UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

A Deep Learning-based approach to VM behavior identification in cloud systems

Matteo Stefanini, Riccardo Lancellotti,
Lorenzo Baraldi, Simone Calderara

University of Modena and Reggio Emilia
Department of engineering "Enzo Ferrari"

Cloud Computing Challenges

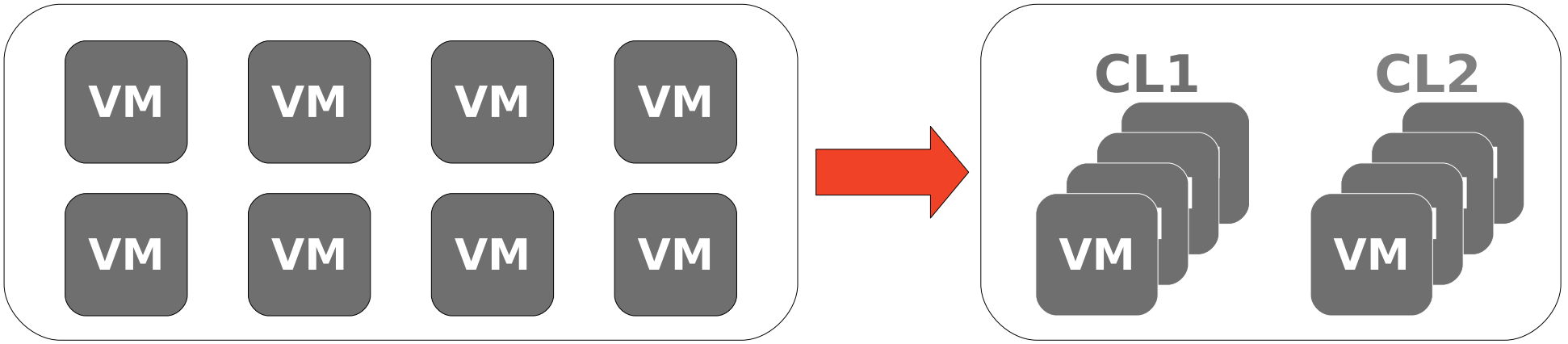
UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

- Critical operations in Cloud data centers
  - Monitoring (overloaded/underutilized VMs and Hosts)
  - Management (huge bin packing problem)
- Main challenge: Scalability
  - Volume of data for monitoring
  - Size (and dimensionality) of optimization problem
- Current solution
  - Oversimplification of the problem

Identification of VMs

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

- Alternative approach:
  - Exploit similarity in VMs: (classes, not instances)
  - Reduced problem size (less data, less VMs)



- Problem: how to classify VMs?
  - Fast and accurate classification

State of the art

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

- Trade-off accuracy/speed
  - Fast classification is not accurate
  - Accurate classification takes time
  - Cannot be applied to on-demand VMs in public Cloud
- Adaptive Gray Area TEchnique (AGATE)
  - Add a confidence value to classification
  - Fast and accurate classification of *some* VMs
  - Still unsatisfactory → Proposal of a different approach

Deep Learning model

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

- Input: time series of W samples of several VMs metrics
- Output: class belonging probabilities
- Multiple layers (number depending on the input size)
- Two models:
  - DeepConv: based on convolutive networks
    Focus on patterns between samples
  - DeepFFT: based on Fast Fourier Transformation
    Focus on spectral domain (novel Deep Learning approach)

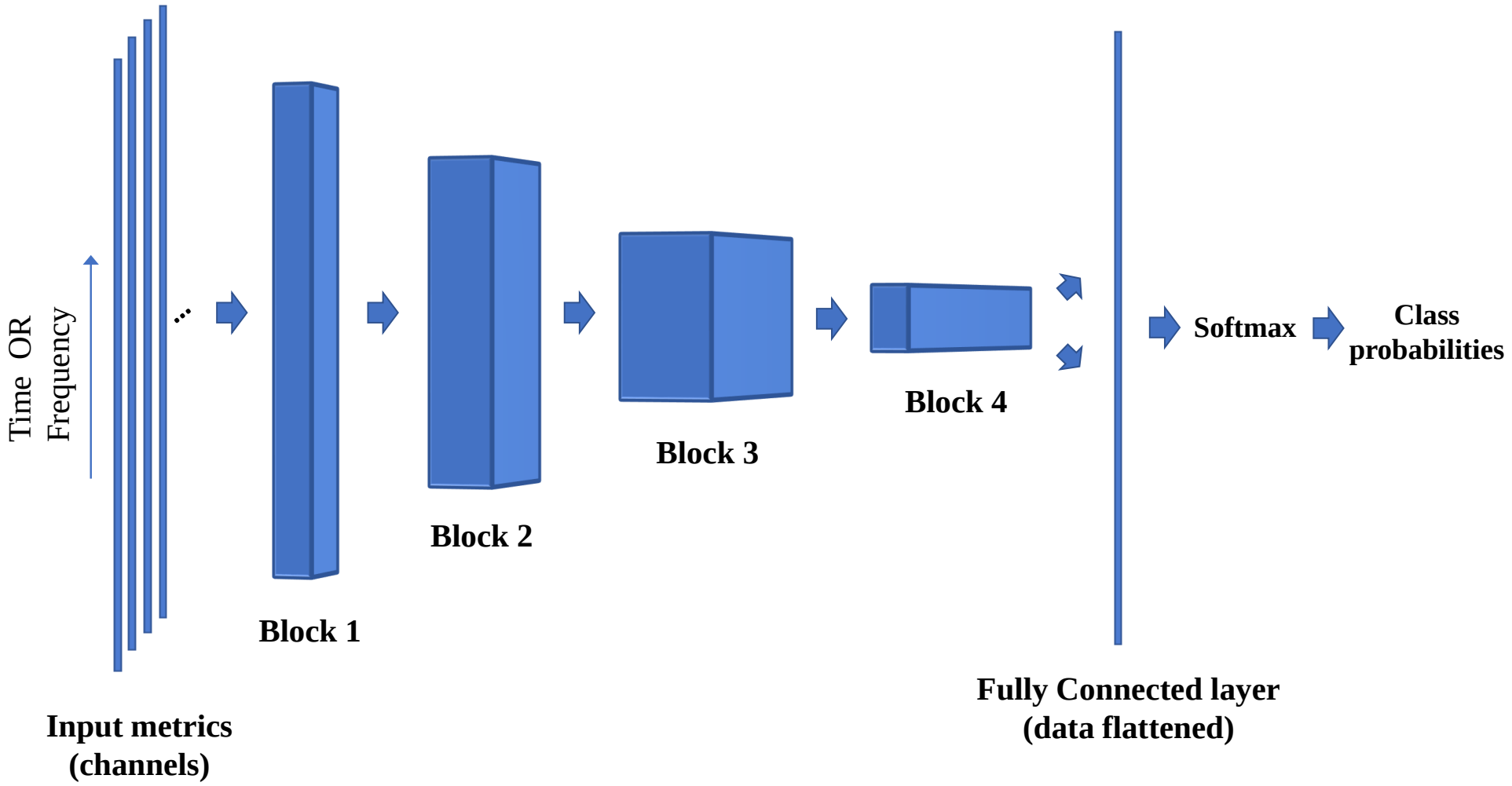
Deep Learning model

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

- General structure:
  - Input layer (pre-processing of samples)
  - Processing blocks (multiple layers)
  - Fully connected layer (and softmax classifier)
- DeepConv:
  - Standard model
- DeepFFT:
  - Performs FFT in input layer

Model representation

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA



Processing block

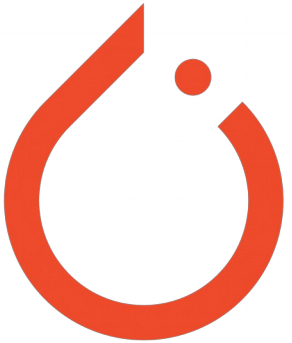
UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

- Each processing block contains
  - Activation function (ReLU)
  - Batch Normalization
  - 1-Dimensional convolution
- Each block:
  - Reduces by 2 the input size (stride=2)
  - Doubles the number of channels
- Number of blocks: $N_b = \max(\log_2(W) - 1, 2)$

Implementation details

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

- Implementation based on Pytorch
  - In-house implementation of FFT
- Source code available
  - Code in git repository
  - See paper for details
- Deployment on CINECA data center



Experimental setup

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

- Data from a real datacenter (e-health app)
- Two classes of VMs:
  - Web servers
  - DBMS
- Traces divided in chunks with different window
  - 1 sample every 5 min
  - 4 samples (20 mins) → 256 samples (21 hrs)

Experimental setup

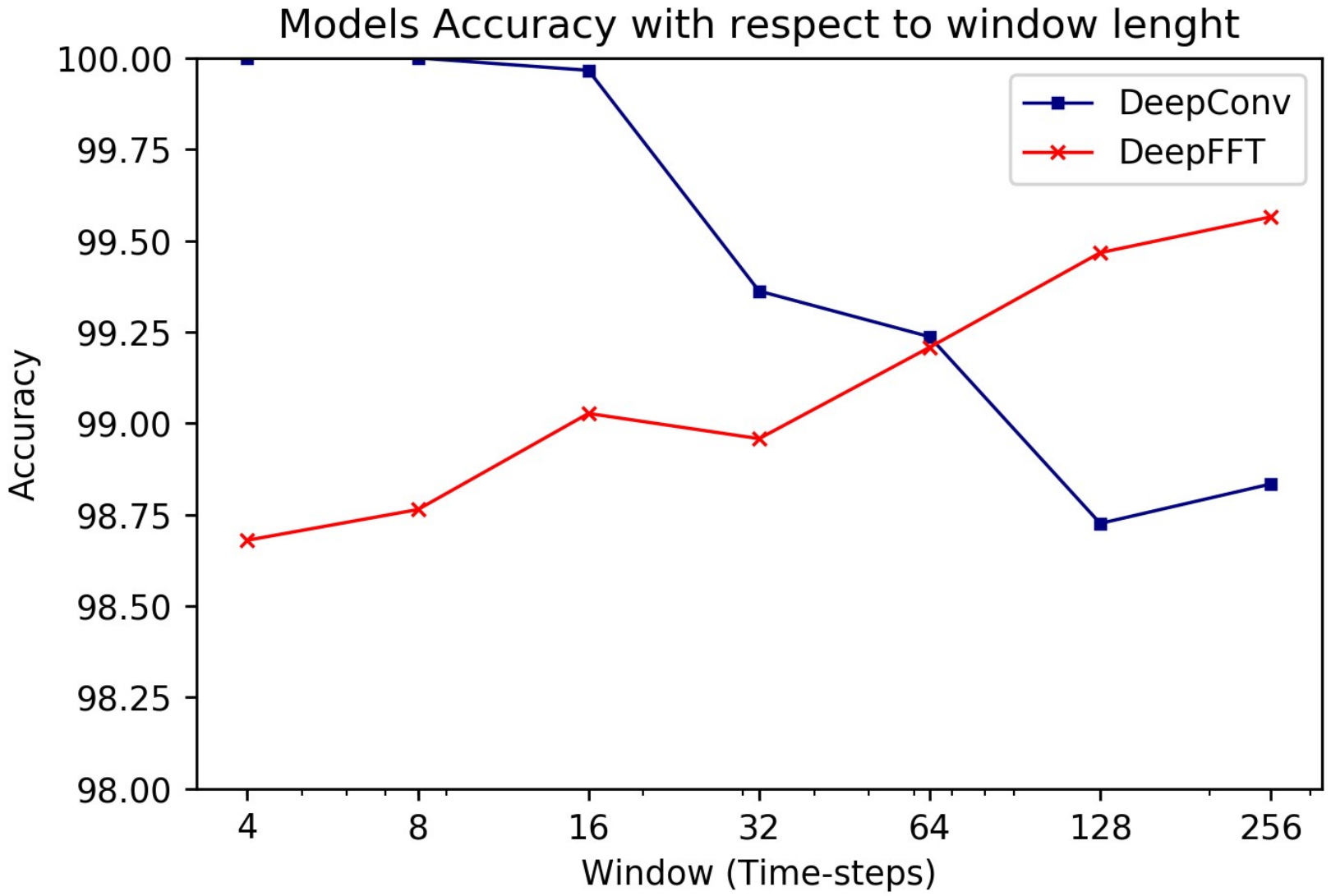
UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

- 16 metrics (virtualized HW / guest OS)

Metric	Description
SysCallRate	Rate of system calls [req/sec]
CPU	CPU utilization [%]
IdleCPU	Idle CPU fraction [%]
I/O buffer	Utilization of I/O buffer [%]
DiskAvl	Available disk space [%]
CacheMiss	Cache miss [%]
Memory	Physical memory utilization [%]
UserMem	User-space memory utilization [%]
PgOutRate	Rate of memory pages swap-out [pages/sec]
InPktRate	Rate of network incoming packets [pkts/sec]
OutPktRate	Rate of network outgoing packets [pkts/sec]
InByteRate	Rate of network incoming traffic [KB/sec]
OutByteRate	Rate of network outgoing traffic [KB/sec]
AliveProc	Number of processes in system
ActiveProc	Number of active processes in run queue
RunTime	Execution time

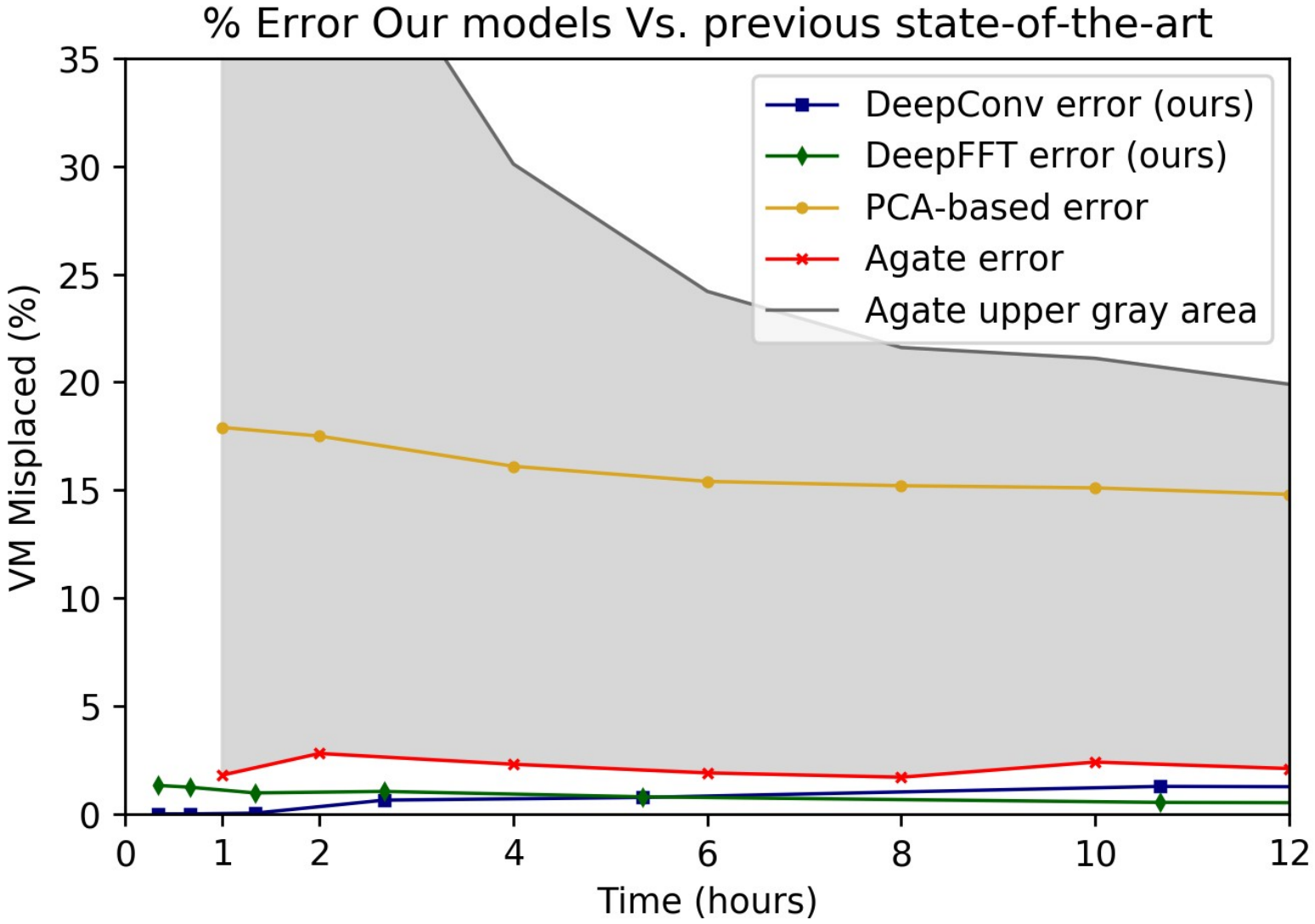
Deep Learning performance

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA



Comparison with AGATE

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA



Concluding remarks

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

- Challenge: scalability of monitoring/management in Cloud data centers → VMs identification
- Complex to achieve fast and accurate identification
- Proposal of a Deep Learning-based approach
- Outperforms state of the art (AGATE)
- Suitable also for on-demand VMs

Future research directions

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

- Thorough evaluation in cases with limited / low-quality data
- Identification of new classes:
  - Auto-encoders / triggers in NN
  - Integration with AGATE
- Generative Adversarial Network for workload generation

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

A Deep Learning-based approach to VM behavior identification in cloud systems

Matteo Stefanini, Riccardo Lancellotti,
Lorenzo Baraldi, Simone Calderara

University of Modena and Reggio Emilia
Department of engineering "Enzo Ferrari"