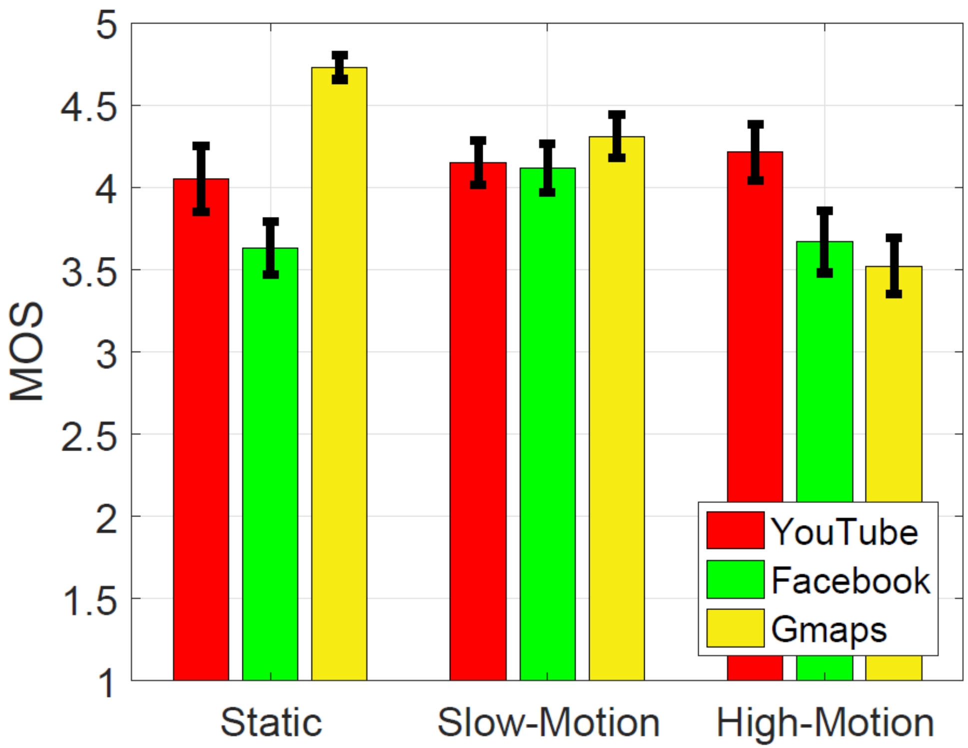
- Features flagged in green/red are positively/negatively correlated to the target metric (nominal features are marked in yellow)
- **TH (avg/max) are the most relevant features**
- Interestingly, **longer YouTube sessions experience a worse QoE** (stalling)
- The **ISP is also relevant**, specially for acceptability → **different ISP performance and user expectations**?

ISPs Benchmarking from a QoE Perspective



- 4 cellular ISPs operate in Vienna, we compare them in terms of MOS and ACC
- **There is a relevant different in terms of QoE among ISPs**
- Interestingly, **ISP A has similar acceptability rate than the best ISPs, with a poorer overall perceived QoE → expectations?**

Impact of Mobility on Smartphone QoE



- **Does mobility impact QoE?**
- We construct **mobility profiles** from user location as follows:
  - **static**: *home* and *office*
  - **slow-motion**: *street*
  - **high-motion**: *car, train, metro*

- There is **no apparent impact of the mobility profiles on both YouTube and Facebook** QoE (**apps are not highly interactive**)
- However, **Google Maps QoE is strongly correlated to mobility**, and the faster one moves, the worse QoE to be expect
- Gmaps task is tested in a very interactive manner, browsing through satellite-view maps to locate specific areas → **network QoS stability has a relevant impact on interactive sessions**

AGENDA

- *Field Trial Description*
- *Monitoring Apps and Machine Learning Models*
- *QoE Assessment Results*
- *Relevant Features for QoE in Smartphones*
- ***Concluding Remarks***

Conclusions

- **QoE is highly relevant to cellular ISPs**, and has even the potential to become a core guiding paradigm for 5G network management (see 5G PPP)

- **Our study is the first one addressing the problem of QoE monitoring, assessment and prediction in cellular networks**, relying **exclusively on in-smartphone QoS** passive traffic **measurements** and QoE crowdsourced feedback

- We have conceived a **two-phase system** which is capable of:
  - **generating a rich dataset of QoS/QoE measurements**, which can be used to train the operational model
  - **predicting QoE in smartphones for popular apps in a distributed fashion**, using **only in-smartphone passive traffic measurements** (**GENERALIZATION and APP-independence**)

- Using a DT model, evaluations show that **the proposed session features and model** can correctly **forecast the individual, per-user overall experience** and service **acceptability of popular apps** in 91%/98% of the monitored sessions

- The preliminary analysis on the **impact of the selected input features on QoE** should be extended to potentially **enhance future applications** for **network diagnosis** issues

Thanks for Your Attention!