5G-Enabled NetApp for Predictive Maintenance in Critical Infrastructures

Sofia Giannakidou
K3Y Ltd.
sgiannakidou@k3y.bg

AUTHORS

Under the H2020 5G-INDUCE Project

Sofia Giannakidou, Panagiotis Radoglou-Grammatikis, Sotirios Koussouris, Minas Pertsetakis, Nikolaos Kanakaris, Alexios Lekidis, Konstantinos Kaltakis, Maria P. Koidou, Chrysi Metallidou, Konstantinos E.Psanis, Sotirios Goudos, Panagiotis Sarigiannidis.

This project has received funding from the European Unions Horizon 2020 research and innovation programme under grant agreement No.101016941 (5G-INDUCE).

Introduction

A predictive maintenance system for critical infrastructure

1

5G NetApp

Composed of 6 modules: On-Premises Apache Kafka Engine, On-Premises Data Collector, Edge Data Collector, Edge Analytics Engine, Cloud Data Collector and Cloud Visualisation Engine

2

Dockerized approach

Each of the components is provided in a container format. Moreover, the containers are orchestrated by an application orchestrator based on Kubernetes.

3

Zero-downtime approach

Timely detection of faulty machine states in energy production environments using historical data.

Related Work

R. Sahal, J. G. Breslin, and M. I. Ali, “Big data and stream processing platforms for industry 4.0 requirements mapping for a predictive maintenance use case,” Journal of Manufacturing Systems, vol. 54, pp. 138–151, 2020.

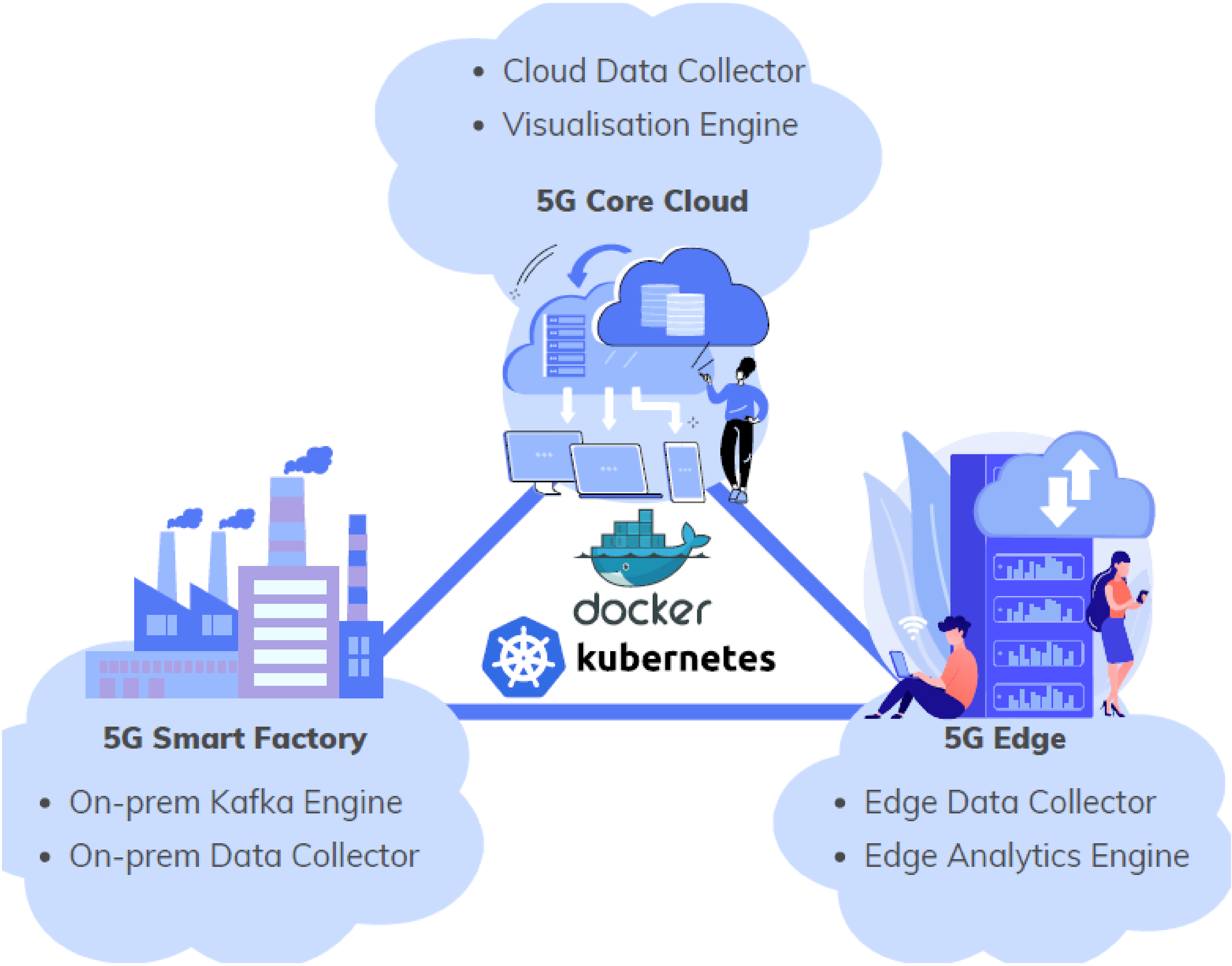
W. Shin, J. Han, and W. Rhee, “Ai-assistance for predictive maintenance of renewable energy systems,” Energy, vol. 221, p. 119775, 2021.

R. Fioravanti, K. Kumar, S. Nakata, B. Chalamala, and Y. Preger, “Predictive-maintenance practices: For operational safety of battery en ergy storage systems,” IEEE Power and Energy Magazine, vol. 18, no. 6, pp. 86–97, 2020.

Y. K. Teoh, S. S. Gill, and A. K. Parlikad, “Iot and fog computing based predictive maintenance model for effective asset management in industry 4.0 using machine learning,” IEEE Internet of Things Journal, pp. 1–1, 2021.

T. Ahmad, D. Zhang, C. Huang, H. Zhang, N. Dai, Y. Song, and H. Chen, “Artificial intelligence in sustainable energy industry: Status quo, challenges and opportunities,” Journal of Cleaner Production, vol. 289, p. 125834, 2021.

PROPOSED NETAPP ARCHITECTURE



ML-Based Detection

Three Main Steps

Preprocessing
Processes appropriately the input data so that it will be in accordance with the corresponding ML model. Usually, data preprocessing methods are applied, such as min max scaling, normalisation, standardisation , robust scaler and max abs scaler

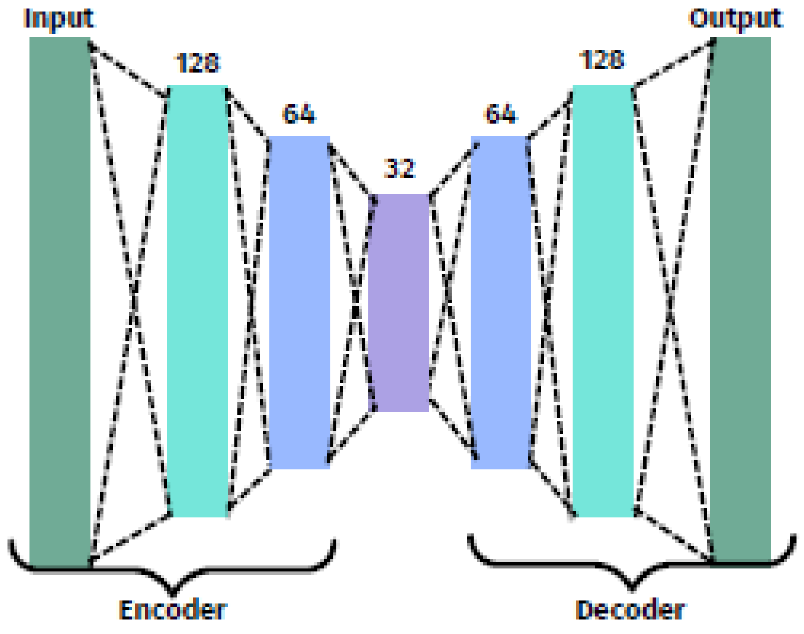
Training
Supervised detection methods, unsupervised/ oulier detection methods and semi supervised/novelty detection methods

Prediction
The ML model can be deployed in order to predict unknown data after the execution of the same pre-processing tasks of the first phase

Outlier Detection Algorithms

Autoencoder

Autoencoder is an unsupervised artificial neural network that learns how to efficiently compress and encode data then learns how to reconstruct the data back from the reduced encoded representation to a representation that is as close to the original input as possible



Principal Component Analysis (PCA)

Principal Component Analysis, or PCA, is a dimensionality-reduction method that is often used to reduce the dimensionality of large data sets, by transforming a large set of variables into a smaller one that still contains most of the information in the large set.

Reducing the number of variables of a data set naturally comes at the expense of accuracy, but the trick in dimensionality reduction is to trade a little accuracy for simplicity. Because smaller data sets are easier to explore and visualize and make analyzing data much easier and faster for machine learning algorithms without extraneous variables to process.

Angle-Based Outlier Detection (ABOD)

This technique is based on the idea of keeping an eye on the angle formed by a set of any three data points in the multi-variate feature space. The variance in the magnitude of the angular enclosure comes out to be different for outliers and normal points. Usually, the observed variance is higher for the inlier points than for outliers, hence such a measure helps us to cluster normal and outlier points differently. The angle-based outlier (ABOD) technique works pretty well in high-dimensional space, unlike other distance-based measures that suffer from the “Curse of dimensionality”. The distance between any two points in the high dimension space is almost similar. In such scenarios, angles can give better picture of closeness.

Outlier Detection Algorithms

One-Class SVM

One Class Support Vector Machine (SVM) aims to find a hyperplane that can separate the vast majority of data from the origin in the projected high dimensional space without making any assumptions about their distribution In particular, One Class SVM separates all the data points from the origin (in feature space) and maximises the distance from this hyperplane to the origin. This results in a binary function, which captures regions in the input space where the probability density of the data lives. The idea of One-Class SVM for anomaly detection is to find a function that is positive for regions with a high density of points, and negative for small densities.

Local Outlier Factor (LOF)

LOF relies on the concept of a local density, where locality is given by k nearest neighbours, whose distance is utilised to estimate the density. By comparing the local density of an object to the local densities of its neighbours, one can identify regions of similar density, and points that have a substantially lower density than their neighbours. These are considered to be outliers.

Isolation Forest

The Isolation Forest algorithm finds anomalies by deliberately "overfitting" models that memorize each data point Since outliers have more empty space around them, they take fewer steps to memorize. The algorithm is using full decision trees (every leaf is a single data point) and we measure the path length between the root and each leaf (data point) The final measure for each data point would be the average path length. Abnormal data points should be classified easily thus the average path should be relatively short.

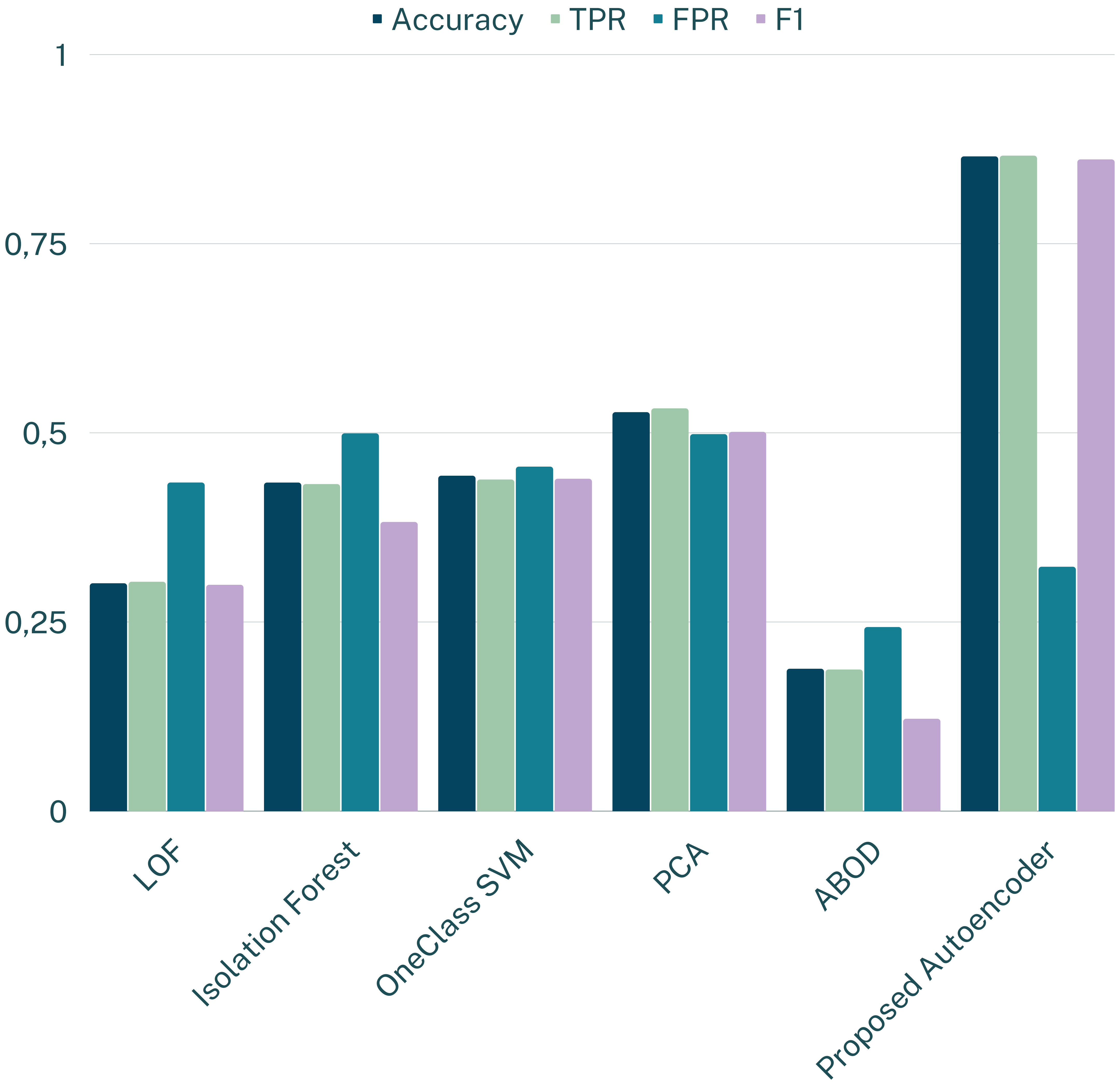
Evaluation

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$TPR = \frac{TP}{TP + FN}$$

$$FPR = \frac{FP}{FP + TN}$$

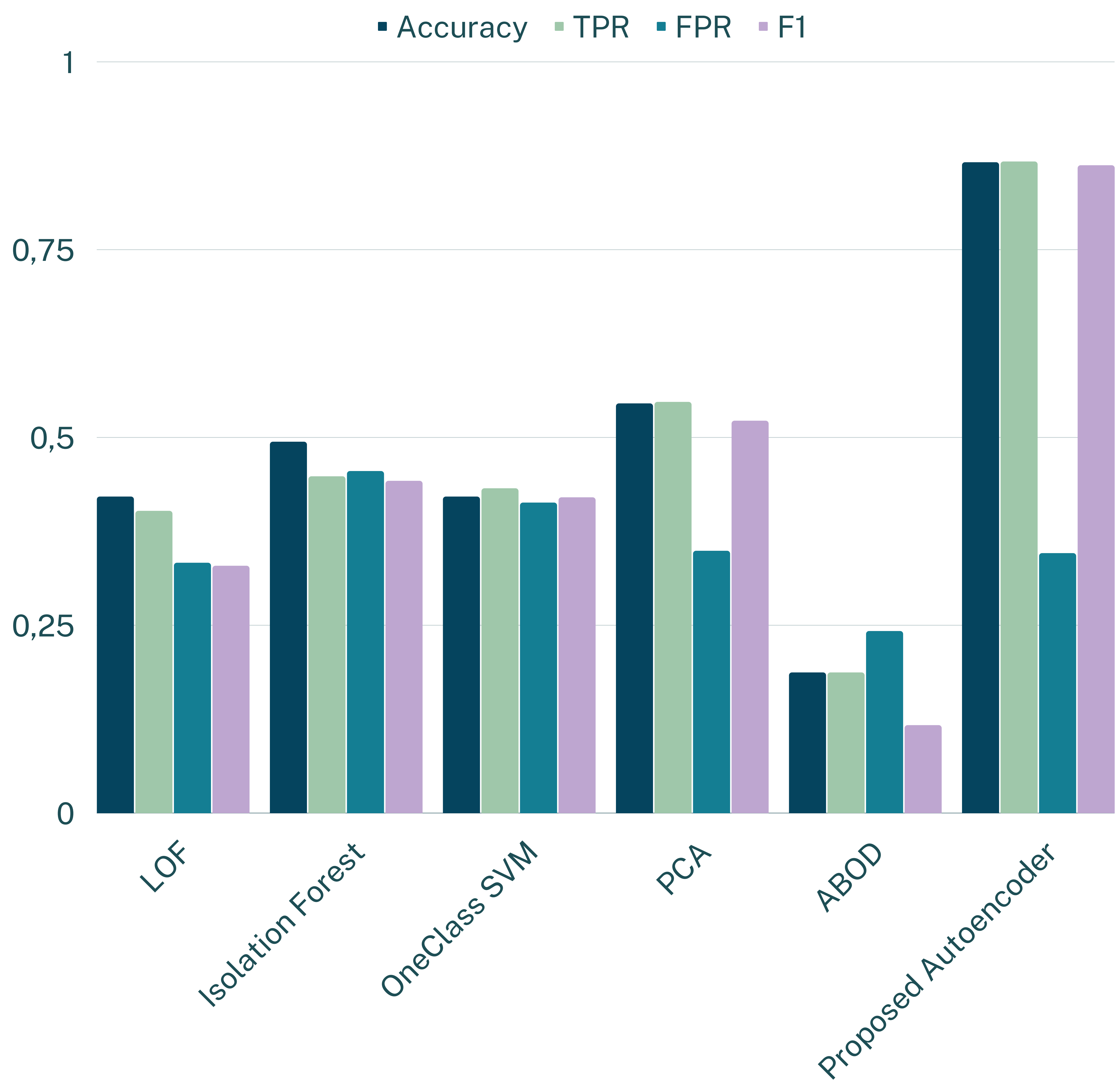
$$F1 = \frac{2 \times TP}{2 \times TP + FP + FN}$$



Results

Anomaly/Outlier Detection based on Energy Operational Data with Window Size 20 AI Model Accuracy

Results



Anomaly/Outlier Detection based on Energy Operational Data with Window Size 40

Thank You

sgiannakidou@k3y.bg

https://www.5g-induce.eu/

https://www.linkedin.com/in/5g-induce-project-7232b320b/