RADIO SIGNALS RECOGNITION WITH UNSUPERVISED DEEP LEARNING

Ljupcho Milosheski

Doctoral Dissertation
Jožef Stefan International Postgraduate School
Ljubljana, Slovenia

Supervisor: Prof. Mihael Mohorčič, Jožef Stefan Institute, Ljubljana, Slovenia
Co-Supervisor: Dr. Carolina Fortuna, Jožef Stefan Institute, Ljubljana, Slovenia

Evaluation Board:
Assoc. Prof. Dr. Andrej Hrovat, Chair, IPS and Jožef Stefan Institute, Ljubljana, Slovenia
Prof. Dunja Mladenić, Member, IPS and Jožef Stefan Institute, Ljubljana, Slovenia
Prof. Adnan Shahid, Member, Ghent University, Belgium

MEDNARODNA PODIPLOMSKA ŠOLA JOŽEFA STEFANA
JOŽEF STEFAN INTERNATIONAL POSTGRADUATE SCHOOL



Ljupcho Milosheski

RADIO SIGNALS RECOGNITION WITH UNSUPERVISED DEEP LEARNING

Doctoral Dissertation

RAZPOZNAVANJE RADIJSKIH SIGNALOV Z UPORABO NENADZOROVANEGA GLOBOKEGA UČENJA

Doktorska disertacija

Supervisor: Prof. Mihael Mohorčič

Co-Supervisor: Dr. Carolina Fortuna

Ljubljana, Slovenia, June 2026

To my family

Acknowledgments

I would like to express my sincere gratitude to my supervisors, Prof. Mihael Mohorčič and Dr. Carolina Fortuna, for their guidance, support, and encouragement throughout my doctoral studies. Their insightful feedback, critical perspective, and openness to discussion were invaluable to the development of this thesis. I particularly appreciate their high standards and clear direction, which not only elevated the quality of my research but also helped me grow as an independent researcher.

I am also grateful to my colleagues and collaborators for their practical assistance, constructive comments, and the many discussions that enriched this work. I would additionally like to acknowledge the technical and administrative staff of the Department of Communication Systems (E6) for their assistance with equipment, procedures, and organizational matters, all of which contributed to the smooth completion of this work.

This research was financially supported by the Slovenian Research Agency (ARIS) under grant P2-0016, and by the Horizon Europe projects NANCY (grant agreement No 101096456), and EnerTEF (grant agreement No 101172887). I gratefully acknowledge their support.

This research was financially supported by the Slovenian Research Agency (ARIS) under grant P2-0016, and by the Horizon Europe projects NANCY (grant agreement No 101096456) and EnerTEF (grant agreement No 101172887). I gratefully acknowledge their support.

Finally, and most importantly, I would like to thank my family for their unconditional support, patience, and understanding throughout this journey.

Abstract

Wireless communication networks have evolved significantly over the past decade, driven by rapidly growing traffic demands and device diversity. This evolution has intensified pressure on the radio-frequency spectrum, a limited medium where usage remains uneven and time-varying, making spectrum optimization a highly relevant research direction. Reliable optimization requires continuous awareness of the dynamic radio environment, encompassing transmission detection, technology recognition, and usage pattern tracking, which can enable closed-loop control of network parameters.

Deep learning-based models have shown remarkable signal recognition abilities over the last decade, significantly outperforming classical approaches in classification tasks. However, their deployment is constrained by the cost and scarcity of labeled data in real-world conditions. The emergence of open radio access network (O-RAN) architectures and open testbeds, exposing large-scale unlabeled spectrum data through standardized interfaces, creates a timely opportunity for unsupervised deep learning approaches that can learn from unlabeled data automatically, reducing or eliminating the necessity of expert annotation.

This dissertation investigates *unsupervised deep learning*, with particular emphasis on deep clustering, as a scalable and practically deployable approach for radio access technology (RAT) recognition from unlabeled spectrum measurements. The main contributions are: (i) a comprehensive review of unsupervised deep learning for wireless signal recognition, linking recognition challenges to suitable unsupervised learning families and identifying open research directions; (ii) adaptation and benchmarking of deep clustering architecture for RAT recognition from spectrograms, based on established deep learning modules; (iii) a label-free spectrum sensing workflow that requires no prior knowledge of active RATs or operating conditions and uses post-hoc cluster-to-class mapping; and (iv) a lightweight CNN integrated into the deep clustering framework and tailored to the fast Fourier transform (FFT)-amplitude representation of RAT transmissions.

The results demonstrate that deep clustering consistently learns more discriminative representations than baseline approaches. Systematic adaptation of image processing modules for spectrogram processing yields reductions in model complexity of up to one order of magnitude while preserving or improving clustering performance by up to 2.5 times. A dedicated CNN module for FFT amplitudes further reduces the parameter count by two orders of magnitude compared to the module for processing spectrograms, and by 22% compared to the state of the art approach. At the same time it reduces computational cost by 50% and improves feature separability and clustering quality by 35 percentage points relative to an alternative architecture using the same modules. Validation across multiple real-world datasets collected in diverse radio environments confirms the consistency and practical viability of the proposed approaches, contributing to data-driven spectrum awareness as a foundational capability for future AI-native wireless communication networks.

Povzetek

Hitro naraščanje podatkovnega prometa in raznolikost naprav sta v zadnjem desetletju spodbudila izrazit razvoj brezžičnih komunikacijskih omrežij, kar povečuje pritisk na radiofrekvenčni spekter. Ta predstavlja omejen naravni vir, katerega izkoriščenost ostaja neenakomerna in časovno spremenljiva, zato je optimizacija rabe spektra zelo aktualna raziskovalna smer. Za zanesljivo optimizacijo je potrebno neprekinjeno zavedanje o dinamičnem radijskem okolju, kar zajema zaznavanje oddanih signalov, prepoznavanje radijskih tehnologij in sledenje vzorcem zasedenosti spektra, s čimer se lahko preko povratne informacije ustrezno prilagajajo parametri omrežja.

Modeli globokega učenja so v zadnjem desetletju pokazali izjemne zmogljivosti prepoznavanja signalov in so pri nalogah razvrščanja znatno presegli klasične pristope. Njihovo uporabo pa omejujeta zahtevnost označevanja in posledično pomanjkanje označenih podatkov v realnih pogojih delovanja. Pojav arhitektur odprtega radijskega dostopovnega omrežja (O-RAN) in odprtih eksperimentalnih okolij, ki prek standardiziranih vmesnikov omogočajo dostop do obsežnih neoznačenih spektralnih podatkov, sta ustvarila priložnost za uporabo pristopov nenadzorovanega globokega učenja, ki se samodejno učijo iz neoznačenih podatkov ter s tem zmanjšujejo ali odpravljajo potrebo po ekspertnem označevanju.

Disertacija preučuje *nenadzorovano globoko učenje* s poudarkom na globokem razvrščanju v gruče (angl. deep clustering) kot razširljivem in praktično uporabnem pristopu za prepoznavanje tehnologij radijskega dostopa (RAT) iz neoznačenih spektralnih meritev. Glavni prispevki so: (i) celovit pregled nenadzorovanega globokega učenja za prepoznavanje brezžičnih signalov, ki povezuje izzive prepoznavanja z ustreznimi skupinami metod nenadzorovanega učenja in opredeljuje odprte raziskovalne smeri; (ii) prilagoditev in primerjalno vrednotenje arhitekture za globoko razvrščanje v gruče, zasnovane na uveljavljenih modulih globokega učenja, za prepoznavanje tehnologij RAT iz spektrogramov; (iii) postopek zaznavanja spektra brez oznak ter brez predhodnega poznavanja aktivnih RAT ali pogojev delovanja, ki temelji na naknadni preslikavi gruč v razrede; ter (iv) vitka konvolucijska nevronska mreža (CNN), vključena v ogrodje globokega razvrščanja in prilagojena amplitudni predstavitvi hitre Fourierjeve transformacije (FFT) signalov RAT.

Rezultati kažejo, da globoko razvrščanje dosledno zagotavlja bolj diskriminativne predstavitve kot referenčni pristopi. Sistematična prilagoditev modulov, razvitih za obdelavo slik, za procesiranje spektrogramov, zmanjša kompleksnost modela do enega velikostnega reda, hkrati pa ohrani ali izboljša učinkovitost razvrščanja v gruče do 2,5-krat. Namensko razviti modul CNN za amplitude FFT signalov dodatno zmanjša število parametrov za dva velikostna reda v primerjavi z modulom za obdelavo spektrogramov in za 22% v primerjavi z najsodobnejšim pristopom, računsko zahtevnost pa zmanjša za 50%. Hkrati izboljša ločljivost značilk in kakovost razvrščanja za 35 odstotnih točk v primerjavi z alternativno arhitekturo, ki uporablja enake module. Validacija na več zbirkah realnih podatkov iz raznolikih radijskih okolij potrjuje doslednost in praktično izvedljivost predlaganih pristopov. S tem disertacija prispeva k podatkovno vodenemu zavedanju o rabi spektra kot osnovi prihodnjih brezžičnih komunikacijskih omrežij z vgrajeno umetno inteligenco.

Contents

List of Figures **xvii**

List of Tables **xix**

Abbreviations **xxi**

1 Introduction **1**
1.1 Motivation . . . 2
1.2 Purpose of the Dissertation . . . 4
1.3 Research Hypotheses . . . 4
1.4 Methodology . . . 4
1.5 Contributions . . . 6
1.6 Organization of the Thesis . . . 7

2 Unsupervised Deep Learning for Recognition of Radio Signals **9**
2.1 Introduction . . . 10
2.2 Related work . . . 12
2.3 Unsupervised deep dearning techniques in wireless communications . . . 15
2.4 Neural architectures for the representation learning modules . . . 17
2.5 Automatic modulation classification . . . 20
2.5.1 Overall AMC comparative overview . . . 21
2.5.2 Overview per approach to AMC . . . 22
2.5.3 Datasets . . . 24
2.5.4 Findings . . . 24
2.6 Signal Sensing . . . 24
2.6.1 Types of signal sensing . . . 25
2.6.2 Overview per approach to SS . . . 26
2.6.3 Datasets . . . 27
2.6.4 Findings . . . 28
2.7 Specific emitter identification . . . 29
2.7.1 Overall SEI comparative overview . . . 29
2.7.2 Overview per approach to SEI . . . 29
2.7.3 Datasets . . . 31
2.7.4 Findings . . . 31
2.8 Anomaly detection . . . 32
2.8.1 Overall AD comparative overview . . . 32
2.8.2 Overview per approach to AD . . . 32
2.8.3 Datasets . . . 33
2.8.4 Findings . . . 33
2.9 Challenges and future directions . . . 34
2.10 Conclusions . . . 35

3 Deep Clustering of Wireless Spectrum Activity Based on Spectrogram Processing **41**
3.1 Introduction . . . 43
3.2 Related work . . . 44
3.3 Architectures of the SSL and baseline systems . . . 45
3.3.1 SSL architectures . . . 46
3.3.2 Baseline systems . . . 47
3.4 Methodolgy . . . 47
3.4.1 Raw training data preparation . . . 47
3.4.2 Feature development and evaluation methodology . . . 48
3.4.3 Cluster development and evaluation methodology . . . 48
3.4.4 Evaluation with labeled data . . . 48
3.5 Experimental results . . . 49
3.5.1 Effect of the dimensionality reduction . . . 49
3.5.2 Clustering tendency of the features . . . 49
3.5.3 Evaluation of clusters quality . . . 50
3.5.4 Analysis of clusters . . . 50
3.5.5 Manual evaluation . . . 51
3.5.6 Evaluation with labeled data . . . 54
3.6 Conclusions . . . 54

4 Spectrum Sensing with Deep Clustering for Label-Free Radio Access Technology Recognition Based on FFT Amplitudes **57**
4.1 Introduction . . . 59
4.2 Related work . . . 61
4.3 Reference scenario and proposed spectrum sensing workflow . . . 62
4.4 Problem statement . . . 63
4.5 Proposed SSL architecture . . . 63
4.5.1 Description of architectural modules . . . 64
4.5.2 Workflow of the SSDC architecture . . . 64
4.5.3 Reference autoencoder-based architectures . . . 65
4.6 Evaluation methodology . . . 66
4.6.1 Training data . . . 66
4.6.2 Training vconfigurations . . . 67
4.6.3 Cluster quality evaluation metrics . . . 67
4.6.4 Transmission detection evaluation metric . . . 69
4.6.5 Computational performance metrics . . . 69
4.7 Results . . . 69
4.7.1 Evaluation results with TCD-L . . . 70
4.7.2 Evaluation results with ITS-L . . . 72
4.7.3 Evaluation results with UNB-L . . . 73
4.7.4 Potential deployment challenges . . . 74
4.8 Conclusion . . . 75

5 Conclusions and Future Work Directions **77**
5.1 Conclusions . . . 77
5.2 Future Work . . . 78

References **81**

Bibliography **85**

Biography **87**

List of Figures

Figure 2.1: Overview of the survey structure. . . . 11
Figure 2.2: Relative performance of classical unsupervised algorithms compared to deep learning approaches. . . . 14
Figure 2.3: General structure of deep clustering methods. . . . 15
Figure 2.4: General structure of contrastive learning methods. . . . 15
Figure 2.5: General structure of reconstruction-based learning methods. . . . 16
Figure 2.6: General structure of conditional generative adversarial networks. . . . 16
Figure 2.7: General structure of predictive methods. . . . 16
Figure 2.8: Unsupervised deep learning approaches (Venn diagram). . . . 17
Figure 2.9: Multilayer perceptron illustration. . . . 17
Figure 2.10: Single layer of 1D convolutional neural network illustration. . . . 18
Figure 2.11: Types of recurrent neural networks. . . . 19
Figure 2.12: Self-attention and multihead attention modules. . . . 20
Figure 2.13: Deep learning modules used across surveyed works (Venn diagram). . . . 20
Figure 2.14: Reported performance vs model size / labeled samples (AMC). . . . 21
Figure 2.15: Reported accuracy of surveyed models in low SNR conditions. . . . 22
Figure 2.16: Size comparison of the models in signal sensing. . . . 26
Figure 2.17: Estimated size comparison of the models in SEI. . . . 30
Figure 2.18: Summary of existing challenges and future research directions. . . . 34
Figure 2.19: Web plots of key common requirements for comparative analysis across four challenges. . . . 35

Figure 3.1: Architecture of the self-supervised CNN system. . . . 46
Figure 3.2: Architectures of the baseline systems. . . . 47
Figure 3.3: Sample of 8 spectrogram segments from the data. . . . 48
Figure 3.4: Cumulative sum of EVR of the PCA-transformed features. . . . 49
Figure 3.5: VAT plots of the different feature vectors. . . . 50
Figure 3.6: Hopkins scores of the extracted features. . . . 50
Figure 3.7: Evaluation curves of metrics for quality of clustering. . . . 50
Figure 3.8: Average spectrograms and frequency histograms of appearance for each cluster. . . . 52

Figure 4.1: SSL-enabled spectrum sensing workflow. . . . 62
Figure 4.2: Proposed SSDC architecture. . . . 63
Figure 4.3: SotA autoencoder-based architecture. . . . 65
Figure 4.4: Samples for each signal type in the TCD-L dataset. . . . 66
Figure 4.5: Samples for each signal type in the ITS-L dataset. . . . 67
Figure 4.6: Evaluation of the model learnt with the SSDC architecture and TCD-L dataset. . . . 70
Figure 4.7: Evaluation of the model learnt with the AEML architecture and TCD-L dataset. . . . 70

Figure 4.8: Evaluation of the model learnt with the DCEC architecture and TCD-L dataset. 71
Figure 4.9: Feature-space and average clusters of the different models. 73

List of Tables

Table 1.1: Mapping of the dissertation contributions to the core challenges in unsupervised RAT recognition, indicating the chapter in which each is addressed. 6

Table 2.1: Summary of related works concerning unsupervised learning for wireless communication. 13
Table 2.2: Summary of unsupervised deep learning approaches and achievements in AMC. 21
Table 2.3: Summary of RadioML datasets and their characteristics. 23
Table 2.4: Summary of works concerning unsupervised deep learning approaches and achievements in signal sensing. 25
Table 2.5: Summary of real-world datasets for wireless signal recognition. 28
Table 2.6: Summary of unsupervised deep learning approaches and achievements for SEI. 30
Table 2.7: Summary of datasets for SEI challenges. 31
Table 2.8: Summary of works concerning unsupervised deep learning approaches and achievements for RF anomaly detection. 33
Table 2.9: Summary of open datasets for anomaly detection. 33

Table 3.1: Related works with main characteristics. 45
Table 3.2: Statistics of the clusters. 51
Table 3.3: Transmission detection performance and complexity comparison. 54

Abbreviations

5G NR	5G New Radio
AD	Anomaly Detection
AE	Autoencoder
AEML	Autoencoder with Modified Loss
AMC	Automatic Modulation Classification
ARI	Adjusted Rand Index
CCS	Clustering completeness score
CH	Calinski-Harabasz score
CHS	Clustering homogeneity score
CNN	Convolutional Neural Network
CPU	central processing unit
C-V2X	Cellular Vehicle-to-Everything
DB	Davies-Bouldin score
DC	Deep Clustering
DCEC	Deep Convolutional Embedded Clustering
DL	Deep Learning
DMD	Dynamic Mode Decomposition
DST	Digital Spectrum Twin
DVB-T	Digital Video Broadcasting – Terrestrial
FC	Fully Connected
FFT	Fast Fourier Transform
FLOPS	floating-point operations per second
GAN	Generative Adversarial Network
GASF	Gramian Angular Summation Field
GFLOPS	one billion floating-point operations per second
GMM	Gaussian Mixture Model
IPS	International Postgraduate School
I/Q	In-phase and Quadrature
ISAC	Integrated Sensing and Communication
ITS	Intelligent Transportation Systems
ITS-L	Intelligent Transportation Systems – Labeled
JSI	Jožef Stefan Institute
KL	Kullback-Leibler divergence
LBT	Listen-Before-Talk
LSTM	Long Short-Term Memory
LTE	Long-Term Evolution
ML	Machine Learning
MLP	Multilayer Perceptron
MSE	Mean Squared Error
Near-RT	Near-real-time

NLP ... Natural Language Processing
NMI ... Normalized Mutual Information
NN ... Neural Network
OFDM.. Orthogonal Frequency-Division Multiplexing
O-RAN. Open Radio Access Network
PCA ... Principal Component Analysis
ppt ... percentage points
PSD ... Power Spectral Density
RAN ... Radio Access Network
RAT ... Radio Access Technology
REM ... Radio Environment Map
RF ... Radio Frequency
RGB ... Red-Green-Blue
RIC ... Radio Intelligent Controller
RSSI ... Received Signal Strength Indicator
RU ... Radio Unit
SEI ... Specific Emitter Identification
SL ... Silhouette score
SNR ... Signal-to-Noise Ratio
SS ... Signal Sensing
SSDC... Self-supervised Deep Clustering
SSL ... Self-Supervised Learning
TCD-L.. Technology Classification Dataset – Labeled
t-SNE... t-distributed Stochastic Neighbor Embedding
UDL ... Unsupervised Deep Learning
UMAP.. Uniform Manifold Approximation and Projection
UNB-U.. Ultra Narrow Band – Unlabeled
ViT ... Vision Transformer

Chapter 1

Introduction

In wireless communication networks, the rapidly growing number [1] and diversity of connected devices, operating across multiple Radio Access Technologies (RATs), together with increasing traffic demands, continuously increase the minimum technical requirements for commercial wireless networks, such as bandwidth and the supported number of simultaneously connected devices in various operational conditions. This naturally shifts the usage of the spectrum towards the higher, less-used radio frequency (RF) bands, which can offer larger bandwidths and thus support higher data rates, thus leading to new, higher frequency spectrum allocations [2]. As a result, two significant limitations emerge. First, the RF spectrum as a transmission medium is a limited resource, especially as only a subset of frequencies is suitable and available for commercial wireless communications. And second, the higher-frequency signals are more sensitive to operating environment conditions, such as obstacles, mobility, and interference from a variety of wireless technologies that operate in the same or adjacent RF bands. Thus, next-generation networks will require more intelligent and adaptive parameter control based on continuous environmental monitoring and feedback, which can lead to better network resources and RF spectrum utilization. Recent urban-area, continuous measurement campaigns report [3] transmissions being conducted only 1–24% of the time in the 3.45–3.55 GHz and 3.55–3.65 GHz bands, indicating substantial headroom for more efficient spectrum sharing even in heavily used bands.

Overcoming these limitations requires monitoring and providing feedback from the radio operating environment, which, in the context of wireless communications, means detection, classification, and interpretation of radio signals present in the environment. Key signal recognition challenges include RAT classification [4], modulation recognition [5], anomaly detection [6], interference identification [7], and specific emitter identification (SEI) [8]. Solutions to these challenges provide awareness of the network, which is important for virtualization, monitoring, and optimization of the use of wireless network resources, including RF spectrum.

Over time, a variety of signal recognition techniques have been explored in the context of RF spectrum sensing, including traditional methods such as energy detection [9], [10], and detection of cyclostationary signal characteristics [11], [12]. More recently, deep learning approaches [13]–[15], particularly supervised convolutional neural networks (CNNs), have shown promising results. However, their effectiveness is often constrained by the reliance on large, labeled datasets. Such labeled datasets are expensive and time-consuming to produce [10] and require domain experts for annotation, which is further complicated by the low overlap between expert annotations [10]. Furthermore, the inherently dynamic and stochastic nature of wireless environments makes it difficult to generalize models trained on static or limited data. These challenges highlight the need for adaptive, data-driven learning approaches that can operate effectively with minimal supervision.

Unsupervised Deep Learning (UDL) methods, such as deep clustering [16]–[18], have emerged as a powerful alternative to supervised approaches, especially in cases where labeled data is scarce. They include Self-Supervised Learning (SSL) approaches, which are meant to drastically reduce the amount of labelled data [19] that is required for the pretraining phase of high-accuracy ML models that are subsequently fine-tuned with supervised approaches. UDL methods learn meaningful, most often compressed, representations from raw data by solving surrogate (pretext) tasks that do not rely on external labels. For the deep clustering approach, the pretext task is training a classifier based on pseudo-labels provided by the clustering method. In the context of wireless communications, UDL can enable systems to autonomously discover structures in unlabelled radio signals, group similar transmission patterns, and assign semantic meaning through downstream tasks like classification or clustering. By leveraging large volumes of unlabelled spectrum data, UDL provides a scalable and label-efficient path to enabling intelligent spectrum sensing capabilities.

Considering the emergence of open datasets and research testbeds, such approaches are expected to become mainstream in the training of ML models for signal recognition in wireless communications. This is aligned with the trends in other closely related domains, such as computer vision and natural language processing (NLP), where UDL techniques are critical in training large world and language models. This makes the investigation, development, and validation of SSL approaches an increasingly important and timely topic in wireless communications research.

1.1 Motivation

Given the dynamic and complex nature of contemporary wireless environments, characterized by a wide variety of devices and multiple RATs, spectrum sensing is envisioned to become a building component of future 6G networks [20]. As such, it will support the progress of Integrated Sensing and Communication (ISAC)-related systems and technologies [21], [22], facilitating their development and integration into the next generation of wireless communication systems. Consequently, spectrum sensing is anticipated to be a key component in the evolution of emerging wireless communication frameworks such as Open Radio Access Network (O-RAN)[23] and Digital Spectrum Twins (DSTs)[24].

In the emerging O-RAN in which a large proportion of the radio functions are realized in software [25], [26], RAT monitoring could play an important role in optimizing radio and network parameters and spectrum utilization, largely through the employment of AI algorithms encapsulated as xApps and rApps applications for near-real-time and non-real-time control, respectively [23]. By continuously assessing the usage and availability of different RATs, O-RAN can utilize intelligent applications to dynamically fine-tune network settings, allocate spectrum resources or manage beam forming in real time. Thus, it can enhance the overall efficiency and performance of the network and accommodate growing traffic demands.

DSTs are digital representations of the real-world characteristics of the RF spectrum in a region based on historical data and measurement updates [27]. They are envisioned as tools to inform, forecast, and enhance spectrum utilization and interference management in wireless networks in a given geographical area. Spectrum sensing is a building component of DST [24]. It provides the measurements input necessary to refine the DST, which is initially developed based on empirical, theoretical, and/or ray-tracing models. Incorporating RAT monitoring into DST could further enhance its capabilities by adding detailed data on the various transmission technologies that are present in the area, such as RAT-specific activity patterns for a given time interval within the operational landscape. This integration could

significantly improve the DST's awareness, accuracy and functionality, facilitating more effective spectrum management and optimization strategies.

The current state-of-the-art research for RAT classification predominantly revolves around supervised CNN-based approaches [28]–[30], which have shown promising results, particularly in diverse Signal-to-Noise (SNR) settings and various environmental conditions [13]. However, these methods have a number of shortcomings. First, they heavily depend on extensive labeled datasets acquired in controlled settings where the transmission parameters need to be recorded in addition to the received signals, requiring a more complex set-up for collection compared to unlabelled data. Second, because they are trained on labeled data from specific conditions, adapting them to new environments, whether characterized by (i) additional RATs not present in the original training data or (ii) different physical obstructions and noise variations, may require collecting and labeling new data for each setting. Third, the majority of the models have been developed and evaluated on a single dataset, leaving open questions regarding their consistency across diverse environments.

Self-supervised deep learning approaches, extensively explored in other application domains [31] such as image and audio processing, are able to address the first two challenges mentioned above, because they do not require labels for training. Collected spectrum data can be used directly to train such models, without the need to record transmission parameters or annotate individual samples. By analogy, this approach should be able to adapt the model in regular retraining periods to new types of RATs that may appear in the area. However, to be able to realize the classification functionality similar to their supervised counterparts, a manual cluster-to-label assignment step is required by a domain expert. Nevertheless, this process is rather efficient as the learnt clusters, together with additional insights, can be presented to a human decision maker for label assignment, reducing the labelling effort to clusters instead of individual samples. Despite growing interest in self-supervised learning for wireless communications, fully label-free approaches, such as [32], remain rare. Most existing methods, including recent transformer-based foundation models [33], still rely on labels at some stage of the pipeline, whether for a supervised pretext objective, for fine-tuning, or for the downstream recognition task itself. Earlier works predominantly build on reconstruction-based approaches [34], [35], which emphasize instance learning, where the learning module (e.g., CNN, Self-attention, MLP) is trained solely or partially on reconstruction loss, focusing on distinguishing individual samples.

Recognizing the need for a more adaptable solution, particularly in scenarios where class information is unknown, in this dissertation, we investigate unsupervised deep learning (UDL), with an emphasis on deep clustering, as a scalable approach for learning informative representations from unlabeled radio frequency (RF) data with minimal expert intervention. We first establish the context through a survey of unsupervised deep learning for wireless signal recognition, considering two unique and complementary perspectives, i.e., wireless signal recognition problems and unsupervised deep learning approaches, while consolidating open datasets and outlining research opportunities toward UDL-based signal recognition.

Throughout this work, the proposed approaches operate on frequency-domain representations of the spectrum, namely spectrograms and one-dimensional FFT amplitudes, rather than on raw in-phase and quadrature (IQ) samples. This is a deliberate choice. Frequency-domain representations match raw IQ in recognition accuracy while being more robust at low signal-to-noise ratios, where IQ-based models degrade due to their higher noise sensitivity [13]. They are also far lighter in terms of data, requiring orders of magnitude lower transfer and storage rates [5], which is decisive for continuous, large-scale monitoring. Because amplitude-based representations discard phase, they retain all task-

relevant information, such as occupancy, bandwidth, timing, and activity patterns, while not preserving the content of the transmissions. However, IQ-based deep clustering [32] remains a valid alternative, motivated by adaptability across technologies without dataset-specific preprocessing.

1.2 Purpose of the Dissertation

This dissertation aims to serve two main purposes. First, it seeks to critically examine existing UDL approaches for radio signal recognition in wireless communications, from the complementary perspectives of prevailing challenges and employed methods, in order to establish a structured and analytical view of the field.

Building on this foundation, the dissertation further aims to advance spectrum sensing through the adoption and validation of existing deep learning architectures, as well as the design of dedicated, domain-specific models tailored to the characteristics of RF data. Ultimately, it seeks to reduce dependence on expert-labeled data and support the development of data-driven spectrum awareness as a foundational capability for future AI-native wireless communication networks.

1.3 Research Hypotheses

The work presented in this dissertation is structured around three hypotheses, stated as follows:

H1: Self-supervised deep learning architectures can effectively learn discriminative representations from unlabeled radio frequency spectrum data, reducing the need for expert annotation to cluster-level labeling rather than individual sample annotation.

H2: Adaptation of the existing unsupervised deep learning architectures to the specific radio frequency spectrum data characteristics can lead to significant complexity reduction of the models and features, while preserving or improving performance on downstream tasks such as classification.

H3: It is possible to develop unsupervised deep learning models that perform consistently across diverse real-world radio environments, characterized by differences in operating frequency and number of operational radio access technologies.

1.4 Methodology

The methodology used in this dissertation consisted of four key phases aimed at designing and studying the feasibility and effectiveness of a deep clustering architecture for spectrum sensing and radio signal recognition in real-world wireless environments.

Phase 1 - Observation: The first phase involved reviewing the existing challenges in the wireless communications domain related to radio signal recognition and their potential implications on the development of the next generation wireless networks, considering the current limitations and trends in the emerging technologies. Furthermore, a thorough analysis of the most relevant ML-based approaches was conducted to address the reviewed challenges, considering their complexity-to-performance trade-offs, explainability, data-specific requirements, and deployment challenges. These insights guided the development of a UDL-based solution tailored to RF signal characteristics.

Phase 2 - Hypotheses Formulation: Considering the initial observations, the key constraints of the existing deep ML-based approaches were identified, considering also the state of the existing technological options for their exploitation. The hypotheses were

proposed related to the advancement of the UDL-based techniques for RF signal recognition, framing it as a superior approach considering the current emerging concepts, such as O-RAN, DST and ISAC.

Phase 3 - Development and Experimentation: Two development strategies were pursued in the thesis, i.e. (1) adopting existing architectures from the computer vision domain (e.g., CNNs and ViTs) and (2) designing new architectures specifically adapted to RF signal data. The contemporary deep clustering techniques based on DeepCluster and autoencoders were developed and evaluated across two types of input formats: 2D spectrograms and 1D FFT amplitude vectors. All models were developed in PyTorch, with preprocessing and clustering supported by scikit-learn and custom Python scripts. The full implementation, including model definitions, training, and evaluation scripts, is openly available at [1], [2]. The ML models were evaluated with multiple parameter setups, considering the parameters that affect the learning process. Considering the complexity of the UDL approaches, their evaluation was performed by multiple quantitative and qualitative evaluation metrics, thus allowing for a comprehensive analysis of the performance from multiple perspectives. Quantitative analysis included standard clustering metrics (e.g., Normalized Mutual Information (NMI), Adjusted Rand Index (ARI), homogeneity, completeness) and distance-based metrics (e.g., silhouette score, Davies-Bouldin index, Calinski-Harabasz index). Qualitative insights were obtained through t-SNE (t-distributed stochastic neighbor embedding) visualizations and cluster-averaged spectrogram inspections to assess the interpretability and structure of the learned embeddings.

Phase 4 - Results Analysis: Evaluation was conducted on three diverse open real-world spectrum datasets collected from research testbeds in different regions, deliberately chosen to span complementary axes of difficulty over a wide frequency range, from below 1 GHz up to nearly 6 GHz. The *Technology Classification Dataset (TCD)*[3] [13] provides a labeled setting with three RATs originally captured at their technology-specific frequencies, namely DVB-T at 482 MHz, LTE at 806 MHz, and WiFi at 2.4/5.2 GHz. These recordings are normalized to a common 1024-bin FFT representation, yielding a single dataset in which the technologies appear as if coexisting in one channel, each with a distinct FFT signature. The *Intelligent Transportation Systems (ITS)* dataset[4] [30], reflects genuine coexistence in the safety-critical ITS 5.9 GHz band, where five technologies (LTE, 5G NR, WiFi, ITS-G5, and C-V2X PC5) share the spectrum. Its samples are captured within short acquisition windows representative of real-time sensing, requiring recognition from brief signal snapshots, and it includes spectrally similar pairs such as LTE and 5G that are more challenging to separate. The unlabeled, continuously sensed *LOG-a-TEC* dataset[5] [10], containing LoRa, IEEE 802.15.4, SIGFOX, and proprietary technologies transmissions in an ultra-narrow 868 MHz segment, most closely reflects real deployment conditions, exposing the models to genuine acquisition artifacts and interference. For the labeled datasets, ground-truth labels were used only for evaluation. This progression from controlled to uncontrolled transmission processes, across varied bands, RAT counts, and signal similarities, enables one of the most extensive evaluations of UDL-based RAT recognition reported to date.

Communication of Findings: The intermediate research results were presented through academic publications and open-source code releases to validate and disseminate the findings. Such engagement with the research community helped obtaining valuable feedback and fostered further developments in the field.

[1] `https://github.com/sensorlab/self-supervised-spectrum-sensing`
[2] `https://github.com/sensorlab/spire`
[3] `https://github.com/ewine-project/Technology-classification-dataset`
[4] `https://gitlab.ilabt.imec.be/mgirmay/tech-rec-its-band`
[5] `https://log-a-tec.eu/datasets.html`

Table 1.1: Mapping of the dissertation contributions to the core challenges in unsupervised RAT recognition, indicating the chapter in which each is addressed.

Challenge	**C1**	**C2**	**C3**	**C4**
Reliance on labeled data and prior knowledge of active RATs		Ch. 3	Ch. 4	Ch. 4
Consistent performance across diverse datasets and conditions	Ch. 2		Ch. 4	Ch. 4
Fragmented evaluation practices and limited dataset diversity	Ch. 2			
Model complexity and deployment constraints		Ch. 3		Ch. 4

Iteration: The proposed methodology was used in an iterative process where insights from experiments and community feedback were used to refine general approaches, model architectures, training procedures, and evaluation strategies to ensure robust and consistent results.

1.5 Contributions

Aligned with the overarching goal of advancing spectrum sensing through the deep clustering approach, this dissertation claims four main scientific contributions:

C1 Systematic review of UDL approaches for wireless signal recognition that, for the first time, jointly maps recognition challenges to unsupervised learning families and conducts a per-challenge analysis of validation approaches and metrics. The review exposes fundamental gaps in the field, including fragmented evaluation practices, limited dataset variation, and the lack of standardized validation protocols, and therefore provides actionable recommendations for future research.

C2 Design and development of a deep clustering architecture for RAT recognition based on spectrogram processing, by systematically adapting existing deep learning modules (ResNet, VGG and ViT) to the spectral domain and its data semantics, resulting in model size reduction of up to an order of magnitude, while simultaneously improving clustering performance by roughly a factor of 2–2.5 across the selected evaluation metrics.

C3 Design of a label-free spectrum sensing workflow that requires no prior knowledge of the active RATs and operating conditions during model training, supporting post-hoc class assignment via manual cluster-to-class mapping.

C4 Design of a lightweight, deep clustering CNN architecture tailored for FFT amplitude processing, achieving a reduction in parameter count and computational cost of up to two orders of magnitude compared to the spectrogram-based architecture, while improving feature separability and clustering quality.

1.6 Organization of the Thesis

This dissertation is based on three published journal articles and organized into five chapters that progress from a broad foundation to two complementary, original research contributions addressing label-free radio signal recognition.

Chapter 1 introduces the problem setting and motivates the scientific research of unsupervised and self-supervised learning in wireless communications, where labeled data are scarce and operational conditions vary widely. It then states the purpose of the dissertation, formulates the working hypotheses, and defines the concrete objectives and methodology used throughout the work. The chapter closes by summarizing the dissertation's contributions and outlining its content.

Chapter 2 is comprised of a journal article *Radio Signals Recognition with Unsupervised Deep Learning: A Survey*. It establishes the scientific context by systematically reviewing recent advances in deep unsupervised and self-supervised learning for wireless signal recognition. The chapter structures the literature along two complementary dimensions: the core recognition challenges (e.g., modulation recognition, signal sensing, specific emitter identification, anomaly detection) and the principal families of unsupervised/self-supervised methods (e.g., clustering-based, contrastive, reconstruction-based, and generative approaches). By consolidating representative datasets and synthesizing key findings, it identifies recurring limitations and articulates open research questions that motivate the methodological developments in the subsequent chapters.

Chapter 3 is comprised of a journal article *Self-supervised Learning for Clustering of Wireless Spectrum Activity Based on Spectrogram Processing*. It translates the insights from Chapter 2 into a practical self-supervised pipeline for discovering and clustering wireless spectrum activities from 2D spectrogram representations. It details the end-to-end methodology, including raw data preparation, the considered SSL and baseline architectures, and the feature and cluster development procedures. The chapter then introduces and applies a dedicated evaluation workflow combining embedding-space quality assessment, clustering metrics, and visualization analysis, to study design choices such as dimensionality reduction and backbone selection, and to validate the approach on representative use cases.

Chapter 4 is comprised of a journal article *Spectrum Sensing with Deep Clustering: Label-Free Radio Access Technology Recognition Based on FFT Amplitudes*, which complements the spectrogram-based approach from Chapter 3. It focuses on label-free RAT recognition from raw 1D FFT amplitude data within an explicit spectrum sensing workflow. It describes the proposed deep clustering architecture and its operational pipeline, and defines evaluation protocols covering clustering quality, transmission detection, and computational efficiency. It reports experimental results across multiple datasets and discusses practical deployment considerations, thereby positioning the proposed workflow as a step toward robust, environment-agnostic spectrum awareness.

Finally, Chapter 5 consolidates the scientific contributions and discusses their broader implications for data-driven spectrum sensing in future wireless systems. It closes by identifying the limitations of the present work and outlining directions for potential future research in unsupervised learning for wireless signal recognition.

Chapter 2

Unsupervised Deep Learning for Recognition of Radio Signals

In this chapter, we aim to provide a review of recent scientific works studying the utilization of unsupervised deep learning algorithms for the purpose of signal recognition. By the umbrella term signal recognition, we refer to four closely related and complementary tasks specific to the wireless communications domain, which are concerned with providing different types of awareness in the radio environment based on deep learning–driven intelligence. These tasks include modulation classification, signal sensing, specific emitter identification (SEI) and anomaly detection. The analysis is conducted from two complementary perspectives, namely, by considering unsupervised deep learning approaches and the domain-specific tasks that are suitable to be addressed with such methods.

This dual perspective offers clear insights into the capabilities and limitations of state-of-the-art models in addressing established signal recognition problems that underpin intelligent radio awareness in wireless networks, while also highlighting emerging challenges and promising directions for future research. Furthermore, through multi-angle analysis, such as small-sample performance vs. model sizes, we highlight the existing trends and emerging challenges resulting from the adoption of these new approaches, and support potential future work by collecting state-of-the-art methods and open datasets necessary for unhindered progress in this area. This is particularly important given that unsupervised deep learning approaches often rely heavily on large amounts of data, significantly more than their supervised counterparts, in order to achieve meaningful results.

This Chapter aquires methodological tools to support addressing hypotheses **H1**-**H3** and contains contribution **C1**.

The work presented in this chapter has been published in the journal article:

- L. Milosheski, B. Bertalanič, C. Fortuna, and M. Mohorčič, "Radio Signals Recognition With Unsupervised Deep Learning: A Survey," IEEE Access, vol. 13, pp. 217769–217798, 2025.

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 13 November 2025, accepted 15 December 2025, date of publication 23 December 2025, date of current version 31 December 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3647686

SURVEY

Radio Signals Recognition With Unsupervised Deep Learning: A Survey

LJUPCHO MILOSHESKI[1,2], (Member, IEEE), BLAŽ BERTALANIČ[1], (Member, IEEE), CAROLINA FORTUNA[1], AND MIHAEL MOHORČIČ[1,2], (Senior Member, IEEE)
[1]Department of Communication Systems, Jožef Stefan Institute, 1000 Ljubljana, Slovenia
[2]Jožef Stefan International Postgraduate School, 1000 Ljubljana, Slovenia

Corresponding author: Ljupcho Milosheski (ljupcho.milosheski@ijs.si)

This work was supported by Slovenian Research and Innovation Agency under Grant P2-0016.

ABSTRACT Optimization of wireless network parameters relies on the awareness of a dynamically changing radio environment, which depends on the presence of active devices characterized by various radio access technologies (RATs), modulation schemes, and overall spectrum usage patterns, and can be determined by advanced radio signal recognition methods. While various supervised machine learning (ML) models have been explored for signal recognition, their actual deployment has been limited so far due to challenges in acquiring labeled datasets. The emergence of Open Radio Access Network (O-RAN) architectures and open experimental testbed setups has enabled access to large-scale, unlabeled data through standardized interfaces, paving the way for unsupervised deep learning methods. These methods, unlike supervised approaches, require minimal labeled data and have shown promising results in domains such as computer vision and time-series processing. However, their application in wireless communications remains relatively unexplored. This survey aims to provide a comprehensive overview of unsupervised deep learning techniques for addressing key challenges for signal recognition in wireless communications, including automatic modulation classification (AMC), signal sensing, specific emitter identification (SEI), and anomaly detection. Specifically, we examine state-of-the-art approaches such as deep clustering, contrastive learning, autoencoder-based reconstruction, and generative models. Additionally, we discuss available open datasets and identify research opportunities to advance this field, leveraging the substantial successes of self-supervised learning in computer vision and natural language processing. By organizing the survey into two key complementary perspectives—wireless communication challenges and unsupervised deep learning solutions—this work provides a roadmap for researchers and practitioners seeking to develop innovative, data-efficient models for the next generation of AI-native wireless networks.

INDEX TERMS AMC, anomaly detection, radio signal recognition, SEI, signal sensing, spectrum sensing, unsupervised deep learning, wireless communications.

I. INTRODUCTION

Radio signal recognition concerns recognizing the presence, type, and characteristics of signals transmitted over the radio frequency (RF) spectrum. It is an essential task in wireless communications and spectrum management, supporting a variety of challenges. In this survey, we focus on four closely related and sometimes overlapping subdomains: modulation classification [1], signal sensing [2], specific emitter identification (SEI) [3], and anomaly detection [4]. These areas have been widely explored by the research community over the past decade, with a significant emphasis on deep learning approaches in the last five years. The awareness of the radio environment is important for optimizing various network parameters, such as selecting appropriate frequency channels, adjusting transmission power, choosing modulation types, and determining suitable radio access technologies. Timely information on the overall spectrum usage and activity from different aspects, such as the number of active devices, used radio access technologies (RATs), modulations, and

The associate editor coordinating the review of this manuscript and approving it for publication was Qiang Li.

© 2025 The Authors. This work is licensed under a Creative Commons Attribution 4.0 License.
For more information, see https://creativecommons.org/licenses/by/4.0/

anomalous signals, could enable more efficient and effective management of wireless networks, ensuring their stable performance and minimizing mutual interference.

Sensing capabilities offer significant benefits for both licensed and unlicensed spectrum networks, as direct feedback from the environment, for the purpose of network parameter control. In unlicensed spectrum, technologies like WiFi, LoRa, and Bluetooth use sensing to inform channel occupancy [5] to guide the selection of less congested channels, and detect interfering transmissions, which is increasingly common with the proliferation of wireless devices in everyday life.

Licensed wireless networks could benefit significantly from such environmental awareness, achieved through the different signal recognition capabilities. These proved the potential to further increase the efficiency of the spectrum usage, following the successfull applications in the past, such as the License Assisted Access (LAA) [6] at 2.4 GHz and 5 GHz providing coexistance of LTE with WiFi, as well as the 4G/5G sharing the 3.5 GHz Citizens Broadband Radio Service (CBRS) band with military radar [7]. This is highly relevant for the developments related to next-generation wireless networks such as 6G, considering that Integrated Sensing and Communication (ISAC) is envisioned as one of their key capabilities [8], [9]. Furthermore, signal recognition outcomes could be used for network parameters control and optimization, for objectives like interference detection (via spectrum sensing), improved security (through SEI), and traffic steering (based on RAT recognition). This is particularly the case in private network deployments of 5G networks, such as office buildings, and factories where the operator has no control of transmissions in similar neighbouring deployments. The utilization of such functionalities is enabled by Open Radio Access Network (O-RAN) [10], considering that they could easily be deployed as software applications within the non-real-time and near-real-time controllers (xApps and rApps) [11], leveraging open and standard interfaces [12] for data collection.

In addition to direct network parameter control, historical network monitoring data obtained through signal recognition functionalities can be used to build detailed environmental maps of network performance. They can form the basis for multimodal Digital Twins [13], i.e., virtual replicas of the network with its components and metrics such as latency and uplink/downlink speeds. These go beyond the existing concepts that provide only signal coverage [14], which is highly important for network planning and monitoring, by including multiple planes of data, such as mapping the uplink and downlink speed to the physical environment, thus paving the way for the emerging throughput intensive and low-latency [15] Extended Reality (XR) technologies [16].

All four subdomain tasks rely on processing radio signals, either in their raw format consisting of an in-phase and a quadrature (I/Q) component, usually utilized for SEI and AMC, or in transformed forms such as spectrograms based on FFT or wavelet transforms. These provide a pictorial format particularly suitable for existing image processing algorithms, thereby facilitating sensing capabilities. ML models are particularly suitable for these tasks [17] due to their capability to capture high-level correlations between temporal patterns, which in such use-cases of radio data are difficult or even impossible to detect with standard energy detection [18], [19] or cyclo-stationary momentum-based detectors [20], [21]. However, this comes at a price of increased computational complexity and lower explainability.

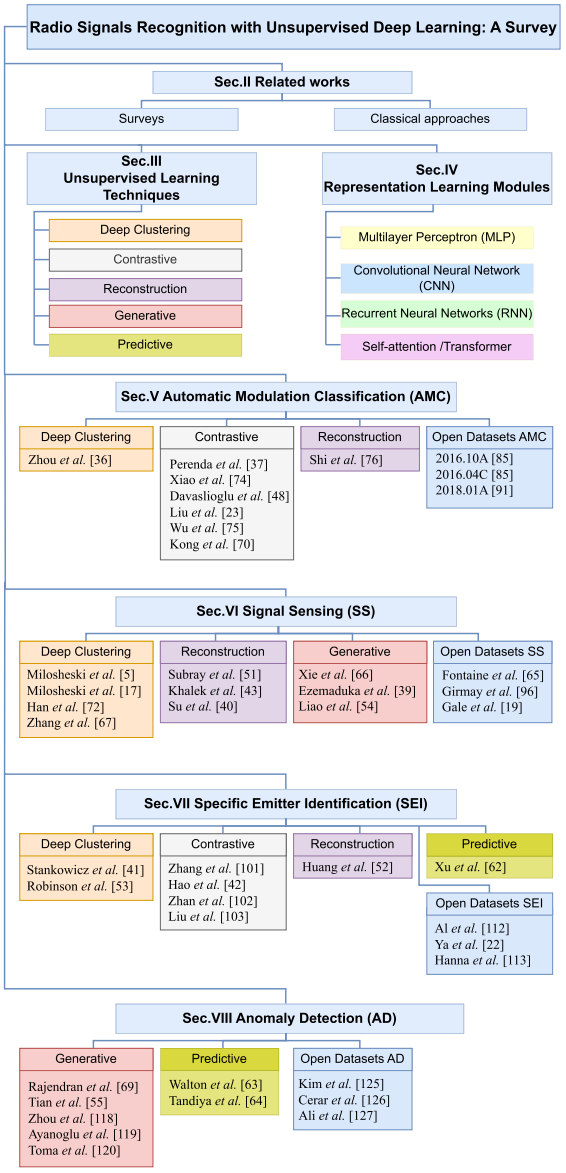


FIGURE 1. Overview of the survey structure, illustrating with color code the relationships between sections and the mapping of the surveyed works based on their employed approaches and addressed challenges.

While the use of machine learning (ML), including Deep Learning (DL), for signal recognition-related tasks is not new, actual deployment in real-world systems is still in its early stages. Supervised learning approaches have been extensively validated, and offer promising results [22], however, they face several limitations. Application is limited by the need for large labeled datasets, which are costly and time-consuming to obtain [19]. Furthermore, they face generalization problems due to the constraints of the labelled dataset sizes. Unsupervised deep learning, which includes several approaches that learn based on pretext tasks derived from the data itself, emerges as a promising alternative. It addresses the generalization and labeling problem by achieving comparable results to supervised methods while requiring only a fraction of the labeled data. This is due to the capability of automatic learning by processing large amounts of unlabeled data. These are now widely accessible, either generated from many available testbeds operating in the unlicensed bands or through the standardized O-RAN interfaces. As a result, unsupervised learning serves as a practical solution for initial model development, with supervised learning reserved for targeted fine-tuning in specific deployment scenarios [23].

A. CONTRIBUTIONS

The unsupervised deep learning approaches have matured in fields such as machine vision [24], [25], and general time series processing [26]. However, it is still gaining traction in the wireless communications domain, with limited existing research, given that its adoption is not straightforward, both in its raw and transformed formats. It requires communications-specific alterations, such as special data augmentation functions or feature extraction kernels, e.g., convolutional neural networks (CNNs), with particular capabilities, which are tailored for the specific semantics contained in the radio signals.

To the best of our knowledge, no existing survey summarizes recent advances in unsupervised deep learning for wireless communications while systematically categorizing learning approaches and their relation to key signal recognition challenges. This work addresses that gap by providing a comprehensive overview of deep unsupervised learning approaches applied to wireless communications, along with open datasets to support and inspire further research. Given the remarkable progress in computer vision and natural language processing through self-supervised models, exploring their potential in this domain is both timely and promising. In the most general sense, our work focuses on a review of the existing work from two perspectives:

1) **Review of core wireless signal recognition-related challenges addressed with unsupervised deep learning**: These challenges include AMC, signal detection, SEI, and anomaly detection. Considering that these challenges are well-known and established research fields in the domain, each backed by a large corpus of publications, we adopt the existing categorization published in [28].
2) **Review of unsupervised deep learning solutions in communications**: Such approaches include deep clustering, contrastive learning, autoencoder-based reconstruction, and generative models in relation to the key signal recognition tasks within wireless communications. Specifically, we further build on the existing segmentation provided in [26]. For each specific research challenge, we examine which unsupervised learning approaches have been applied to improve performance.

These two analytical review perspectives ultimately lead to the final and most important contribution of our work as a survey, i.e., the **identification of open research problems and future directions** regarding the application of unsupervised deep learning to the wireless communications domain. More specifically, in addition to surveying current challenges and approaches, our work highlights existing gaps and unresolved issues where unsupervised learning methods could offer innovative solutions.

B. STRUCTURE OF THE PAPER

In Figure I, we summarize the structure of the entire survey, indicating the main parts and their relations, which will be followed through the discussions in the text. Section II reviews existing surveys and related work, highlighting their contributions and limitations while positioning our study within this research landscape. Section III introduces the fundamental unsupervised deep learning approaches applied in wireless communications, including deep clustering, contrastive learning, reconstruction-based methods, generative models, and predictive learning. Section IV describes the commonly used types of learning modules for those approaches. Sections V-VIII review the most recent works, report on open datasets, and discuss potential future work regarding signal recognition and classification for AMC, signal sensing, SEI, and anomaly detection. Section IX highlights the observed challenges and potential future work directions, while Section X presents the conclusions of this study.

II. RELATED WORK

Several surveys have addressed ML applications in radio signal recognition, focusing on various aspects. However, a closer look shows that the unsupervised deep learning is treated only marginally, either as a subcategory of deep learning, [27], [28], [30], [31], [32], or completely omitted due to the focus on particular application domains such as dynamic spectrum access [29], and attacks modeling [33]. This leaves a notable gap in the coverage of unsupervised deep learning methods for radio frequency signals, which in recent years have become increasingly relevant, considering their potential for automatic signal

representation learning, as well as the availability of large open datasets.

A. GENERAL ML IN COMMUNICATIONS

Among the most comprehensive surveys in the field, [30] provides an extensive overview of deep learning algorithms applied to mobile and wireless networking, along with the enabling technologies supporting their deployment. The authors systematically review state-of-the-art models and offer a fine-grained categorization linking specific algorithms to the problems they address, thereby serving as a valuable reference for researchers and practitioners.

While slightly dated, [31] remains an influential and insightful contribution, offering a broad overview of deep learning applications in wireless networks from an application perspective. The survey organizes prior work according to the layers of the communication network stack, systematically analyzing how deep learning techniques have been applied to the physical, data link, network, and higher layers. This layered categorization provides a useful framework for understanding the scope and diversity of deep learning use cases in wireless systems. A related effort by [32] adopts a similar structure but focuses on applications within the Internet of Things (IoT) domain, addressing comparable challenges from a connectivity and device-integration standpoint. While both surveys introduce the main algorithmic paradigms, including supervised and unsupervised learning, their treatment of unsupervised deep learning remains high-level and mostly conceptual.

In [28], the authors provide a clear categorization of the existing challenges that are being addressed in the domain and a comprehensive overview of existing works from the perspective of the used dataset. They also provide insightful recommendations on the creation of future radio datasets. We take into consideration this categorization and complement it by reviewing the approaches that are used to address the discussed challenges, focusing on the unsupervised learning approaches, which we believe are highly relevant.

B. RADIO SIGNAL RECOGNITION

In the recent work [27], the authors provide a broad overview of machine learning (ML) approaches for spectrum sensing, focusing on supervised learning techniques, while paying limited attention to unsupervised deep learning, restricted to a brief mention of autoencoders. Such coverage overlooks other major unsupervised paradigms, such as contrastive learning and deep clustering, which have recently demonstrated strong potential for representation learning without labeled data [36], [37]. Furthermore, authors highlight the importance of large-scale radio frequency (RF) datasets for developing pre-trained models and enabling transfer learning; however, the discussion remains general and lacks concrete dataset references. As summarized in Figure I, we analyze these aspects in depth throughout the sections of this survey.

In [29], the authors provide a concise overview of the sensing-related challenges in the radio frequency (RF) domain addressed through machine learning (ML), focusing on cognitive radio functions such as spectrum sensing and dynamic spectrum access. The review includes supervised and reinforcement learning approaches, and, unlike our work, they offer no examination of unsupervised deep learning approaches from the perspective of their relevance for the pretraining phase of supervised approaches. This aspect is covered in our work in Sections V and VII.

In a similar manner, [35] provides systematic coverage of AI-based methods for the tasks of automatic modulation classification (AMC), signal detection, channel estimation, and MIMO beamforming. The work systematically reviews AI-based methods and highlights their potential for enabling intelligent and adaptive physical-layer processing. However, it gives limited attention to unsupervised deep learning approaches, which we systematically analyze in Section III and review in Sections V to VIII of this work as depicted in Figure I.

The most closely related work to ours is [34], which provides one of the first comprehensive surveys dedicated to unsupervised ML techniques in networking. Although dated (2019), it remains a valuable reference, as it captures fundamental challenges and methodological directions that are still relevant to current research. However, given the rapid evolution of deep learning in recent years, the survey is gradually becoming outdated, considering the recent emergence of transformer-based architectures and the contrastive learning paradigm, both of which have driven significant progress in self-supervised representation learning for radio signal processing.

C. OUR WORK

Considering the existing categorization of the domain-related challenges in prior reviews, the classification of unsupervised deep learning methods from closely related fields [24], [33], [38], and the identified gap in reviewing the recent advances in their use in radio signal recognition, we are focusing on these approaches from the perspective of wireless networks communications.

Summary of the related works, considering their key contents and shortcomings, as well as how we complement them in our work, is given in Table 1.

D. CLASSICAL UNSUPERVISED METHODS

While recent research in unsupervised learning increasingly emphasizes deep learning techniques due to their superior performance in feature extraction, classical unsupervised methods such as Energy Detector (ED), K-means clustering and its variants, and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) remain highly relevant. These traditional approaches are frequently used as benchmarking baselines [5], [36] or as components within more advanced frameworks, such as deep clustering [17], where they are combined with neural network-based feature extractors. The following section provides a concise overview of recent works that deploy non-deep learning unsupervised methods.

TABLE 1. Summary of related works concerning unsupervised learning for wireless communication.

Ref.	Focus Area	Strengths	Limitations	Our Work
[27]	ML-based spectrum sensing with emphasis on supervised learning	Comprehensive overview of supervised learning and importance of RF datasets for pre-trained models	Brief coverage of unsupervised learning (only mentions Auto-encoders); lacks references to datasets/testbeds	Provides a detailed review of unsupervised learning techniques (e.g., contrastive learning, deep clustering) and their application to RF signal recognition challenges
[28]	Challenges in spectrum sensing and dataset usage	Clear categorization of challenges; recommendations for future dataset creation	Limited discussion on unsupervised deep learning	Builds on the dataset categorizations and challenges outlined, focusing on how unsupervised learning addresses these issues
[29]	Spectrum sharing in cognitive radio	Relevant to environmental adaptiveness and 6G ISAC concept	Short and shallow mention of unsupervised deep learning	Highlights the potential of unsupervised learning to enhance adaptability and efficiency in 6G spectrum sharing
[30]	Deep learning algorithms for wireless communication	Most comprehensive survey; fine-grain categorization of ML problems	Outdated coverage of unsupervised learning due to publication year	Expands on unsupervised learning advances (e.g., generative models, self-supervised techniques) not covered extensively in prior works
[31]	ML applications in network architecture	Covers relevant research topics from the application perspective	Outdated (6 years old); lacks specifics on unsupervised deep learning	Connects unsupervised learning to well-defined application challenges
[32]	IoT perspective on ML in wireless communication	Domain-related insights; discussion on algorithmic approaches	Fundamental coverage without details on unsupervised deep learning	Focuses on unsupervised deep learning approaches, considering the latest advances and existing datasets for further research relevant for radio signal recognition in a broader context
[33]	Adversarial Machine Learning in wireless communications	Insights into unsupervised ML for modeling security threats	Focuses on the specific use case of attacks modeling	Reviews unsupervised learning approaches for signal recognition in wireless communications
[34]	Domain challenges and unsupervised learning	Covers unsupervised approaches; highlights challenges and data-driven models	Slowly becoming outdated due to recent advancements concerning transformers and contrastive learning	Highlights gaps in most recent works; emphasizes open datasets crucial for development of unsupervised learning
[35]	Intelligent signal processing in wireless communications	Covers four topics for the physical layer, modulation classification, signal detection, beamforming, and channel estimation	Covers machine learning approaches in a wider sense, without focus on unsupervised deep learning	Focuses on unsupervised deep learning approaches for specific wireless communications signal recognition challenges

The Energy Detector is one of the simplest and most widely used methods for signal detection, especially in spectrum sensing. It operates by measuring the energy of a received signal over a predefined observation interval and comparing it to a threshold. If the measured energy exceeds this threshold, the presence of a signal is assumed; otherwise, the channel is considered idle. Its performance is highly sensitive to noise uncertainty and can degrade significantly in low Signal-to-Noise Ratio (SNR) conditions. In recent works, it is usually deployed as a baseline approach for benchmarking purposes [39], [40].

K-means clustering is a partition-based clustering algorithm that aims to divide a dataset into K non-overlapping clusters. It is simple to use and requires only the number of clusters as a single parameter for initialization, which is independent of the feature space. Thus, it is the most frequently used non-deep learning approach, either as part of more complex deep-learning architectures [5], [17] such as deep clustering, or as a benchmarking baseline in its original form [5], [36], [41], [42]. Other clustering algorithms, for instance DBSCAN [41], [42], GMM [43], Agglomerative and Spectral clustering [36], are typically used as baselines due to their higher complexity initialization parameters that require feature-space specifics, such as distance between samples.

In Figure 2, we show the performance of the most commonly used classical unsupervised methods given as relative performance to the deep learning-based counterparts, proposed in the corresponding works. On the x-axis are the reference numbers of related works. On the y-axis is the relative performance given as percentages of the top-performing deep-learning counterpart (marked on top of each bar) evaluated in similar conditions, considering the reported metrics (also marked on top of each bar) in the corresponding works, for each of the three reviewed radio signal recognition tasks. No such data was available for the reviewed works in the Anomalies Detection task; thus, we consider only the AMC, SS, and SEI tasks. As can be seen from the plot, the classical approaches consistently show lower performance compared to the deep learning-based approaches across all three tasks and a variety of metrics. Only two cases ([39] and [43]) reach a significant 80% of the performance of the deep learning counterpart proposed in the corresponding work. However, such performance is observed only in the binary classification problems (note that the metric is Pd or Pd_{AUC}), while for the more complex tasks with multiple classes, the performance is significantly worse.

III. UNSUPERVISED DEEP LEARNING TECHNIQUES IN WIRELESS COMMUNICATIONS

Clear categorization of the Unsupervised Deep learning algorithms is still an open discussion, depending on the perspective from which the approaches are being observed. For example, in [44], authors make a distinction based on the way the pseudo-labels are generated, while in [26], authors

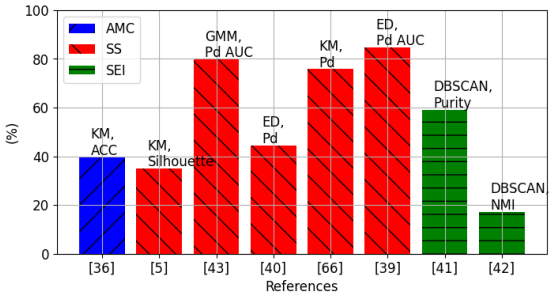


FIGURE 2. Relative performance of classical unsupervised algorithms compared to the deep learning approaches in the surveyed works as reported in the literature. Labels on bar charts: Method, Metric. Methods: KM=K-means, GMM=Gaussian Mixture Model, ED=Energy Detector, DBSCAN=Density-Based Spatial Clustering of Applications with Noise. Metrics: Pd=Probability of Detection, NMI=Normalized Mutual Information, ACC=Clustering Accuracy, Purity=Clustering Purity.

consider the final training phase only. Considering these two publications and the pool of relevant surveyed works, we distinguish five different unsupervised deep learning approaches utilized in signal recognition in the wireless communications domain: *deep clustering*, *reconstruction-based*, *generative*, *contrastive*, and *predictive*. In the following, we briefly introduce each of them, without going into fine-grained details, for which some of the early original works are referenced.

A. DEEP CLUSTERING METHODS

Deep clustering methods [45], illustrated in Figure 3, integrate deep learning with clustering techniques to extract semantically meaningful representations from high-dimensional RF data and to group similar signal patterns in an unsupervised fashion. In this context, neural networks are trained to map raw or preprocessed RF inputs, such as I/Q samples, spectrograms, or just averaged FFT amplitudes, into a latent space where clustering algorithms (e.g., K-means or Gaussian Mixture Models) can distinguish between signal classes [17]. The training process jointly optimizes two objectives: (1) representation learning, which structures the latent space to capture key features of RF signal variability (e.g., hardware impairments in the I/Q components for emitter identity, or amplitudes patterns at different frequencies for signal sensing), and (2) unsupervised clustering, which promotes the separation and compactness of signal clusters in the learned embedding space [17].

B. CONTRASTIVE METHODS

Self-supervised contrastive learning [46], [47] is a powerful representation learning paradigm that enables the extraction of discriminative features from unlabeled data (including RF data) by contrasting positive and negative signal instances. The core idea is to maximize the similarity between positive pairs, which are derived from the same RF signal sample through domain-specific augmentations (e.g., zero-masking, time-shifts, noise injection [48]), while minimizing similarity with negative pairs, which correspond to embeddings of different RF signals. Both augmented views of the same signal are passed through a shared neural encoder to generate feature embeddings, followed by a projection head that maps these into a latent space where a contrastive loss function (e.g., InfoNCE [49]) is applied, as depicted in Figure 4. This contrastive objective encourages embeddings of similar RF patterns (e.g., same emitter or modulation scheme) to cluster together while pushing dissimilar signals apart [42], thereby enabling robust, generalizable representations useful for downstream tasks such as AMC and SEI.

C. RECONSTRUCTION-BASED METHODS

Reconstruction-based methods [50] are unsupervised learning techniques particularly well-suited for RF signal processing. The goal is to learn compact and informative representations by reconstructing input RF signal data from compressed or latent representations. The basic model is the autoencoder (AE), a symmetric encoder-decoder architecture, visualized with violet in Figure 5a. An encoder transforms raw RF signals, such as raw I/Q samples or spectrograms, into a lower-dimensional latent space, and a decoder attempts to reconstruct the original signal from this compressed representation. The objective is to minimize the reconstruction error, ensuring the learned representations capture the most salient features of the RF data, such as modulation, temporal pattern, or energy burst shapes across the frequency domain. These approaches are widely used for signal recognition tasks, such as RAT classification [51], primary user (PU) recognition [43], and denoising [40]. By optimizing the reconstruction, these baseline autoencoders provide a robust framework for learning compact and informative representations suitable for downstream tasks, such as SEI [52], after further tuning of the encoder in a supervised manner.

However, relying solely on reconstruction loss, the autoencoder may not necessarily produce a compact and interpretable feature space of the input RF signals. This lack of structure can make it difficult to analyze or utilize the encoded representations in a completely unsupervised

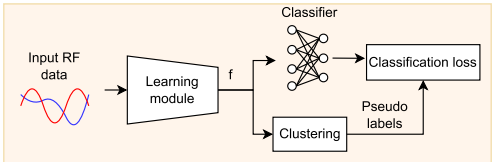


FIGURE 3. General structure of *Deep clustering* methods.

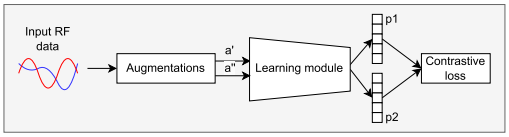
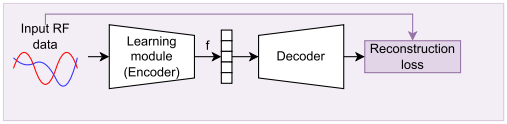
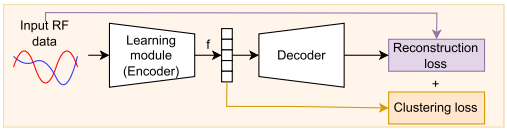


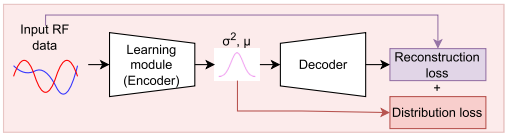
FIGURE 4. General structure of *Contrastive* learning methods.



(a) Autoencoder.



(b) Autoencoder-based Deep Clustering.



(c) Variational autoencoder.

FIGURE 5. General structure of *Reconstruction-based* learning methods.

setup [5], [19], where fine-tuning is not an option. To address this, introducing a clustering loss in the feature space f, as in [36], can guide the encoding process toward a more cluster-friendly distribution of samples. This adjustment makes the learned representations more suitable for further analysis, especially when labels are entirely absent. Such modifications in the learning process give rise to reconstruction-based deep clustering methods that are used for self-supervised RF signal feature learning as pretraining, for downstream tasks such as clustering for novel (unseen in training) device detection [41], [53]. Its functional structure is highlighted in orange in Figure 5b.

An important extension of the reconstruction-based approach for RF signal representation is the Variational Autoencoder (VAE), depicted in red in Figure 5c. Unlike standard autoencoders, which map each input RF signal to a fixed point in latent space, VAEs model the latent space as a multidimensional probability distribution, typically Gaussian. For each input, such as an I/Q sample stream [39] or spectrograms [54], [55], the encoder outputs parameters of a distribution (mean and variance) rather than a single deterministic code. The decoder then reconstructs the signal by sampling from this learned distribution. This probabilistic framework provides a continuous and structured latent space, which enables smooth interpolation between different RF signal samples and supports generative capabilities such as the synthesis of novel RF-like samples.

D. GENERATIVE METHODS

Generative methods [56] are a class of machine learning techniques designed to model the underlying data (including RF data) distribution and generate new, realistic samples that resemble the original dataset. These methods are widely used for tasks like image synthesis, data augmentation, and representation learning. Among the most prominent generative approaches besides VAEs discussed in the previous subsection are Generative Adversarial Networks (GANs). GANs generally consist of a generator that produces synthetic data and a discriminator that evaluates their authenticity, training them in an adversarial manner to improve the quality of generated samples as depicted in Figure 6.

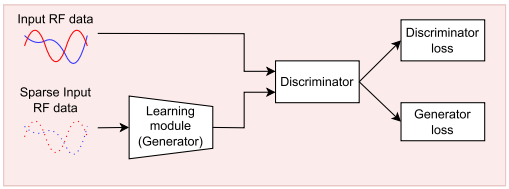


FIGURE 6. General structure of *Conditional Generative Adversarial Networks*.

The GAN architecture has seen wide application in the physical layer of wireless communications [57], [58], usually in its conditional variant, cGAN, illustrated in Figure 6. In cGAN, instead of with random input, the generator is provided with sparse points of the actual input data. The use of GANs mainly focuses on pre-processing of radio signals, such as enhancing [59] and generating new data samples [60] for data-constrained use-cases. Such approaches could contribute to improving signal recognition [61]; however, many such solutions require labeled data for training, such as modified spectrograms in [54], thus positioning them as edge cases with regard to the scope of this survey.

E. PREDICTIVE METHODS

Self-supervised predictive learning is a representation learning method that leverages unlabeled data by training models to predict certain aspects or features of the data itself. As illustrated in Figure 7, they typically consist of a representation learning module followed by a classifier and a classification loss module.

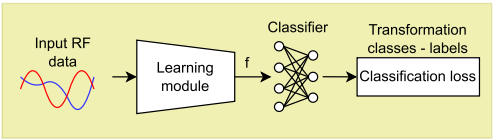


FIGURE 7. General structure *Predictive* methods.

In the RF domain, predictive self-supervised tasks are designed to exploit the temporal, spectral, and structural continuity inherent in RF signals. These tasks involve learning to predict transformations or future states of the signal, encouraging the model to internalize meaningful representations without labeled data. Typical examples include: predicting artificially applied transformations such as rotation classification [62]; forecasting the next I/Q symbol in a time series to model fine-grained temporal dependencies [63]; and predicting the next spectrogram frame to detect anomalies as deviations from expected spectral behavior learned during training [64]. These pretext tasks promote the learning

of statistical regularities—such as modulation structure, spectral patterns, and temporal correlations in I/Q data—making the resulting representations highly informative for downstream applications like anomaly detection, AMC, or SEI.

F. OVERVIEW

We summarize the discussed unsupervised deep learning methods in Figure 8 using a Venn diagram to illustrate the relationship between them and to map the corresponding works in the wireless communications domain that will be surveyed through Sections V-VIII.

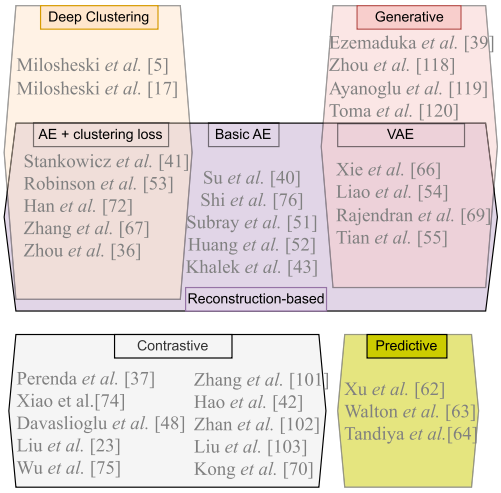


FIGURE 8. Unsupervised deep learning approaches. Color coding corresponds to Fig. I.

Each approach includes specific types of architectures, yet there are overlapping cases, particularly with autoencoders, which can be categorized differently depending on how their loss functions are modified. Standard autoencoders, as a reconstruction-based approach, consist of an encoder-decoder structure optimized using reconstruction loss. This provides a straightforward way to feature extraction in an unsupervised manner, capturing essential data representations in a low-dimensional latent space. Modifications to their loss functions improve different aspects of their functioning, thus placing them into different categories, even though they retain their fundamental structure and functionality of encoding input data into a lower-dimensional representation and reconstructing it at the output. Introducing a clustering loss in the latent space, in addition to the standard reconstruction loss (such as a distance-based loss), leads to the formation of more cluster-friendly feature spaces. This adaptation aligns autoencoders more closely with deep clustering methods. Alternatively, replacing the latent space loss with a statistical loss, while incorporating reparameterization to ensure gradient flow during backpropagation, transforms the model into a generative framework. This modification enables the generation of new samples by sampling from a smooth, continuous latent distribution, making it a VAE-based generative model.

As shown in Figure 8, contrastive and predictive learning approaches remain distinct categories that do not yet overlap with reconstruction-based, deep clustering, or generative models.

IV. NEURAL ARCHITECTURES FOR THE REPRESENTATION LEARNING MODULES

In this section, we shortly outline commonly used types of learning modules as part of previously described techniques in Section III.

A. MULTILAYER PERCEPTRON (MLP)

Multilayer Perceptrons are the most fundamental form of neural networks, consisting of fully connected layers where each neuron in one layer connects to all neurons in the next layer. They are composed of an input layer, one or more hidden layers, and an output layer, as visualized in Figure 9. Each layer performs a linear transformation followed by a nonlinear activation:

$$\mathbf{y} = \sigma(\mathbf{W}\mathbf{x} + \mathbf{b}), \tag{1}$$

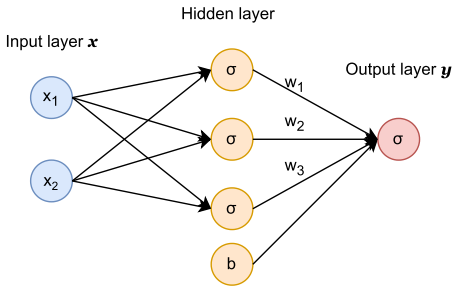


FIGURE 9. Multilayer Perceptron illustration.

where $\mathbf{x}$ is the input, $\mathbf{W}$ the weight matrix, $\mathbf{b}$ the bias, and $\sigma(\cdot)$ is the activation function (e.g., ReLU, Sigmoid).

In the wireless communications domain, MLPs are rarely used for direct processing of raw data, and usually only for low-density inputs, such as Received Signal Strength Indicator (RSSI) measurements. This is mainly because MLPs exhibit a high parameter count that increases rapidly with input dimensionality, resulting in substantial computational and memory demands. Furthermore, they are weak at capturing local dependencies, since they give equal weight to each input. However, they still offer decent performance [65], [66] if appropriate inputs, such as statistical moments, are provided. In recent works [5], [51], [67], the MLPs are mostly used as building blocks of more complex architectures, such as classification layers for the CNN-provided feature spaces, as will be discussed later in this section and visualized in Figure 13. Regarding the setup requirements, the MLPs are fairly easy to utilize since they require specification of a relatively small number of parameters. Besides the standard parameters which are common for most

of the backpropagation-based learning architectures, such as optimizer, learning rate, batch size, etc., these include the activation function and the number of layers and neurons per layer.

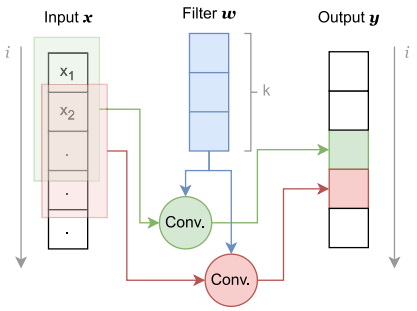


FIGURE 10. Single layer of 1D Convolutional Neural Network illustration.

B. CONVOLUTIONAL NEURAL NETWORK (CNN)

CNNs are designed to process grid-like data, such as images or time series, by applying convolution operations to extract local features. A convolution layer applies filters (also referred to as kernels) that slide over the input, capturing spatial hierarchies. In the following, we formulate a one-dimensional (1D) CNN, also depicted in Figure 10, which is suitable for sequence processing. The extension to 2D, which is also commonly used when a sequence to image transformation is applied on the input data, is straightforward. Given an input sequence x and a filter (kernel) w of length k, the 1D convolution operation at position i is calculated as:

$$y(i) = \sum_{j=0}^{k-1} w(j) \cdot x(i+j), \tag{2}$$

where $y(i)$ is the output feature map at position i, $x(i+j)$ is the input value at position $i+j$ in the sequence and $w(j)$ is the weight of the kernel at position j. This operation is repeated across the entire sequence to form a complete output $\mathbf{y}$.

CNNs are the most widely used learning modules in the wireless communications domain, potentially as part of any of the discussed unsupervised deep learning approaches in Section III, due to their ability to capture and distill spatial correlations. Their hierarchical feature extraction provides strong generalization for structured inputs. However, their performance depends on more parameters, which have to be defined and are application dependent, such as filter sizes, stride, padding, and pooling strategy, which makes them more complex to set up compared to the MLPs. On the other side, they can be directly used on the raw data, avoiding manual feature engineering as it is usually needed with MLPs, which makes them more flexible and applicable.

There is a variety of existing implementations of CNN, which include different filter configurations and interconnection of layers (Visual Geometry Group (VGG), Residual Network (ResNet), GoogLeNet, etc.), enabling solutions of various complexity. While they exhibit great performance for time series or pictorial data processing and extraction of spatial correlations, they lack the ability to capture long-term dependencies, which are specific to I/Q signal components in wireless communications data. One way to address this is by increasing the receptive field by introducing dilated convolutions [68], or simply increasing the filter sizes.

C. COMMONLY USED TYPES OF RECURRENT NEURAL NETWORKS (RNN)

RNNs, depicted in Figure 11a are tailored for sequential data processing with variable length, where each neuron's output depends not only on the current input but also on the previous hidden state. Theoretically, they could be used in any of the unsupervised deep learning approaches discussed in Section III; however, there are no works adopting them in the Deep Clustering approach (Section III-A). There are works with contrastive (Section III-B) [23], generative (Section III-D) [69], predictive (Section III-E) [63] and reconstruction-based (Section III-C) [40] approaches. Their formal definition is given as:

$$h_t = \sigma\left(W_h h_{t-1} + W_x x_t + b\right), \tag{3}$$

where W_h is the weight matrix for the previous hidden state, W_x is the weight matrix for the current input, b is the bias vector and σ is an activation function. The output y_t is the same as the hidden state h_t. While they can capture temporal dependencies, their performance suffers from vanishing gradients, leading to forgetting longer-term dependencies. For the sake of simplicity, the bias vectors which are existing for each weight matrix are not visualized in Figure 11.

1) LONG SHORT-TERM MEMORY NETWORKS (LSTMS)

LSTMs, depicted in Figure 11b, address the limitations of standard RNNs by introducing gating mechanisms that control the flow of information through a cell state. The gates include input (i_t), forget (f_t), and output (o_t) gates.

The forget gate f_t controls which part of the information is discarded from the previous cell state c_{t-1}, modeled as:

$$f_t = \sigma\left(W_{fh} h_{t-1} + W_{fx} x_t + b_f\right), \tag{4}$$

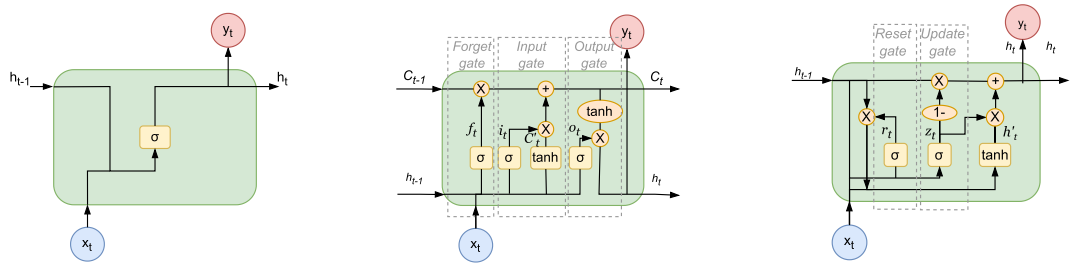
where σ is the sigmoid activation function, W_{fh} and W_{fx} are weight matrices for the hidden state and the input, and b_f is the bias term.

The input gate i_t controls which values will be updated in the cell state, i.e.

$$i_t = \sigma\left(W_{ih} h_{t-1} + W_{ix} x_t + b_i\right). \tag{5}$$

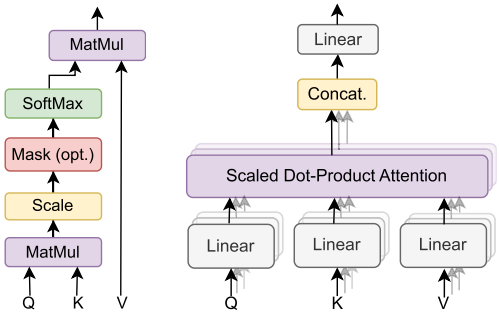
The candidate cell state $\tilde{c}_t$ generates the candidate values that can be added to the cell state, according to:

$$c_t' = \tanh\left(W_{ch} h_{t-1} + W_{cx} x_t + b_c\right). \tag{6}$$



(a) Basic RNN. (b) LSTM with forget gate. (c) GRU.

FIGURE 11. Types of recurrent neural networks.



(a) Self-attention module. (b) Multihead attention module.

FIGURE 12. Illustration inspired by [71].

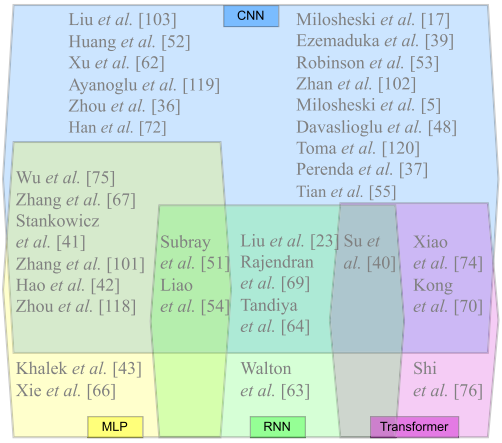


FIGURE 13. Venn diagram of the surveyed works with regard to the corresponding deep learning module used. Color coding corresponds to Fig. I.

The new cell state c_t is a combination of the old cell state and the candidate values:

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t. \tag{7}$$

The output gate o_t determines which parts of the cell state contribute to the hidden state:

$$o_t = \sigma\left(W_{oh} h_{t-1} + W_{ox} x_t + b_o\right). \tag{8}$$

And finally, the hidden State h_t is the actual output from the LSTM cell:

$$h_t = o_t \cdot \tanh\left(c_t\right). \tag{9}$$

The introduction of the three gates enables LSTMs to capture and model longer-term temporal dependencies. Thus, these models are suitable for time-series processing in general, including wireless communications signals. This approach is also one of the most common approaches for addressing classification problems in wireless communications. However, they are computationally more expensive to train compared to the basic RNNs.

2) GATED RECURRENT UNITS (GRUS)

GRUs, depicted in Figure 11c, are a simplified variant of LSTMs that combine the forget and input gates into a single update gate and use a reset gate.

The reset gate r_t determines how much of the past information to forget, according to:

$$r_t = \sigma\left(W_{rh} h_{t-1} + W_{rx} x_t + b_r\right), \tag{10}$$

where σ is a sigmoid activation function, W_{rh}, W_{rx} are the weight matrices for the previous hidden state and the current input, respectively, and b_r is the bias term for the reset gate.

The update gate z_t controls how much of the previous hidden state h_{t-1} is retained versus updated, according to:

$$z_t = \sigma\left(W_{zh} h_{t-1} + W_{zx} x_t + b_z\right). \tag{11}$$

The candidate hidden state $\tilde{h}_t$ incorporates the reset gate to decide which part of the past information to forget before applying a non-linear transformation, following:

$$h'_t = \tanh\left(W_{hh}\left(r_t \cdot h_{t-1}\right) + W_{hx} x_t + b_z\right). \tag{12}$$

The final hidden state h_t is a combination of the previous hidden state and the candidate hidden state, controlled by the update gate, calculated as:

$$h_t = (1 - z_t) \cdot h_{t-1} + z_t \cdot h'_t. \tag{13}$$

GRUs are computationally more efficient than LSTMs while achieving comparable performance. Besides the basic training parameters, such as learning rate and activation function, these types of networks also require specifying

the number of layers, the number of hidden units per layer, and the sequence length, which are the main drivers of the final model's complexity. Due to recurrent dependencies, the setup requirements are typically more complex but still comparable to those of CNNs. Furthermore, they are more time-consuming to train because they process data sequentially, which limits parallelization.

D. TRANSFORMERS (ATTENTION MECHANISMS)

The transformer architecture, originally designed for natural language processing, has proven highly effective for time series modeling due to its ability to capture both short- and long-term dependencies. While this learning module could be used in any of the approaches discussed in Section III, the existing works are only concerned with the contrastive (Section III-B) [70] and reconstruction-based (Section III-C) [40] learning approaches. The relatively low number of works utilizing transformers is due to the freshness of such architectures in the wireless communications domain, thus being still in the early adoption stage.

Transformers rely entirely on attention mechanisms rather than recurrence or convolutions. The self-attention mechanism computes the importance of different parts of a sequence relative to each other, given as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V. \quad (14)$$

As illustrated in Figure 12a, the core component is the scaled dot-product attention, where query (Q), key (K), and value (V) matrices are processed through a series of linear transformations and attention mechanisms. The attention scores are computed by taking the dot product of Q and K, scaled by the square root of the key dimension d_k, and passed through a softmax function to produce attention weights. These weights are applied to V to obtain the context-aware output.

To enhance the model's capacity, multi-head attention, illustrated in Figure 12b, runs several attention layers in parallel, allowing the model to focus on different representation subspaces simultaneously. This parallelism and flexibility make transformers particularly suitable for capturing complex temporal patterns in time series data without relying on recurrence. The main parameters defining the configuration of an attention mechanism are the embedding dimension, number of attention heads, number of layers, and feed-forward network size. Self-attention models are generally more complex than both LSTMs and CNNs, particularly in terms of computational and memory costs, due to a quadratic increase in complexity with sequence length. Unlike LSTMs, attention mechanisms are highly parallelizable, which makes them faster to train on GPUs despite their higher theoretical complexity.

E. MIXED LEARNING MODULES

Besides the standalone usage of the learning modules in the approaches discussed in Section III, their combinations proposed in multiple works [40], [62], [72] prove to be even more suitable for unsupervised deep learning. The use of different types of neural networks as learning modules in combination allows for capturing different types of dependencies that could exist in the input signals. For instance, consecutive CNN and LSTM [73] can be utilized for capturing fine-grain as well as temporal dependencies, while in a combination of CNN and a transformer [74], CNN provides the temporal dependencies encoded as input vectors for the transformer.

F. OVERVIEW

In Figure 13, we present a Venn diagram visualization of the surveyed works based on the learning models employed. As expected, the vast majority utilize CNNs, likely due to their widespread adoption in related fields, straightforward implementation, and competitive performance. The second most commonly used models are Multi-Layer Perceptrons (MLPs); however, they typically appear as components within more complex architectures that primarily rely on CNNs. This preference is largely driven by the high data density of radio signals (e.g., I/Q samples or spectrograms), which poses processing challenges for standard neural networks. Interestingly, only a small number of works employ Recurrent Neural Networks (RNNs), despite their suitability for sequential data. This may be attributed to their higher complexity and comparable performance to CNNs. Currently, transformer-based models (e.g., multi-head attention mechanisms) are the least frequently adopted. Although they demonstrate superior performance, their limited use may be due to their recent introduction to wireless communication signal processing and the associated computational complexity constraints. As it is clearly depicted, many works utilize multiple models to combine their individual benefits.

V. AUTOMATIC MODULATION CLASSIFICATION

AMC is a technique in wireless communications that enables automatic identification of a signal's modulation format without prior knowledge of the transmitted signal. Standard modulation formats, such as PSK, QAM, and OFDM, encode information onto carrier signals. AMC facilitates their recognition by leveraging signal processing methods, recently based on machine learning (ML), both supervised [77], [78], [79], [80], [81] and unsupervised [82], [83], [84]. Fast and accurate AMC on the receiver side could allow for adaptiveness on the transmitter side of the terminal, changing the modulation schemas based on the perceived channel conditions. This enables better spectrum efficiency and throughput, similar to traditional adaptive coding and modulation techniques.

A. OVERALL AMC COMPARATIVE OVERVIEW

In Table 2, we summarize the unsupervised AMC works. As can be seen from the second column and also highlighted with different colors, most of them rely on contrastive learning approaches (Section III-B), two rely on reconstruction

methods (Section III-C), while one employs deep clustering (Section III-A). This distribution shows the growing dominance of contrastive paradigms for feature extraction and representation learning in AMC tasks.

Representation learning module: As per the third column, it is evident that the CNN-based (see Section IV-B) backbones are the most frequently adopted, appearing in eight studies. These are often combined with additional components, such as attention mechanisms (see Section IV-D), which are apparent in four references, including one standalone. The LSTM and MLP (see Section IV-C and IV-A) modules are integrated less frequently and appear only once in combination with CNN. The fourth column of the table reveals that all works rely on I/Q data, meaning the different modulations are more distinguishable in the raw data than in the derivations, such as spectrograms or higher-order statistics. As a consequence, CNNs remain the main tool for capturing spatial correlations in I/Q signal representations, due to their capability of processing raw data as well as application flexibility (see Section IV-B). Recurrent and attention layers are typically used to enhance temporal or contextual feature modeling as part of the hybrid approaches.

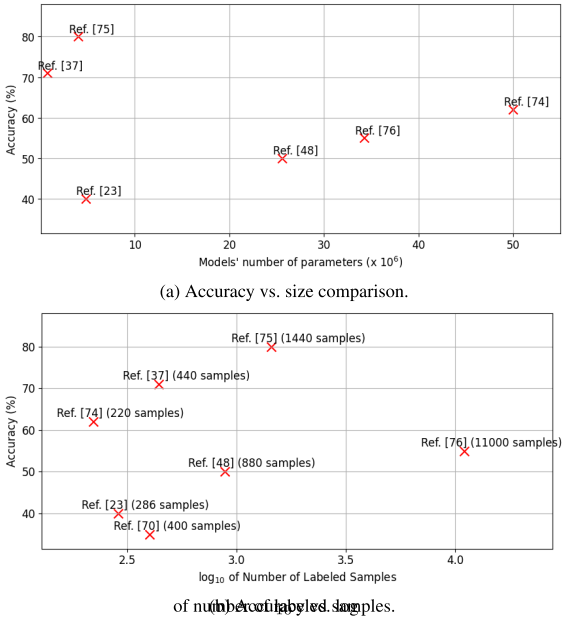
Experimental setup: The small-sample analysis column shows a consistent focus on label efficiency, with most approaches reporting results under extremely limited supervision, ranging from 0.5% to 10% labeled data or as few as 1–2 samples per class. Finally, the results and SNR/channel conditions columns reveal that evaluations are primarily conducted under Additive White Gaussian Noise (AWGN) conditions, with several studies also testing robustness under Rayleigh and Rician fading.

1) REPORTED PERFORMANCE ANALYSIS

Reported accuracies in AMC works vary from around 35% to 80% depending on the proportion of labeled data and the network configuration, illustrating the performance–supervision trade-off typical in semi- and unsupervised AMC. Although difficult to compare, due to the different evaluation setups, considering different numbers of classes, labeled samples, and channel conditions, in Figure 14 we show approximate plots of *accuracy-complexity* and *accuracy-number of labeled samples*.

Figure 14a, which shows the dependency of accuracy on model size, indicates no clear monotonic relationship between the number of parameters and classification performance. Notably, [75] achieves the highest accuracy (80%) with a relatively compact CNN+MLP model; however, it uses the largest number of labeled samples. The work in [37] with a small architecture attains an accuracy of around 70% using only 440 samples. In contrast, larger models such as [74] (Attention+CNN with ≈50M parameters) and [76] (Attention-based AE with ≈35M parameters) reach moderate accuracies near 55-60%.

Figure 14b, which shows the dependency of accuracy on the number of labeled samples given in logarithmic scale, highlights the performance sensitivity to labeling effort. Here, [75] again leads with ≈80% accuracy using



(a) Accuracy vs. size comparison.

of number of labeled samples.

FIGURE 14. Reported performance in the reviewed works considering models' sizes and the number of labeled samples.

only 1% of labeled data (1440 samples), underscoring the effectiveness of its contrastive formulation in leveraging unlabeled data. In contrast, [37] and [74] maintain reasonable accuracies (≈62–70%) with substantially fewer labels (220–440 samples), reflecting strong data efficiency under minimal supervision. At the opposite end, [76] achieves lower accuracy (≈55%) despite employing the largest number of labeled samples (≈11,000), which may be attributed to the reconstruction-based learning objective and usage of GAF-transformed data being less discriminative.

Overall, these results reinforce that contrastive learning architectures, particularly those integrating CNNs with auxiliary modules (e.g., MLPs or attention layers), tend to outperform reconstruction-based or clustering counterparts on the AMC task, even with fewer labels and smaller model sizes.

Furthermore, we compare the performance of the surveyed models under low SNR conditions, specifically at –6dB, which represents the most common low-SNR evaluation point among the reviewed works. This comparison aims to highlight the robustness of the models in challenging noise environments. A direct and comprehensive comparison across all studies is difficult due to varying validation conditions, including the use of custom datasets, different modulation sets, and varying proportions of labeled samples. Therefore, we limit our analysis to works that report results on the common RML2018.01A dataset [85], and we compare their classification accuracy, as illustrated in Figure 15.

It is important to note that the evaluation protocols differ even within this subset of studies. Some works [23], [37], [70], [76] employ fully supervised training using

TABLE 2. Summary of unsupervised deep learning approaches and achievements in AMC.

Ref.	Approach	Module	Input	Small Sample Analysis	Evaluation type	Num. class.	Results (Accuracy / Labels)	Data	Code	SNR [dB]	Channel
[36]	Deep clustering (AE)	CNN	I/Q	–	unsupervised	6	0.796 ACC, 0.88 NMI at 10dB	✓	–	−18 to 20	AWGN
[37]	Contrastive + Reconstr. + Classif.	CNN	I/Q	n=1, 2, 5, 10, 20, 30, 40, 50, 100	supervised	11, 20	70.54%, 55.84% / 2 labels per class (440, 800)	✓	–	−20 to 18	AWGN
[74]	Contrastive	Attention + CNN	I/Q	n=1, 2, 5, 10, 20, 30, 40, 50, 100	supervised	11	62% / 1 label per class per SNR (220 samples)	✓	–	−20 to 18	AWGN / Rayleigh / Rician
[48]	MoCo-v3 Contrastive	CNN	I/Q	n=0.5%, 1%, 5%, 10%	supervised	11	50.4% / 0.5% labeled (880 samples)	–	–	−6 to 20	AWGN / Rayleigh / Rician
[75]	SimCLR Contrastive	CNN + LSTM	I/Q	n=1, 2, 5(1%), 10, 20, 30, 40, 50(10%)	supervised	11	40% / 2 label per class (286 samples)	–	–	−6 to 20	AWGN / Rayleigh / Rician
[76]	Contrastive	CNN + MLP	I/Q	n=1%	supervised	9	80% / 1% (1440 samples) labeled data	✓	–	−20 to 18	AWGN
[70]	Contrastive	Attention + CNN	I/Q	n=1, 2, 5, 10, 20, 30, 40, 50, 100	supervised	11	35% / 2 per class per SNR (400)	✓	–	−20 to 18	AWGN
[77]	Reconstruction (AE)	Attention	GAF-I/Q	n = 5%, 10%, 20%	supervised	11	54.85% / 5% (11000) samples	✓	–	−20 to 18	AWGN

Legend: - Deep Clustering, - Contrastive Learning, - Reconstruction

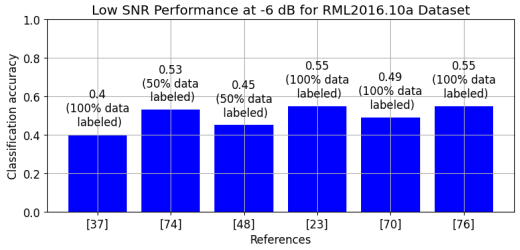


FIGURE 15. Reported accuracy of the surveyed models in low SNR conditions.

100% of the labeled data, whereas others [48], [74] use only 50% of the labels. Nevertheless, this comparison still provides meaningful insight into relatively low-SNR performance. Among the evaluated models, [74] demonstrates the best label-efficiency, achieving a classification accuracy of 0.53 fine-tuned on 50% labeled data. This strong performance is likely due to the model's large capacity, which allows it to effectively capture signal-specific features during pretraining. The top performing models at −6dB are reported in [23] and [76] while using 100% of the data labeled in the evaluation. Notably, all three of these leading models incorporate sequence-processing architectures, LSTM [23] and attention-based networks [74], [76], which suggests that such temporal or contextual modeling mechanisms exhibit superior robustness under low-SNR conditions.

B. OVERVIEW PER APPROACH TO AMC

1) DEEP CLUSTERING

In [36], the only study with a deep clustering approach in this section, the authors developed a custom clustering loss function for favoring better separation in the feature space provided by the encoding (middle) layer of the autoencoder. They combined it with the standard Mean Squared Error (MSE) reconstruction loss for training the models, thus enabling simultaneous feature learning and clusterability improvement in the feature space. The general constraint of such approaches is the combined loss function tuning. The separate components do not necessarily contribute to the general goal, considering that lowering the reconstruction loss does not always mean improving the clusterability. This could lead to complex shapes of clusters [17], which are impractical for post-processing and knowledge extraction.

2) CONTRASTIVE LEARNING

For the task of AMC, most of the recent works that rely on deep unsupervised learning [23], [37], [48], [70], [74] use contrastive learning architectures (see Section III-B), which are utilized in completely unsupervised setup or in semi-supervised [23] setup where the contrastive learning is only part of the pretraining phase. In [48], authors use the contrastive learning paradigm in one of its original setups, known as MoCo-v3 [86], using the ResNet50 CNN module (see Section IV-B), with special data augmentations relevant for the I/Q samples of radio data. They show that such a model is consistently better compared to other out-of-the-box contrastive learning frameworks, such as SimCLR [87].

In [37], the authors propose a weighted sum of three losses for the training architecture of three parts: reconstruction, contrastive, and classification, using weak and strong augmentations instead of the original signal and a single augmentation. They achieve 70.54% averaged accuracy over 11 classes and 55.84% averaged accuracy over 20 classes, trained with only 2 labelled samples per class, without

unlabeled samples, under AWGN noise in the range between −6 and 20dB. Further work could consider performance validation with unsupervised pre-training so that the full potential of such an approach is explored.

The authors in [23] introduce contrastive learning to AMC by utilizing SimCLR [87] to train a feature extractor (encoder) in the pre-training phase and subsequent fine-tuning on labeled data. They show that pretraining of the encoder leads to better performance on the downstream task compared to a supervised model using the same amount of data.

In [75], contrastive learning is part of a more complex architecture that enables the model to learn robust and generalized representations of radar signals during the semi-supervised pre-training stage. They explore the effect on the overall performance of a small subset of iterative labeled data injection in the fine-tuning procedure, and achieve better performance compared to the baseline MoCo [88] approach. A similar approach is considered in [70], where the learning module is realized with a transformer (see Section IV-D) instead of a CNN. They also provide extensive evaluation regarding the influence of each augmentation technique, the feature embedding power of the self-attention module, and the influence of the number of labeled samples in the fine-tuning stage.

Considering that data augmentations take a central role in the performance of contrastive learning, authors in [74] propose more domain-specific, semantics-preserving augmentations that mimic the environmental influence that affects the signal's shape, including Time Warping, Carrier Frequency Offset, and Phase Offset. Additionally, they use strong and weak augmentations of the input signal instead of the signal itself and one augmentation for the contrastive setup, marked as a' and a'' in Figure 4. For the learning task, they employ a Temporal Convolutional Network (TCN) [89] with masking for capturing the short-term features in the raw signal and a consequent transformer for capturing the long-term features. Together, they build the learning module of the contrastive learning setup. While the basic contrastive setup, as depicted in Figure 4, rates equally all of the negative samples, it introduces a weighting function in the loss calculation. This function takes into consideration the distances between the encodings of the samples by favouring closer negatives in the feature space, which leads to improved performance on a downstream classification task. They evaluate the proposed architecture against multiple approaches, including supervised CNN and LSTM (Sections IV-B and IV-C), and prove that it outperforms the competing approaches.

3) RECONSTRUCTION

Autoencoders (see Section III-C) are yet another unsupervised deep learning approach utilized for AMC. Compared to the Contrastive Learning works (see Section III-B), the AE seems less capable, but they have the advantages of being simpler and do not require data-specific augmentations. The different AE solutions vary regarding the learning module that is used for the encoder and decoder functionalities [40] and modifications in the loss functions. In [40], authors combine GRU, self-attention, and residual CNN blocks in the encoder and decoder modules to capture fine-grained (CNN) and time-dependent (GRU) features with a low-dimensionality representation (self-attention). The resulting architecture provides a higher score than the regular off-the-shelf CNN-based autoencoders, such as ResNet-based, on the task of low-signal detection with various modulations (RadioML data) across different low levels of signal-to-noise ratio (SNR).

Another AE-based, innovative approach is proposed in [76], where I and Q data streams are transformed to Gramian Angular Field (GAF) images, and an autoencoder is trained to reconstruct the fused image (combination of both I and Q components) based on the masked input image. The encoder and decoder are realized with a transformer architecture. While the baseline concept is the same, the usage of image-like data instead of raw I/Q time series makes the approach more flexible regarding the selected encoder-decoder module, considering there are many proven image processing architectures in open-source libraries such as [90].

C. DATASETS

All of the reviewed works consider the well-known RadioML dataset with its three variations RADIOML 2016.04C [85], RADIOML 2018.01A [91], and RADIOML 2016.10A, summarized in Table 3, as an evaluation baseline.

TABLE 3. Summary of RadioML datasets and their characteristics.

Ref.	Characteristics	Specific Features
[86]	RADIOML 2016.10A Synthetic I/Q signal samples across 11 modulation types (3 analog, 8 digital)	Includes labeled SNR increments, (-20dB to +18dB) with samples containing 128 I/Q data points and balanced classes.
[86]	RADIOML 2016.04C Synthetic I/Q signal samples across 11 modulation types (3 analogs, 8 digital)	Includes labeled SNR increments, (-20dB to +18dB) with samples containing 128 I/Q data points and unbalanced classes.
[92]	RADIOML 2018.01A Synthetic I/Q signal samples across 24 modulation types (analog and digital)	Contains 1,024 I/Q data points per sample across 26 SNR levels (-20dB to +30dB), including AWGN and Rayleigh fading.

Both RADIOML 2016.10A and RADIOML 2016.04C data consist of synthetic I/Q signal samples across 11 modulation types, including three analog and eight digital schemes in varying SNR conditions. RADIOML 2016.04C has additional variation among the samples with labeled SNR increments, allowing for a more detailed analysis of the benchmarked models.

The RADIOML 2018.01A dataset is the most comprehensive one, containing synthetic I/Q signal samples from 24 digital and analog modulation types. Each sample consists of 1,024 I/Q data points across 26 SNR levels ranging from −20dB to +30dB, simulating realistic wireless channel conditions with additive white Gaussian noise (AWGN) and Rayleigh fading.

All of the related papers perform evaluations based on these three datasets. While these three datasets are a good base for direct model-to-model comparisons of the different approaches for AMC, real-world evaluation is lacking. The creation of real-world measurement datasets could provide great insights about the actual deployment capabilities of the developed models. As such, they could contribute to building models for actual existing challenges instead of models for the available data.

D. FINDINGS

Based on the reviewed literature in this domain (Table 2), contrastive learning appears as the most promising approach, given that more than half of the recent research works propose it as the superior solution for AMC. Thus, future efforts should explore more domain-specific data augmentations, such as those mimicking environmental influences, to improve model robustness in diverse operating scenarios. The combination of multiple methods in sequential or parallel learning phases may help train more capable models regarding their generalization to unseen environmental conditions.

1) EVALUATION PRACTICES

Nearly all surveyed studies rely on simulated datasets, with the RadioML corpus and its variants being the most frequently used benchmarks. While these datasets facilitate controlled experimentation, they often fail to capture the full complexity of real-world environments. Thus, an important direction for future work is the development and open dissemination of real-world RF benchmark datasets. These would reveal challenges beyond the signal-to-noise ratio (SNR) variability, such as channel fading models and environmental dependencies, which could be obtained by spatial Radio Environment Maps (REMs) [92]. In such a way they would support validation of the proposed models in operational conditions that are more representative of the actual deployment.

Moreover, there is a pressing need for a standardized evaluation framework, tailored to domain-specific characteristics such as viable I/Q sample length for actual deployment, considering the time delays of the decision-making and the system itself, SNR levels, labeling ratios, and augmentation techniques. Although many studies assess model performance under limited labeled data conditions, there is significant inconsistency in the selection of sample percentages or absolute quantities, making direct comparisons between works problematic. Currently, comparisons often center around performance with only one or two labeled samples, but the diversity in evaluation setups obscures the relative progress across methods. A consistent benchmarking methodology, such as enforcing a fixed number of labeled samples, using standardized activation functions in the final fully connected layers, or defining common augmentation pipelines, would greatly enhance transparency, reproducibility, and comparability across studies.

2) ROAD AHEAD

In standardized wireless systems where a fixed set of modulation schemes is used, supervised learning models have already demonstrated high classification accuracy [80]. However, recent advances suggest a shift towards machine-learned physical layers, where deep learning models replace traditional components of the communication chain [93]. These approaches enable the design of custom modulations, dynamically adapted to specific channel conditions. Such a paradigm shift will demand self-adaptive, label-free learning models capable of generalizing to new, previously unseen signal formats and operating conditions. This challenge points to unsupervised deep learning (see Section III) as a key enabler, allowing models to extract general signal features from large-scale unlabelled data and to adaptively refine their internal representations based on deployment-specific observations.

VI. SIGNAL SENSING

While AMC, discussed in the previous section, focuses on the categorization of signal-related features for a given signal, this section addresses signal sensing, which involves detecting, analyzing, and interpreting signals within a specific radio spectrum or environment to determine their presence, characteristics, and properties. Signal sensing provides general awareness in the time and frequency domains regarding the operational patterns of various technologies and devices. It processes sensed data that may or may not contain signals, making the problem more general and closely tied to the physical layer of communication systems. The main purpose of such awareness is the smart utilization of the radio spectrum as a scarce resource by multiple devices that rely on different RATs, thus avoiding interference. This is especially important in the unlicensed radio spectrum bands such as the Industry, Scientific and Medical (ISM) [94] and 5.9 GHz Intelligent Transportation Systems (ITS) [95], [96] in which multiple technologies share similar frequency bands.

Signal sensing in the following is focused on the general monitoring of radio spectrum occupancy only by analyzing the patterns of spectral activity in the concerned frequency bands. This involves extracting knowledge about activity patterns in frequency and time domains with the main goal of using the free spectrum and time resources in a dynamic manner based on historical knowledge, complementary to other sensing techniques, such as AMC and SEI.

A. TYPES OF SIGNAL SENSING

For clarity and consistency, we group the reviewed works into three complementary research directions aimed at understanding and managing spectrum usage, namely Primary User (PU) detection, Radio Access Technology (RAT) recognition, and Jamming Signal Recognition (JSR). Each category addresses a specific aspect of spectrum awareness. More specifically, they identify active licensed transmissions, characterize the operating technology and detect intentional or unintentional interference.

1) PRIMARY USER DETECTION (PUD)

PUD focuses on determining whether a licensed or prioritized user is transmitting within a given frequency band. Although related to Specific Emitter Identification (SEI), PUD operates at a broader level, as it does not aim to identify individual transmitters but rather to detect the presence of active users or user groups. While direct processing of raw in-phase and quadrature (I/Q) signal samples could achieve this task, such methods raise security and privacy concerns with respect to potentially revealing user-specific information. As a result, existing approaches often rely on short I/Q sequences, typically from the signal preamble [96], to infer PU activity without accessing sensitive data, or some derived data representation such as spectrograms [54] or energy levels [66].

2) RADIO ACCESS TECHNOLOGY (RAT) RECOGNITION

RAT recognition represents a key component of signal sensing, aimed at identifying the type of radio access technology active in a monitored frequency band. This task involves analyzing spectral activity patterns across time and frequency to distinguish among different communication standards. The extracted knowledge could enable information for dynamic spectrum access by predicting and exploiting unused spectrum resources based on historical activity. RAT recognition, therefore, complements techniques such as AMC and SEI, providing a higher-level insight into spectrum utilization and technology coexistence.

3) JAMMING SIGNAL RECOGNITION (JSR)

We also categorize the *jamming signal detection* as part of the general signal recognition since it requires distinguishing between the wanted and unwanted signals operating in the same portion of the radio spectrum. For this challenge, the unsupervised deep learning models are particularly suitable for two reasons. Firstly, there is usually a lot of sensed data that is difficult to label since the jamming signals are typically unknown. Secondly, the radio spectrum environment may be very dynamic, where there could be many unknown sources of radiation in the monitored band, especially in the unlicensed bands.

B. OVERALL SS COMPARATIVE OVERVIEW

In Table 4, we summarize the works employing unsupervised deep learning for signal sensing. Three types of methods are equally used: deep clustering (Section III-A), reconstruction (Section III-C), and generative (Section III-D), as can be seen from the second column of the table and also highlighted with different colors. As can be seen from the third column, seven architectures contain CNN modules (see Section IV-B) at least as part of the complete solution, four contain MLPs (see Section IV-A), while LSTM, MLP, and GRU (see Section IV-C) appear twice each.

Regarding the data type, I/Q samples are most frequently utilized, in six of the surveyed works, which is expected considering that it is the highest density data and thus provides the richest features. However, for the signal sensing challenges, low data intensity solutions also exist, such as energy levels of the monitored channel for PU detection [43], [66] and FFT-derived data such as spectrograms for RAT classification [5], [17].

PU detection is the most frequently explored task according to the reviewed works. Most approaches rely on AE or VAE architectures applied to I/Q data [39], [40], [43]; however, simpler energy detection solutions based on MLP are also viable [66]. AE and VAE are the most common approaches for the PU task, which means that reconstruction and generative models provide robust detection capabilities under low supervision and variable noise conditions. Interestingly, the PU detection is the only task that considers the channel model.

RAT recognition tasks are addressed in [5], [17] and [51], mainly through deep clustering and AE-based approaches. This confirms that clustering-based feature learning is well-suited for RAT differentiation, as the RAT-specific patterns are distinct and sufficiently captured through frequency-domain representations.

Jamming recognition appears in two of the reviewed works [67], [72], both relying on AE architectures and considering the Jamming to Signal ratio in the experimentation. This shows that AEs, although being simple models, are well-suited for the task of JSR.

In terms of evaluation and reproducibility, most studies conduct unsupervised validation, with a few complementing it by supervised fine-tuning for benchmarking. Reported class numbers range from two to eight, depending on task specifics, which makes potential direct comparison of the methods difficult.

1) REPORTED PERFORMANCE ANALYSIS

Figure 16 compares the model sizes across the reviewed signal sensing works, categorized by task type. The results show that RAT (in blue) and jamming (in red) recognition models tend to be the largest, with parameters ranging from approximately 0.13 M to 11.2 M, reflecting the need for richer spectral feature extraction in multi-signal environments. In contrast, PUD models (in green) are considerably smaller, mostly below 1 K parameters, except for the cGAN-based approach [54] which reaches 3 M parameters due to its generative nature. This trend indicates that PU detection tasks can be effectively addressed with compact architectures, while more complex sensing problems, such as RAT classification, benefit from deeper or wider networks.

C. OVERVIEW PER APPROACH TO SS

1) DEEP CLUSTERING

a: RAT

The performance of deep clustering (see Section III-A) with different 2D CNNs and transformer deep learning modules is explored in [5] for wireless technology recognition in the license-free 868 MHz frequency band by using spectrogram

TABLE 4. Summary of works concerning unsupervised deep learning approaches and achievements in signal sensing.

Ref.	Approach	Module	Input	Problem type	Evaluation type	Num. class.	Performance	Data	Code	Noise conditions	Channel
[5]	DeepCluster	CNN	Spectrogram	RAT clustering	unsupervised supervised	unkn.	0.6 Silhouette score / 0.84 F1 score	✓	✓	–	Real-world
[17]	DeepCluster	CNN	FFT Amp.	RAT clustering	unsupervised supervised	unkn. /3 /6	0.6 Silhouette score / 0.99 F1 score avg.	✓	✓	–	Real-world
[72]	Autoencoder + Classifier	CNN	I/Q	Jamming recognition	unsupervised	7	>0.95 F1 score	–	–	JSR (-30 to 20)dB	–
[67]	Autoencoder	CNN + MLP	I/Q	Jamming recognition	unsupervised	8	0.95 Accuracy	–	–	JSR (10 to 15)dB	–
[51]	Autoencoder	CNN, MLP, LSTM	I/Q	RAT classification	supervised	2	0.93 F1 score	–	–	–	–
[43]	Sparse Autoencoder (SAE)	MLP	I/Q	PU detection	unsupervised	4	Detection prob. 0.8 at 1W and 0.1 False alarm probability	–	–	Emiting power (0.1 to 1)W	AWGN
[40]	Autoencoder	GRU+ CNN+ Atten.	I/Q	PU detection	unsupervised	8	Detection prob. 0.85	✓	–	(-18 to -5)dB	Rayleigh
[66]	VAE	MLP	Energy Levels	PU detection	unsupervised	2	Clustering accuracy 86.02%	–	–	AWGN, Laplacian (-15 to 5)dB	Rayleigh fading
[39]	VAE	CNN + MLP + GRU	I/Q	PU detection	unsupervised	2	Detection probability 71% at 0.1 False alarm prob.	✓	–	AWGN (-18 to -5)dB	Rician model
[54]	cGAN	CNN	Spectrogram	PU localization	supervised	2	Error (33 to 15)m	–	–	AWGN (10 to 50)dB	–

Legend: - Deep Clustering, - Contrastive Learning, - Reconstruction

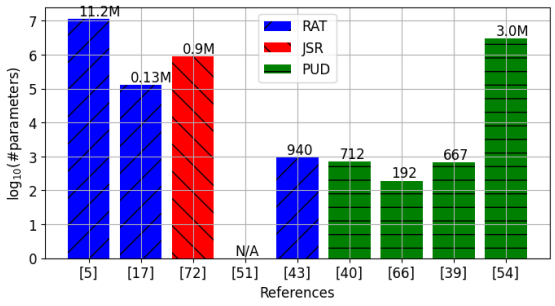


FIGURE 16. Size comparison of the models in signal sensing. Blue - RAT recognition, Red - JSR, Green - PUD.

data. Complexity-performance tradeoffs are analyzed, with a conclusion that the utilization of large CNN and Vision Transformer (ViT) modules for feature learning brings only minor performance improvement while significantly increasing the model size, thus imposing constraints on its application potential. The reason is the low-content pictorial data in the spectrogram images, which is captured by lower complexity models such as ResNet11 with comparable performance to the significantly larger models, such as VGG16 and ViT.

b: RAT

Another work for RAT categorization based on deep clustering is published in [17], where the target is even further lowering the model size. 1D FFT amplitude data is considered instead of spectrograms, thus allowing for significantly shorter end-to-end inference time of the 1D CNN-based models, such as depicted in Figure 10. The increased recognition performance was validated on three different datasets, each targeting a specific deployment challenge. These included the environment influence, the increased number of operating RATs, and the spectrum monitoring data of continuous operation with many unknown signals and noise. The proposed model was shown to be robust and easy to set up compared to the then-state-of-the-art AE-based models. Although the labeled validation datasets are from real-world transmissions, the very high performance (up to 0.99 F1 score) could be a result of using very clean data. The evaluation with the unlabeled, live stream spectrum data could be considered the most informative for the actual performance, considering the three separate evaluations.

c: JSR

In [72], AE architecture (see Section III-C) and classifier networks are jointly used to learn feature extraction from known (labeled) samples of jamming signal during training. In the deployment stage, only the feature extractor is used, combined with two-stage distance-based classification. Firstly, the binary classification of known and unknown signals is performed, and a more granular classification of different patterns is performed afterwards. A joint objective function containing center loss, cross-entropy,

and reconstruction loss is employed with corresponding weighting coefficients. The distance-based classification is actually clustering signals based on Euclidean distance while also taking into consideration the known classes. The main advantage compared to the original implementation is obtained when classifying three unknown classes of jamming signals. However, the reported results appear pretty high with the accuracy of 100% for three known classes and one unknown class.

d: JSR

Similar to the approach in [72], and on a similar problem of unknown Jamming signal recognition, authors of [67] propose an AE-based architecture that can extract semantic attribute features. The only difference is in the definition of the clustering loss. They compare their approach to a significantly simpler (three-layer) MLP trained in a supervised setup and show the superior performance of the proposed method. Additionally, they also provide an ablation analysis of each of the losses contributing to the performance of the method.

2) RECONSTRUCTION

a: RAT

In [51], the authors used deep AEs (see Section III-C) on I/Q data, which contains samples from different modulations, in an attempt to address the problem of RAT classification between the LTE and a combination of WiFi signals (IEEE 802.11ax and IEEE 802.11ac). Interestingly, they show that the deep AE with CNN layers outperforms the LSTM-based and VAE architectures, although LSTM was meant for longer-term patterns apparent in time-series data such as I/Q sequence.

b: PUD

The PU detection task is addressed in [43] with Sparse AE (SAE) combined with Gaussian Mixture Model (GMM). Energy levels sensed by the SUs are used as input data. A cooperative setup is considered where the secondary users (SU) send their sensing data to the central coordination node, where PUs and SUs are spatially distributed in the operating environment. In the evaluation setup, it is considered that SUs can appear and disappear due to their contribution to the signal measurements of the central node. For the considered scenario, the authors prove that their proposed SAE-GMM model achieves performance comparable to a supervised neural network on the task of binary classification of active and inactive PUs in the monitored channel. The solution is realized by introducing sparsity in the embedding layer, by constraining the average activation rate of the neurons of the embedding layer of the AE architecture. The idea is that forcing only a small number of active neurons during training will lead to learning compressed and more meaningful representations. The sparsity is controlled by incorporating the Kullback-Leibler divergence between the required average activations and the actual average activations as part of the loss function, which originally contains only MSE loss. It is important to highlight that the claimed performance is achieved using a very simple, fully connected neural network with only five layers.

3) GENERATIVE

a: PUD

In [66], the authors propose using VAE (see Section III-D), a representative of generative methods, combined with GMM for distinguishing PU signals given unlabeled data of PU transmission signals and free channel signals. VAE is used for feature extraction, and GMM for clustering in the latent space. While the training is performed completely unsupervised, a small labeled subset of data is required for threshold adjustment and cluster identification. With this approach, they achieve comparable results to the supervised CNN counterpart, which requires significantly more labeled data for training.

b: PUD

VAE is also used for unsupervised learning of features in [39] for the PU detection in various single and multiple PU setups for the Non-Orthogonal Multiple Access (NOMA) technique. The feature learning with VAE is performed on denoised radio data with so-called recorrupted-to-recorrupted [97] denoising based on a proposed GRU-based (see Section IV-C) AE architecture for radio signals. The output classification is performed with the K-means++ algorithm. It is important to note that in this work, the authors propose a shallow encoder and corresponding decoder for the AE and VAE structures adapted to the characteristics of the radio signals, such as strong, stationary periodic properties and high frequency. Such architecture is shown to outperform the referenced approaches, which also contain a supervised CNN model, proving the significant contribution of the denoising function by the proposed GRU-based model.

c: PUD

In [54], the authors propose a two-stage solution based on GAN and U-shaped NN (U-Net) for privacy-aware localization of spectrum violators in the United States' Citizen Broadband Radio Service (CBRS) [98]. In their work, GAN architecture (Section III-D) is used for masking PUs in spectrograms acquired with multiple sensors, and U-NET [99] is used in an image-to-image translation setup based on labels. Thus, instead of the standard signal recognition setup where a given signal is classified, GAN is used for masking the privacy-sensitive PU transmissions in spectrograms, which implicitly requires learning their appearance in spectrograms. While the original setup of GAN is unsupervised, the way it is used in this work requires some sort of labeled data, consisting of spectrograms with and without PU transmissions.

D. DATASETS

In Table 5, we summarize some of the open, real-world datasets that could be used for future research in signal sensing challenges.

TABLE 5. Summary of real-world datasets for wireless signal recognition.

Ref.	Characteristics	Specific Features
[65]	Labeled dataset consisting of transmissions from LTE, WiFi, and DVB-T technologies operating in the 2.4 GHz band.	Data collected from 7 locations (6 in Ghent, Belgium, and 1 in Dublin, Ireland), enabling analysis of environmental effects on signal recognition.
[97]	Labeled dataset with signals from LTE, 5G, WiFi, ITS-G5, and C-V2X technologies operating in the ITS 5.9 GHz band.	Suitable for exploring model performance with coexisting RATs, including technologies with different or similar spectral shapes.
[19]	Partially labeled dataset from the LOG-a-TEC testbed in Ljubljana, Slovenia, covering the 868 MHz SRD band.	Contains spectrum traces with at least four technologies (LoRa, IEEE 802.15.4, SigFox, proprietary technologies). Evaluates model performance in scenarios with high variability and interference.

A labeled real-world dataset from [65] consists of transmissions of three RATs, namely LTE, WiFi, and DVB-T, which operate in the 2.4 GHz band. The complete data was collected from 7 different locations, including 6 in the city of Ghent, Belgium, and one in Dublin, Ireland. This dataset allows for a detailed analysis of the environmental effects on signal recognition models' performance.

Another dataset [96] of real-world measurements was collected in Antwerp, Belgium, and includes signals from five RATs: LTE, 5G, WiFi, ITS-G5, and C-V2X, all operating in the ITS 5.9 GHz band. This dataset provides an insight into scenarios with more coexisting RATs compared to the previous one, including technologies with significantly different or similar spectral shapes. This makes it suitable for exploring how the coexistence of more technologies affects the models' performance.

Yet another partially labeled dataset was collected from the LOG-a-TEC testbed in Ljubljana, Slovenia [19]. It comprises spectrum traces in the 868 MHz Short-Range Device (SRD) band with a bandwidth of 192 kHz and a sampling frequency of 5 Power Spectrum Density measurements per second, using 1,024 FFT bins. At least four technologies appear in this data: LoRa, IEEE 802.15.4, SigFox, and proprietary technologies. This dataset is suitable for evaluating model performance in real-world scenarios with significant variability, artifacts, and interference.

E. FINDINGS

1) PUD

PU detection is commonly formulated as a binary classification problem, focused on identifying the presence of a signal of the primary users. This formulation lends itself to relatively simple evaluation using standard metrics such as accuracy, Receiver Operating Characteristic / Area Under the Curve (ROC-AUC), and F1-score, which facilitates straightforward benchmarking across studies. While such simple setups allow for analysis on the model performance, considering also SNR and channel state, they are overly simplistic and do not reflect actual deployment conditions where many more users are present in the environment. Thus, following the examples of [40] and [43], multiple-user scenarios should be considered in future research.

2) JSR

Similarly to the PUD task, JSR is also addressed as a binary detection problem. Therefore, moving beyond simple presence detection to the classification of jamming types as in [67] is an important research direction. Such classification would support the development of targeted mitigation strategies, making this task increasingly relevant for operational deployments. Furthermore, setting the jamming recognition as an Open-Set problem, where the appearance of unseen jamming patterns is expected, should be a favoured approach in future works, following the examples in [67], [72]. Thus, further exploration of unsupervised deep learning approaches is highly relevant in this direction of work.

3) RAT

From a signal recognition perspective, RAT recognition parallels the challenges of JSR, where systems must identify and group multiple signal types without labeled data [5], [17], while favouring even faster decision making, considering potential applications in near-real-time sensing and network parameters' control. RAT recognition requires the identification of multiple signal classes or clusters without or with a very small set of labeled samples. This makes unsupervised evaluation often a necessity in estimating the performance, considering clustering metrics such as Adjusted Rand Index (ARI) and Normalized Mutual Information (NMI). When clean datasets are used [96], and the signals exhibit distinct frequency-domain features (e.g., unique FFT shapes), unsupervised learning can show extremely good performance [17], [40], which could be misleading. Therefore, future works in this direction should consider both supervised and unsupervised evaluations, and multiple metrics due to the uncertainty in the latter.

4) ROAD AHEAD

While supervised evaluation is well-established, the unsupervised, which is highly relevant in this direction of work, requires a more unified and even standard approach that uses a well-defined set of metrics, which could be beneficial for future benchmarking and development. Regarding the data, even when generated with over-the-air (OTA) capturing methods in a controlled manner, some datasets [65], [96] fail to fully reflect the dynamics, variability, and noise present in real-world RF environments. This results in significant performance differences compared to live continuously sensed data [17]. We recommend that in future research, both types of data, from controlled and uncontrolled environments, should be used for providing better performance estimation,

especially for applications intended for deployment in live, dynamic wireless networks [19].

From a deployment perspective, signal sensing must be extremely fast, often within a few milliseconds, to meet the real-time requirements of modern wireless systems [96]. This constraint encourages the placement of sensing algorithms close to the radio unit, typically at the edge or within distributed units (DUs) in O-RAN architectures, where both latency and computational resources are limited. These limitations promote the use of lightweight models that minimize data movement and processing overhead. As a result, averaged FFT representations of the signal are often preferred over raw I/Q samples, leading to computational reductions by up to three orders of magnitude [2].

The rise of distributed applications (dApps) in edge computing and their increasing integration into O-RAN as standard components [100] will further support the development of near-real-time, low-level machine learning models for signal sensing.

VII. SPECIFIC EMITTER IDENTIFICATION

SEI relies on recognizing device-specific signal alterations caused by hardware imperfections [104], [105]. Deep learning techniques, especially supervised models, have shown exceptional performance [106] in distinguishing known devices, even in networks with thousands of transmitters [107], [108], [109]. However, real-world scenarios often lack labeled data and may also contain transmissions of unknown transmitters, which could pose security concerns. Thus, significant attention is paid to the development of models that can learn from data from known transmitters and distinguish potential unknown transmitters. The challenge could be addressed by using CNN-based feature extractors [84], [105], which are trained on known devices and aim to group unseen emitters in the latent space via clustering algorithms. So far, existing solutions approach the training of the feature extractor through a classification task [53], [110]. Unsupervised deep learning fits as a promising approach for this challenge [110], considering the ability to learn from large amounts of unlabeled data.

A. OVERALL SEI COMPARATIVE OVERVIEW

In Table 6, we summarize the works concerning SEI. Similar to AMC (Section V), most of them rely on contrastive learning approaches (Section III-B), two rely on deep clustering methods (Section III-A) while one employes reconstruction (Section III-C) and one predictive (Section III-E) approach, as can be seen from the second column of the table and also highlighted with different colors. As per the third column, all architectures contain CNNs (see Section IV-B) at least as part of the solution, and three also include MLPs (see Section IV-A). The fourth column of the table reveals that all works rely on I/Q data, which is expected since the goal is to capture the hardware-induced I/Q irregularities in the radio signal, specific to each device.

With regards to the *Problem type* column, SEI evaluations are performed in a zero-shot and few-shot settings, which directly corresponds to the unsupervised and supervised *Evaluation type*, accordingly. Small-sample evaluations are of great interest for the SEI task due to the similarity to the actual deployment conditions. The *Performance* column demonstrates consistently strong results, with detection and clustering accuracies typically above 90%. This confirms that unsupervised representations can effectively capture hardware-specific signal alterations. However, each publication reports evaluation with a different dataset and with different numbers of samples in the few-shot setup, which constrains direct comparison. According to the *Data* and *Code* columns, there is limited public availability of datasets and implementations, with only a few works providing accessible repositories. Regarding the *SNR* and *Channel* columns, only three of the studies consider the noise effect. Finally, regarding the channel model, only one study evaluates with antenna replacement, which means currently most of the focus is on the model's design and simplistic validations with varying numbers of seen and unseen devices.

1) REPORTED PERFORMANCE ANALYSIS

For the surveyed SEI works it is impossible to extract the performance comparison with respect to model size due to the large variety of evaluation conditions, such as different dataset classes, number of samples per class, and different technologies. Therefore, Figure 17 only compares estimated model sizes used across the SEI studies for zero-shot (blue) and few-shot (red) evaluation setups. Most of the proposed models in both setups have comparable sizes of up to 1 M parameters, with the exception of [42] and [52], reaching 11 M and 4.53 M parameters, respectively. Overall, the figure shows that SEI models achieve strong generalization with relatively lightweight architectures, and that model size scales primarily with the degree of supervision and the complexity of the experimental setup.

B. OVERVIEW PER APPROACH TO SEI

1) DEEP CLUSTERING

The study in [41] focuses on novel device discovery with deep clustering (see Section III-A) by extracting features

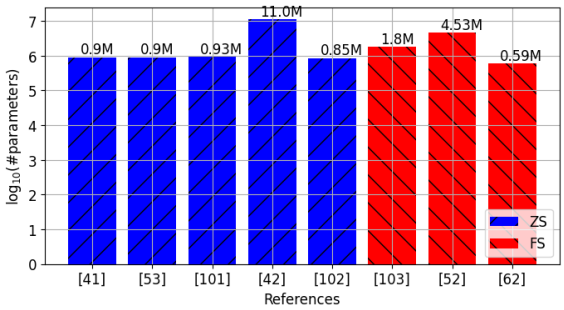


FIGURE 17. Estimated size comparison of the models in SEI. Blue - Zero-shot (ZS) evaluation, Red - Few-shot (FS) evaluation.

TABLE 6. Summary of unsupervised deep learning approaches and achievements for SEI.

Ref.	Approach	Modul	Input	Prob. Type	Evaluation Type	Performance	Data	Code	Small sample analysis	SNR (dB)	Channel
[41]	Deep Clustering	CNN+ MLP	I/Q	Zero-shot	Unsupervised	Clustering Purity WiFi: 0.564 ADS-B: 0.899	–	–	–	–	–
[53]	Deep Clustering	CNN	I/Q	Zero-shot	Unsupervised	Single known emitter detection Acc. 89.1%	–	–	n = 0, 50, 100, 200, 400, 800	–	–
[103]	Contrastive Learning	CNN+ MLP	I/Q	Zero-shot	Unsupervised	15 known 1 unknown, detection Acc. 91.79%	✓	–	0–30%	–	Antenna replacement
[42]	Contrastive Clustering	CNN+ MLP	I/Q	Zero-shot	Unsupervised	6 known 1 unknown, detection Acc. 91.21%	–	–	n = 1, 5, 10, 20, 30, 50	−10 to 20 by 10 incr.	AWGN
[104]	Meta-Contrastive	CNN	I/Q	Zero-shot	Unsupervised	30 known vs 15 unknown Accuracy: 90.5%	✓	✓	–	–	–
[105]	Contrastive	CNN	I/Q	Few-shot	Supervised	WiFi: 10 vs 6, ADS-B: 90 vs 10, 10-shot Acc. 86.33%, 62.73%	✓	✓	5 to 30 by 5 incr.	–	–
[52]	MAE (Reconstruction)	CNN	I/Q	Few-shot	Supervised	8 emitters, 10-shot Acc.: 41%	–	–	20 to 100 by 20 incr.	0 to 10 by 1 incr.	Real in-lab data.
[62]	Phase Prediction+ Decoupling	CNN	I/Q	Few-shot	Supervised	10 vs 6, Accuracy: ∼62%	✓	✓	5 to 30 by 5 incr.	0 to 30 by 5 incr.	–

Legend: - Deep Clustering, - Contrastive Learning, - Reconstruction, - Predictive

from known devices and clustering these features. The proposed architecture outperforms the baseline in metrics such as clustering purity, completeness, and isolation. For comparison, the baseline involves direct clustering of I/Q samples, with both approaches using DBSCAN as the clustering algorithm. The latent representations are taken from an intermediate (late but not final) network layer to avoid class-specific information. The clustering is performed on 2D data obtained via Uniform Manifold Approximation and Projection (UMAP) from 100-dimensional vectors provided by the MLP output layer.

Zero-shot learning is also performed in [53], where the proposed approach assumes adding clustering loss on the feature vectors extracted by the same architecture as used in supervised classification [111], with an additionally defined feature representation layer, thus creating a deep clustering method (Section III-A). Clustering loss is a complex function consisting of three separate parts, enforcing the separability of the clusters during training. Known and unknown transmitters are apparent in the training set, helping the model to create clear cluster boundaries. Single, known device detection is performed with 89.1% accuracy.

2) CONTRASTIVE

Contrastive learning is a recently introduced method in the SEI domain. Although some of the approaches require labels [102], [112] for the pre-training, they are suitable for few-shot learning, which is the most often used approach for evaluation of the unsupervised approaches [52], [62], [103]. Thus, we consider these methods as borderline cases and still include them in this review, as they rely on contrastive loss and meta-learning strategies, i.e., concepts which are applicable in unsupervised learning, and require a small amount of labeled data [102].

In [101], the authors propose a lightweight CNN-based architecture for contrastive learning (see Section III-B). They validate it in several scenarios on the open dataset with 20 transmitters, including the open-set recognition of new unknown transmitters [112]. Their method shows stable performance with the accuracy above 70%, considering various numbers of unknown transmitters (up to 5). It is important to note that, to the best of our knowledge, this is the only work investigating how the disappearance of known transmitters affects the performance. One significant constraint regarding clustering is the necessity of labeled datasets for the initial phase of training the feature extractor.

Contrastive clustering adaptation to I/Q signals for the SEI challenge is proposed in [42]. They use a CNN-based (Section IV-B) feature extractor and evaluate it with K-means on the extracted feature space. Such a model outperforms the baseline clustering methods, existing deep clustering, and generative approaches in novel device detection tasks. However, there is a constraining requirement of knowing the number of classes (transmitters) in the training phase.

Meta-contrastive learning method for SEI is proposed in [102], where the adaptation to the new unseen devices is approached as general domain adaptation, designing a dedicated loss function that enforces both device-specific features development and also general features. While the

model shows remarkable performance, it is worth noting that labeled data is necessary for the training phase. This is an important direction of work, considering that there are many image-processing and time-series [26] processing frameworks that could be utilized with such domain adaptation.

Contrastive learning as a pretraining step of a deep CNN is also performed in [103], where, besides the common data augmentations such as rotation and flipping of samples, additional model-driven feature-space augmentation is proposed. Custom augmentation works by introducing offsets in the feature space and mapping the perturbed feature back to the original sample. Such an innovative approach allows for custom augmentation, which introduces maximal changes to the samples while still maintaining a stable feature space. It outperforms the standard contrastive learning methods with only standard augmentations by a significant margin in scenarios with a very small amount of labeled data.

3) RECONSTRUCTION

Masked AE (MAE) (see Section III-C) architecture for SEI is proposed in [52] for unsupervised pre-training of a feature encoder as part of the development of an SEI classifier in a scenario with a limited amount of training samples. The authors evaluate the approach using multiple masking techniques. They prove that the introduction of MAE leads to consistently better classification performance compared to the purely supervised models with different numbers of training samples, evaluating both simulated and real data.

4) PREDICTIVE

In [62], the authors propose a three-part pretext task for training an encoder on unlabeled data. Each of the three parts, I/Q sample rotation classification (see Section III-E), reconstructing the original sample from the rotated sample, and contrastive learning for distinguishing noise from real samples, has its own loss calculation, together building the total loss. The final model is capable of achieving significant classification accuracy above 90% on downstream tasks using only 30 labeled samples, significantly outperforming the purely supervised model on the same number of labeled samples. However, this work does not provide performance comparisons with other existing approaches, such as the pure contrastive ones proposed in [42] and [102].

C. DATASETS

In this section, we report on some of the most relevant datasets for exploring and addressing the SEI challenge, also summarized in Table 7. While there are many datasets,[1][2] available for this task, we will consider the most comprehensive ones.

In [112], the authors publish a large 7TB real-world dataset containing signals of 20 devices with similar hardware communicating over Wi-Fi. The data are collected in three setups, namely, in an anechoic chamber, in the wild, and using a cable for transmission between the transmitter and receiver. This is a very versatile dataset that could be used to explore how environmental conditions affect SEI performance, such as the multi-path and the channel effect. Considering the amount of data, it makes a valuable starting point for the development of large unsupervised learning models.

TABLE 7. Summary of datasets for SEI challenges.

Ref.	Characteristics	Specific Features
[114]	Real-world dataset of signals from 20 devices communicating over Wi-Fi. Data was collected in three setups: anechoic chamber, in the wild, and cable transmissions.	A versatile 7TB dataset that explores environmental effects on SEI performance, such as multi-path and channel effects. Suitable for developing large unsupervised models.
[22]	ADS-B airplane signals operating at 1090 MHz. Includes short and long signals with 1713 and 1670 device categories.	Automatic data collection procedure. Evaluation of deep learning models on the dataset could guide feature extraction module selection for unsupervised approaches.
[115]	RF fingerprinting dataset with transmissions of 174 Wi-Fi transmitters and 41 USRP receivers in the 2.4 GHz band. Captures made over a month.	Collected in a closed space using the ORBIT[1] testbed. Includes small subsets for fast experiments tackling challenges like channel variations and the effect of multiple receivers.

Over-the-air captured data of airplanes, using the ADS-B system operating at 1,090 MHz, is described in [22]. The authors propose an automatic data collection procedure and demonstrate it by creating two subsets of data, with short and long signals, with 1,713 and 1,670 categories/devices. They also evaluate various deep learning models on the data. Although they consider only supervised approaches, such an evaluation could be useful in the selection of the deep learning module, which most unsupervised approaches also contain for the feature extraction process.

Another open and large RF fingerprinting dataset is reported in [113], containing transmissions of 174 Wi-Fi transmitters and 41 USRP receivers, with the captures made over a month. The transmissions are made in the 2.4 GHz band. The data is captured in a closed space and controlled environment using the ORBIT[4] testbed. The authors also provide small subsets of the data that could be used for fast experiments, tackling different challenges, such as channel variation influence over time and the effect of utilization of multiple receivers.

D. FINDINGS

Unsupervised deep learning shows significant potential for tackling the SEI challenge, especially given the promising early results achieved with architectures adapted from other domains such as contrastive learning [101] and predictive learning [62]. These approaches demonstrate the capacity to extract meaningful representations from unlabeled RF data and distinguish among signal sources with minimal supervision.

[1] https://genesys-lab.org/mldatasets
[2] https://ece.northeastern.edu/wineslab/datasets.php
[4] https://orbit-lab.org/

1) EVALUATION PRACTICES

In general, two dominant evaluation strategies have emerged in the literature. The first corresponds to zero-shot setups, where models are tested on devices unseen during training, with or without the use of labels in evaluation. The second covers few-shot learning approaches, which involve fine-tuning with a small number of labeled samples (e.g., 5, 10, or 30) before evaluating classification accuracy. While both strategies are useful for assessing generalization and sample efficiency, there is substantial inconsistency in dataset selection, the number of devices, and task configurations. Many studies limit their evaluations to a small number of transmitters, often fewer than 10 to 20. This does not reflect the growing density and heterogeneity of wireless deployments in real-world indoor and outdoor environments. Although this limitation does not require a fundamental change in evaluation methodology, scaling the number of devices in benchmark evaluations is essential. Future research should emphasize large-scale validation experiments involving tens or even hundreds of devices, which would bring current research efforts closer to realistic deployment scenarios.

2) ROAD AHEAD

Despite the encouraging technical results, most existing solutions remain at the proof-of-concept stage, primarily evaluated in laboratory settings or fixed testbeds with static channel conditions and controlled interference [113]. While such setups are useful for dissecting the impact of specific channel or device-level features (e.g., hardware impairments), they fail to capture real-world complexities such as device mobility, spectrum congestion, and dynamic multipath environments. Consequently, the robustness of these models to environmental variation remains underexplored. There is a clear need to expand evaluations to real-world deployments, such as those examined in [114], [115] and [116], and to consider different channel models [117]. They should include both indoor and outdoor scenarios, particularly in use cases involving mobile devices, such as uncrewed aerial vehicles, for which the channel variation is very significant. Incorporating such conditions into benchmarking would better inform both model limitations and design priorities for real-world deployment.

A further critical issue is the lack of reproducibility. Many studies do not release their datasets, code, or full experimental details, impeding fair comparison across approaches and hindering progress toward standardized evaluation. This fragmentation not only limits scientific transparency but also slows the development of unified benchmarks. To address this, future research should prioritize the use and extension of open, community-maintained datasets, such as WISIG [113] and RF-DNA [116], and adhere to transparent evaluation protocols. Establishing widely accepted benchmarks will facilitate fair comparison, reproducibility, and ultimately, progress toward deployable and robust unsupervised SEI systems.

VIII. ANOMALY DETECTION

Anomaly detection in radio signal sensing helps identify unusual events and irregular patterns. These anomalies may indicate general malfunctions, equipment failures, or potential security threats that require immediate and appropriate actions [121]. However, in real-world settings, the presence of anomalies is sparse, so we often lack labeled data [122]. To address this challenge, unsupervised learning techniques have gained prominence in anomaly detection for radio signal sensing. These methods can learn normal patterns from unlabeled data and subsequently identify deviations that may correspond to anomalous activities [123], [124]. Moreover, the dynamic and often unpredictable nature of radio environments adds another layer of complexity to anomaly detection. Factors such as interference, signal attenuation, and device heterogeneity can lead to fluctuations that resemble anomalies but are not necessarily indicative of security threats or system failures. In general, reconstruction or generative methods are mostly utilized for unsupervised anomaly detection.

A. OVERALL AD COMPARATIVE OVERVIEW

In Table 8 we summarize the unsupervised anomaly detection works, noticing that most of them rely on generative methods (Section III-D), while two rely on predictive methods (Section III-E) as can be seen from the second column of the table and also highlighted with different colors. According to the third column, six of the architectures contain CNNs (see Section IV-B), out of which two are in combination with LSTM (see Section IV-C), and one is in combination with MLP (see Section IV-A. The fourth column reveals that four of the works rely on spectrogram data, three on I/Q data, and one on Stockwell Transform amplitudes, which are pictorial data.

Overall, the reviewed studies demonstrate that unsupervised deep learning methods can effectively detect radio frequency anomalies even under challenging SNR conditions. Most of the evaluated systems achieve high detection accuracy or strong anomaly discrimination (e.g., AUC above 0.9) using synthetic or limited real-world data. However, differences in input representations, SNR ranges, and evaluation metrics make direct quantitative comparison difficult. The limited availability of public datasets and code repositories, as indicated in the last two columns of Table 8, also constrains reproducibility and benchmarking across studies. In the subsequent sections, we further analyze the generative and predictive approaches in terms of their underlying architectures, training objectives, and generalization capabilities under variable noise and interference conditions.

B. OVERVIEW PER APPROACH TO AD

1) GENERATIVE

The reconstruction error of GANs and AEs (see Section III-D) can also serve as a metric for anomaly detection, allowing for pinpointing unusual behaviors effectively. In [69], the authors propose SAIFE, an adversarial

TABLE 8. Summary of works concerning unsupervised deep learning approaches and achievements for RF Anomaly detection.

Ref.	Approach	Module	Input	SNR	Num. clas.	Performance	Data	Code	Num. params.
[69]	AE + GAN	CNN + LSTM	I/Q	-20dB to 20dB	2	Accuracy: synthetic data 92.86%, real data 100%	✓	-	-
[55]	VAE	CNN	Spectrogram	-	4	PER [55] score AUC 0.98	✓	✓	≈22.5M
[120]	Enhanced GAN (E-GAN)	CNN + MLP	Spectrogram	-20dB to 20dB	2	Probability of detection 1 at −3dB ISR	-	-	-
[121]	GAN	CNN	FFT amplitudes TS	-20dB to 20dB	2	F1 score > 0.95	✓	-	17.5M
[122]	GAN and VAE	CNN	Stockwell Transform (amplitudes)	-	2	AD AUC > 0.97	-	-	-
[63]	Predictive learning	LSTM + MDN	I component (I/Q)	40dB	2	/	-	-	-
[64]	Predictive learning	CNN + LSTM	Spectrogram data	>5dB	3	Probability of detection 90%	-	-	4.7M

Legend: - Generative, - Predictive

AE-based model for wireless spectrum anomaly detection, achieving high accuracy with interpretable feature learning. Similarly, [55] demonstrated that VAE reconstruction can be used to detect anomalies in unauthorized bands by identifying the rise in the spectrogram's noise floor after reconstruction.

The study in [118] proposes a radio anomaly detection algorithm using a modified GAN with an encoder (E-GAN). It applies the Short Time Fourier Transform (STFT) to convert RF signals into spectrograms and detects anomalies based on reconstruction error and discriminator loss. This method also enables anomaly localization in the time-frequency domain. In [119], a similar result was shown where the authors demonstrated the potential of GANs (See Section III-D) in detecting anomalies in the spectrum and mitigating security attacks. Additionally, [120] also applied unsupervised approaches on spectrum anomaly detection in mmWave radios; more specifically, they utilised the Conditional GAN and Auxiliary Classifier, GAN, and VAE.

2) PREDICTIVE

In [63], authors proposed an unsupervised anomaly detection method using a combination of LSTM and mixture density networks (MDNs), i.e., LSTM-MDN, for time-series data in digital radio transmissions. The model learned the expected signal distributions and detected anomalies using negative log-likelihood. On the other hand, the authors in [64] utilized a Deep Predictive Coding Neural Network (see Section III-E) on spectrograms of RF signals to detect anomalies such as jamming, chirping of transmitters, spectrum hijacking, and node failure.

C. DATASETS

In this section, we report on some of the openly available datasets, summarized in Table 9 for exploring and addressing the anomaly detection.

In [125], a wideband spectrum monitoring dataset is introduced with annotated anomalous signals. This dataset uses licensed frequency bands and STFT spectrograms to capture both time and frequency characteristics. Next, [126] proposed an RSSI-based synthetic anomaly detection dataset. The authors injected 4 different types of anomalies into the Rutgers dataset and also provided code for their generation. Finally, the dataset used in [127] focuses on RF jamming scenarios. Experimentally measured spectral scan data provides a baseline for the evaluation of jamming detection algorithms.

TABLE 9. Summary of open datasets for Anomaly detection.

Ref.	Characteristics	Specific Features
[127]	Real-world dataset of 5G spectrum data	This dataset comprises anomalous signals that have been manually annotated and leverages licensed frequency bands along with STFT spectrograms to capture both temporal and spectral features.
[128]	Synthetic dataset of RSSI wireless signals	Provides a script for generating "unlimited" samples of 4 different anomaly types.
[129]	Measured RF jamming data	Comprises of spectrum data with jamming anomalies

D. FINDINGS

Predictive and reconstruction-based approaches each have distinct strengths and challenges when applied to anomaly detection in the wireless spectrum. Predictive methods, such as LSTM-MDN and Deep Predictive Coding Neural Networks [63], [64], are particularly effective in detecting temporal anomalies by learning expected signal behaviors and identifying deviations. These methods excel in time-series analysis and can capture anomalies that emerge as disruptions in the expected signal evolution. However, the predictive methods are, in general, less robust, hence the reconstruction-based techniques appear more frequently in the literature.

Reconstruction-based techniques, including AEs, VAEs, and GANs [55], [69], [118], [119], [120], leverage the

inability of generative models to accurately reconstruct anomalous signals. These models have demonstrated robustness against varying noise levels and the ability to operate in unsupervised settings. The advantage of using GAN-based and VAE-based methods lies in their ability to model complex signal distributions. Additionally, these approaches do not require any labeled data.

1) ROAD AHEAD

Despite their advantages, both approaches face challenges in real-world deployments. Predictive models may struggle with unseen spectrum events that do not follow expected patterns, while reconstruction-based models are harder to train and learn proper latent representation of non-anomalous examples. Future research could focus on hybrid models that integrate predictive and reconstruction-based mechanisms to leverage the strengths of both paradigms, improving anomaly detection reliability in dynamic spectrum environments.

IX. CHALLENGES AND FUTURE DIRECTIONS

In this section, we concisely summarize the identified challenges and potential future research directions, based on the analysis in each of the four research topics tied to signal recognition in wireless communications. Furthermore, we also provide a visual summary highlighting the most important aspects in Figure 18.

1) AUTOMATIC MODULATION CLASSIFICATION

In AMC, research covers a broad range of SNR levels, channel conditions, and datasets, yet most studies still rely on simulated benchmarks such as RadioML, which only partially reflect real-world variability. Future work should focus on evaluations using over-the-air (OtA) collected data that reflect realistic propagation, hardware, and interference effects. Another important direction is small-sample analysis, where current inconsistencies in labeled data ratios constrain the direct comparison. Establishing standardized benchmarks with fixed sample sizes and unified training protocols would improve reproducibility and comparability. Among unsupervised methods, contrastive learning stands out for its strong performance and ability to learn discriminative features through data-specific augmentations; thus, designing effective, domain-relevant augmentations is another promising research direction. Finally, hybrid architectures that include sequence-processing components such as RNNs or attention mechanisms have shown superior accuracy and label efficiency, warranting further exploration using real-world datasets under practical operating conditions.

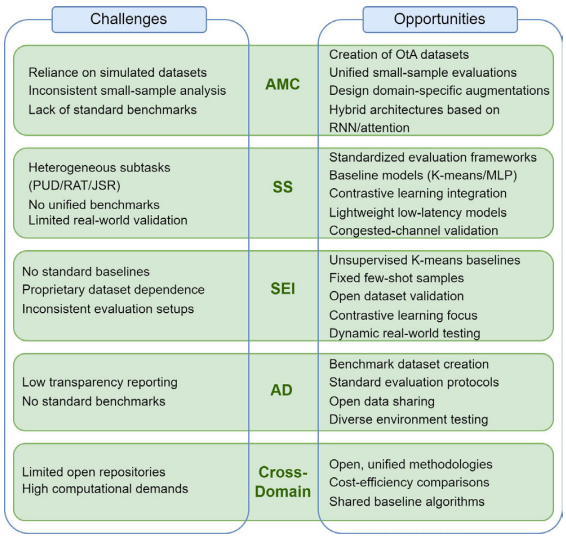


FIGURE 18. Summary of existing challenges and future research directions for each domain of work.

2) SPECTRUM SENSING

Spectrum sensing encompasses three complementary tasks: PU detection, RAT recognition, and JSR, each with distinct challenges that complicate unified evaluation. Developing benchmark datasets and standardized evaluation frameworks for these tasks would greatly enhance comparability across studies. Including simple reference models, such as K-means for unsupervised and MLP for supervised settings, along with common metrics like the Silhouette score, could streamline benchmarking and reduce dependence on highly application-specific designs. Future work should also explore contrastive learning, which has shown strong feature extraction capability in AMC and SEI, as a promising direction for spectrum sensing. Given the prevalence of interference and noise in this domain, rigorous validation under congested channel conditions with multiple PUs is essential for assessing deployment readiness. In terms of architectures, lightweight models are preferable for low-latency PU and RAT detection.

3) SPECIFIC EMITTER IDENTIFICATION

For the SEI task, both unsupervised and supervised evaluation setups are common, corresponding to zero-shot and few-shot learning, respectively. Establishing a standard unsupervised baseline, such as K-means, together with consistent clustering metrics, would help define lower performance boundaries and improve comparability across studies. For the supervised few-shot setting, standardizing the number of labeled samples would enhance transparency and facilitate fair evaluation, similar to practices in AMC research. Although most studies already use real-world data collected from proprietary testbeds, transparency could be improved by including validation with open datasets, such as the ones reviewed in this work. In terms of learning approaches, contrastive learning remains the most widely adopted and effective paradigm, while hybrid architectures that include sequence-processing modules such as LSTM or self-attention show strong potential for capturing temporal and contextual features. Future research should emphasize real-world evaluations under dynamic conditions involving mobile devices and changing propagation environments, and should aim at complementing proprietary experiments with

open, standardized datasets to ensure reproducibility and comparability.

4) ANOMALIES DETECTION

The AD domain currently exhibits the least transparency, with limited information available regarding model architectures, datasets, and code accessibility. This gap highlights an urgent need for the introduction of standard benchmark datasets, baseline algorithms, and evaluation methodologies to enable comparability and reproducibility.

5) FINAL REMARKS

Across all four domains, a recurring challenge is the lack of open science practices such as open-source code, open data repositories and standardized benchmarks. This limits reproducibility and slows progress, as researchers must repeatedly replicate experimental setups that could otherwise serve as shared baselines. Adopting standardized methodologies, including open data, baseline algorithms, unified metrics, and public access to implementation materials, would greatly benefit future research. Together with the channel conditions and the SNR analysis, they represent the key common, domain-agnostic resources based on which a transparent and fair comparison of different models is possible for each of the research topics surveyed in this work. In Figure 19, we show the degree of availability of the corresponding resources, relative to the number of surveyed works, for each specific topic. As can be seen, and also noticed before, the AMC and SEI domains are the most thoroughly analysed, and the AD domain is the least transparent one.

It is important to recognize that all unsupervised deep learning models depend on the intensive processing of large volumes of unlabeled data, which is often taken for granted in the literature. The substantial computational and memory resources required for such training should be carefully considered in future work, besides the size of the models. If a supervised model can achieve comparable performance while requiring only a fraction of the computational cost, the unsupervised approach may not always represent the most efficient or practical solution.

Finally, we note that current works lack considerations regarding throughput/latency trade-offs, largely due to the focus on the model's performance. However, such considerations that are more related to a model operation once it is deployed, may affect the performance of network applications, therefore should be considered in the future, together with energy consumption and CO_2 footprint of models.

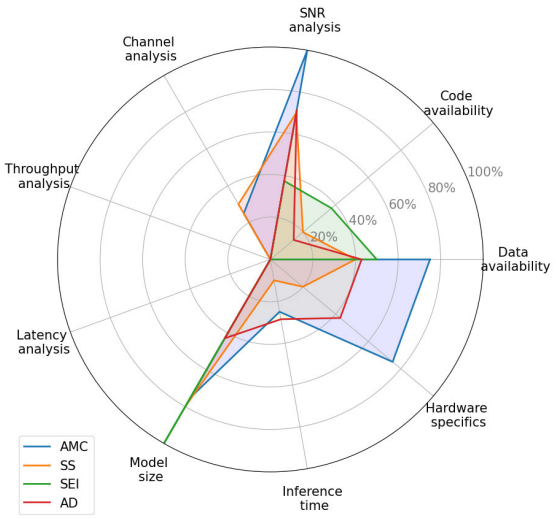


FIGURE 19. Web plots of the key common requirements for comparative analysis of the published works across the four signal recognition challenges.

X. CONCLUSION

This survey has provided a comprehensive overview of the current state of unsupervised deep learning techniques applied to wireless signal recognition, focusing on key challenges, i.e., AMC, signal sensing, SEI, and anomaly detection. We have systematically reviewed deep clustering, contrastive learning, autoencoder-based reconstruction, generative models, and predictive learning, highlighting their potential to address the limitations of supervised methods, particularly the reliance on large labeled datasets. We have complemented this method-oriented review with an overview of the most representative openly available datasets that can be utilized for further research in the area of a given wireless signal recognition challenge.

By presenting both a challenge-driven and a method-oriented perspective, this survey aims to inspire further research and provide a roadmap for the development of data-efficient, self-supervised models that can enhance the adaptability and intelligence of future wireless networks.

Despite notable progress in recent years, the application of unsupervised deep learning in wireless communications remains relatively underexplored compared to some other domains such as computer vision and natural language processing. Significant research opportunities exist in improving the scalability and adaptability of these models to dynamic and complex radio environments, some of them also identified in discussions within sections corresponding to key challenges.

REFERENCES

[1] Z. Ke and H. Vikalo, "Real-time radio technology and modulation classification via an LSTM auto-encoder," *IEEE Trans. Wireless Commun.*, vol. 21, no. 1, pp. 370–382, Jan. 2022.

[2] S. Rajendran, W. Meert, D. Giustiniano, V. Lenders, and S. Pollin, "Deep learning models for wireless signal classification with distributed low-cost spectrum sensors," *IEEE Trans. Cognit. Commun. Netw.*, vol. 4, no. 3, pp. 433–445, Sep. 2018.

[3] J. Robinson, S. Kuzdeba, J. Stankowicz, and J. M. Carmack, "Dilated causal convolutional model for RF fingerprinting," in *Proc. 10th Annu. Comput. Commun. Workshop Conf. (CCWC)*, Jan. 2020, pp. 0157–0162.

[4] I. G. A. Poornima and B. Paramasivan, "Anomaly detection in wireless sensor network using machine learning algorithm," *Comput. Commun.*, vol. 151, pp. 331–337, Feb. 2020.

[5] L. Milosheski, G. Cerar, B. Bertalanič, C. Fortuna, and M. Mohorčič, "Self-supervised learning for clustering of wireless spectrum activity," *Comput. Commun.*, vol. 212, pp. 353–365, Dec. 2023.

[6] 3GPP. (2016). *Release 13*. Accessed: Jul. 11, 2025. [Online]. Available: https://www.3gpp.org/release-13

[7] G. Reus-Muns, P. S. Upadhyaya, U. Demir, N. Stephenson, N. Soltani, V. K. Shah, and K. R. Chowdhury, "SenseORAN: O-RAN-based radar detection in the CBRS band," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 2, pp. 326–338, Feb. 2024.

[8] I. R. Sector, *Framework and Overall Objectives of the Future Development of IMT for 2030 and Beyond*, document ITU-R M.2160, 2023.

[9] *Representative Use Cases and Key Network Requirements for Network 2030*, document NET2030-O-027, 2020.

[10] S. D'Oro, L. Bonati, M. Polese, and T. Melodia, "OrchestRAN: Network automation through orchestrated intelligence in the open RAN," in *Proc. IEEE Conf. Comput. Commun.*, May 2022, pp. 270–279.

[11] L. Bonati, M. Polese, S. D'Oro, S. Basagni, and T. Melodia, "Open, programmable, and virtualized 5G networks: State-of-the-art and the road ahead," *Comput. Netw.*, vol. 182, Dec. 2020, Art. no. 107516.

[12] M. Polese, L. Bonati, S. D'Oro, S. Basagni, and T. Melodia, "Understanding O-RAN: Architecture, interfaces, algorithms, security, and research challenges," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 2, pp. 1376–1411, 2nd Quart., 2023.

[13] G. D. Durgin, M. A. Varner, N. Patwari, S. K. Kasera, and J. Van der Merwe, "Digital spectrum twinning for next-generation spectrum management and metering," in *Proc. IEEE 2nd Int. Conf. Digit. Twins Parallel Intell. (DTPI)*, Oct. 2022, pp. 1–6.

[14] S. Tadik, K. M. Graves, M. A. Varner, C. R. Anderson, D. M. Johnson, S. K. Kasera, N. Patwari, J. Van der Merwe, and G. D. Durgin, "Digital spectrum twins for enhanced spectrum sharing and other radio applications," *IEEE J. Radio Freq. Identificat.*, vol. 8, pp. 376–391, 2024.

[15] R. Schwarz. (2023). *The Metaverse and Extended Reality—Implications for Wireless Communications*. Accessed: May 14, 2025. [Online]. Available: https://www.rohde-schwarz.com/us/solutions/wireless-communications-testing/wireless-standards/5g-nr/extended-reality-xr-testing/white-paper-the-metaverse-and-extended-reality_257899.html

[16] I. F. Akyildiz and H. Guo, "Wireless communication research challenges for extended reality (XR)," *ITU J. Future Evolving Technol.*, vol. 3, no. 2, pp. 273–287, 2022.

[17] L. Milosheski, M. Mohorčič, and C. Fortuna, "Spectrum sensing with deep clustering: Label-free radio access technology recognition," *IEEE Open J. Commun. Soc.*, vol. 5, pp. 4746–4763, 2024.

[18] S. Atapattu, C. Tellambura, and H. Jiang, *Energy Detection for Spectrum Sensing in Cognitive Radio*, vol. 6. Cham, Switzerland: Springer, 2014.

[19] T. Gale, T. Šolc, R.-A. Mosoi, M. Mohorcic, and C. Fortuna, "Automatic detection of wireless transmissions," *IEEE Access*, vol. 8, pp. 24370–24384, 2020.

[20] K. Kim, I. A. Akbar, K. K. Bae, J.-S. Um, C. M. Spooner, and J. H. Reed, "Cyclostationary approaches to signal detection and classification in cognitive radio," in *Proc. 2nd IEEE Int. Symp. New Frontiers Dyn. Spectr. Access Netw.*, Apr. 2007, pp. 212–215.

[21] F. Paisana, N. Prasad, A. Rodrigues, and R. Prasad, "An alternative implementation of a cyclostationary detector," in *Proc. 15th Int. Symp. Wireless Pers. Multimedia Commun.*, Sep. 2012, pp. 45–49.

[22] T. Ya, L. Yun, Z. Haoran, J. Zhang, Y. Wang, G. Guan, and M. Shiwen, "Large-scale real-world radio signal recognition with deep learning," *Chin. J. Aeronaut.*, vol. 35, no. 9, pp. 35–48, Sep. 2022.

[23] D. Liu, P. Wang, T. Wang, and T. Abdelzaher, "Self-contrastive learning based semi-supervised radio modulation classification," in *Proc. IEEE Mil. Commun. Conf. (MILCOM)*, Nov. 2021, pp. 777–782.

[24] J. Gui, T. Chen, J. Zhang, Q. Cao, Z. Sun, H. Luo, and D. Tao, "A survey on self-supervised learning: Algorithms, applications, and future trends," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 12, pp. 9052–9071, Dec. 2024.

[25] Y. Chen, M. Mancini, X. Zhu, and Z. Akata, "Semi-supervised and unsupervised deep visual learning: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 3, pp. 1327–1347, Mar. 2024.

[26] Q. Meng, H. Qian, Y. Liu, Y. Xu, Z. Shen, and L. Cui, "Unsupervised representation learning for time series: A review," 2023, *arXiv:2308.01578*.

[27] S. N. Syed, P. I. Lazaridis, F. A. Khan, Q. Z. Ahmed, M. Hafeez, A. Ivanov, V. Poulkov, and Z. D. Zaharis, "Deep neural networks for spectrum sensing: A review," *IEEE Access*, vol. 11, pp. 89591–89615, 2023.

[28] L. J. Wong, W. H. Clark, B. Flowers, R. M. Buehrer, W. C. Headley, and A. J. Michaels, "An RFML ecosystem: Considerations for the application of deep learning to spectrum situational awareness," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 2243–2264, 2021.

[29] A. Upadhye, P. Saravanan, S. S. Chandra, and S. Gurugopinath, "A survey on machine learning algorithms for applications in cognitive radio networks," in *Proc. IEEE Int. Conf. Electron., Comput. Commun. Technol. (CONECCT)*, Jul. 2021, pp. 01–06.

[30] C. Zhang, P. Patras, and H. Haddadi, "Deep learning in mobile and wireless networking: A survey," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2224–2287, 3rd Quart., 2019.

[31] Q. Mao, F. Hu, and Q. Hao, "Deep learning for intelligent wireless networks: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 2595–2621, 4th Quart., 2018.

[32] J. Jagannath, N. Polosky, A. Jagannath, F. Restuccia, and T. Melodia, "Machine learning for wireless communications in the Internet of Things: A comprehensive survey," *Ad Hoc Netw.*, vol. 93, Oct. 2019, Art. no. 101913.

[33] D. Adesina, C.-C. Hsieh, Y. E. Sagduyu, and L. Qian, "Adversarial machine learning in wireless communications using RF data: A review," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 1, pp. 77–100, 1st Quart., 2023.

[34] M. Usama, J. Qadir, A. Raza, H. Arif, K.-L.-A. Yau, Y. Elkhatib, A. Hussain, and A. Al-Fuqaha, "Unsupervised machine learning for networking: Techniques, applications and research challenges," *IEEE Access*, vol. 7, pp. 65579–65615, 2019.

[35] Q.-V. Pham, N. T. Nguyen, T. Huynh-The, L. B. Le, K. Lee, and W.-J. Hwang, "Intelligent radio signal processing: A survey," *IEEE Access*, vol. 9, pp. 83818–83850, 2021.

[36] H. Zhou, J. Bai, Y. Wang, J. Ren, X. Yang, and L. Jiao, "Deep radio signal clustering with interpretability analysis based on saliency map," *Digit. Commun. Netw.*, vol. 10, no. 5, pp. 1448–1458, Oct. 2024.

[37] E. Perenda, S. Rajendran, G. Bovet, M. Zheleva, and S. Pollin, "Contrastive learning with self-reconstruction for channel-resilient modulation classification," in *Proc. IEEE Conf. Comput. Commun.*, May 2023, pp. 1–10.

[38] K. Zhang, Q. Wen, C. Zhang, R. Cai, M. Jin, Y. Liu, J. Y. Zhang, Y. Liang, G. Pang, D. Song, and S. Pan, "Self-supervised learning for time series analysis: Taxonomy, progress, and prospects," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 10, pp. 6775–6794, Oct. 2024.

[39] N. Liao, Y. Zhang, Y. Wang, and Y. Liu, "Deep denoising and clustering-based cooperative spectrum sensing for non-orthogonal multiple access," *IEEE Trans. Cognit. Commun. Netw.*, vol. 10, no. 5, pp. 1831–1842, Oct. 2024.

[40] Z. Su, K. C. Teh, S. G. Razul, and A. C. Kot, "Deep non-cooperative spectrum sensing over Rayleigh fading channel," *IEEE Trans. Veh. Technol.*, vol. 71, no. 4, pp. 4460–4464, Apr. 2022.

[41] J. Stankowicz and S. Kuzdeba, "Unsupervised emitter clustering through deep manifold learning," in *Proc. IEEE 11th Annu. Comput. Commun. Workshop Conf. (CCWC)*, Jan. 2021, pp. 0732–0737.

[42] X. Hao, Z. Feng, R. Liu, S. Yang, L. Jiao, and R. Luo, "Contrastive self-supervised clustering for specific emitter identification," *IEEE Internet Things J.*, vol. 10, no. 23, pp. 20803–20818, Dec. 2023.

[43] N. A. Khalek and W. Hamouda, "DeepSense: An unsupervised deep clustering approach for cooperative spectrum sensing," in *Proc. IEEE Int. Conf. Commun.*, May 2023, pp. 1868–1873.

[44] L. Ericsson, H. Gouk, C. C. Loy, and T. M. Hospedales, "Self-supervised representation learning: Introduction, advances, and challenges," *IEEE Signal Process. Mag.*, vol. 39, no. 3, pp. 42–62, May 2022.

[45] S. Zhou, H. Xu, Z. Zheng, J. Chen, Z. Li, J. Bu, J. Wu, X. Wang, W. Zhu, and M. Ester, "A comprehensive survey on deep clustering: Taxonomy, challenges, and future directions," *ACM Comput. Surv.*, vol. 57, no. 3, pp. 1–38, Mar. 2025.

[46] J.-B. Grill, F. Strub, F. Altché, C. Tallec, P. Richemond, E. Buchatskaya, C. Doersch, B. A. Pires, Z. Guo, M. G. Azar, B. Piot, K. Kavukcuoglu, R. Munos, and M. Valko, "Bootstrap your own latent—A new approach to self-supervised learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 21271–21284.

[47] H. Hu, X. Wang, Y. Zhang, Q. Chen, and Q. Guan, "A comprehensive survey on contrastive learning," *Neurocomputing*, vol. 610, Dec. 2024, Art. no. 128645.

[48] K. Davaslioglu, S. Boztas, M. C. Ertem, Y. E. Sagduyu, and E. Ayanoglu, "Self-supervised RF signal representation learning for NextG signal classification with deep learning," *IEEE Wireless Commun. Lett.*, vol. 12, no. 1, pp. 65–69, Jan. 2023.

[49] A. van den Oord, Y. Li, and O. Vinyals, "Representation learning with contrastive predictive coding," 2018, *arXiv:1807.03748*.

[50] P. Li, Y. Pei, and J. Li, "A comprehensive survey on design and application of autoencoder in deep learning," *Appl. Soft Comput.*, vol. 138, May 2023, Art. no. 110176.

[51] S. Subray, S. Tschimben, and K. Gifford, "Towards enhancing spectrum sensing: Signal classification using autoencoders," *IEEE Access*, vol. 9, pp. 82288–82299, 2021.

[52] K. Huang, J. Yang, H. Liu, and P. Hu, "Deep learning of radio frequency fingerprints from limited samples by masked autoencoding," *IEEE Wireless Commun. Lett.*, vol. 14, no. 12, pp. 3842–3846, Dec. 2025.

[53] J. Robinson and S. Kuzdeba, "Novel device detection using RF fingerprints," in *Proc. IEEE 11th Annu. Comput. Commun. Workshop Conf. (CCWC)*, Jan. 2021, pp. 0648–0654.

[54] C. Ezemaduka and A. A. Abouzeid, "Privacy-aware deep learning based localization of spectrum violators," in *Proc. IEEE 20th Int. Conf. Mobile Ad Hoc Smart Syst. (MASS)*, Sep. 2023, pp. 98–106.

[55] Y. Tian, H. Liao, J. Xu, Y. Wang, S. Yuan, and N. Liu, "Unsupervised spectrum anomaly detection method for unauthorized bands," *Space, Sci. Technol.*, vol. 2022, 2022, doi: 10.34133/2022/9865016.

[56] A. Jabbar, X. Li, and B. Omar, "A survey on generative adversarial networks: Variants, applications, and training," *ACM Comput. Surv.*, vol. 54, no. 8, pp. 1–49, Nov. 2022.

[57] N. Van Huynh, J. Wang, H. Du, D. T. Hoang, D. Niyato, D. N. Nguyen, D. I. Kim, and K. B. Letaief, "Generative AI for physical layer communications: A survey," *IEEE Trans. Cognit. Commun. Netw.*, vol. 10, no. 3, pp. 706–728, Jun. 2024.

[58] C. Zhao, H. Du, D. Niyato, J. Kang, Z. Xiong, D. I. Kim, X. Shen, and K. B. Letaief, "Generative AI for secure physical layer communications: A survey," *IEEE Trans. Cognit. Commun. Netw.*, vol. 11, no. 1, pp. 3–26, Feb. 2025.

[59] X. Zhou, Z. Sun, and H. Wu, "Wireless signal enhancement based on generative adversarial networks," *Ad Hoc Netw.*, vol. 103, Jun. 2020, Art. no. 102151.

[60] J. Carmack, A. Bhatia, J. Robinson, J. Majewski, and S. Kuzdeba, "Neural network generative models for radio frequency data," in *Proc. IEEE 12th Annu. Ubiquitous Comput., Electron. Mobile Commun. Conf. (UEMCON)*, Dec. 2021, pp. 0577–0582.

[61] J. Wang, H. Du, D. Niyato, J. Kang, S. Cui, X. Shen, and P. Zhang, "Generative AI for integrated sensing and communication: Insights from the physical layer perspective," *IEEE Wireless Commun.*, vol. 31, no. 5, pp. 246–255, Oct. 2024.

[62] L. Xu, W. Shi, X. Fu, H. Xu, Y. Wang, B. Adebisi, and G. Gui, "Few-shot specific emitter identification method using rotation feature decoupling for secure 6G," in *Proc. IEEE 23rd Int. Conf. Commun. Technol. (ICCT)*, Oct. 2023, pp. 490–494.

[63] M. Walton, M. Ayache, L. Straatemeier, D. Gebhardt, and B. Migliori, "Unsupervised anomaly detection for digital radio frequency transmissions," in *Proc. 16th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2017, pp. 826–832.

[64] N. Tandiya, A. Jauhar, V. Marojevic, and J. H. Reed, "Deep predictive coding neural network for RF anomaly detection in wireless networks," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, May 2018, pp. 1–6.

[65] J. Fontaine, E. Fonseca, A. Shahid, M. Kist, L. A. DaSilva, I. Moerman, and E. De Poorter, "Towards low-complexity wireless technology classification across multiple environments," *Ad Hoc Netw.*, vol. 91, Aug. 2019, Art. no. 101881.

[66] J. Xie, J. Fang, C. Liu, and L. Yang, "Unsupervised deep spectrum sensing: A variational auto-encoder based approach," *IEEE Trans. Veh. Technol.*, vol. 69, no. 5, pp. 5307–5319, May 2020.

[67] N. Zhang, J. Shen, Y. Shi, and Y. Li, "CNN-zero: A zero-shot learning framework for jamming identification," in *Proc. IEEE 22nd Int. Conf. Commun. Technol. (ICCT)*, Nov. 2022, pp. 1126–1131.

[68] S. Kuzdeba, J. Robinson, and J. Carmack, "Transfer learning with radio frequency signals," in *Proc. IEEE 18th Annu. Consum. Commun. Netw. Conf. (CCNC)*, Jan. 2021, pp. 1–9.

[69] S. Rajendran, W. Meert, V. Lenders, and S. Pollin, "SAIFE: Unsupervised wireless spectrum anomaly detection with interpretable features," in *Proc. IEEE Int. Symp. Dyn. Spectr. Access Netw. (DySPAN)*, Oct. 2018, pp. 1–9.

[70] W. Kong, X. Jiao, Y. Xu, B. Zhang, and Q. Yang, "A transformer-based contrastive semi-supervised learning framework for automatic modulation recognition," *IEEE Trans. Cognit. Commun. Netw.*, vol. 9, no. 4, pp. 950–962, Aug. 2023.

[71] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2025, pp. 5998–6008.

[72] H. Han, W. Li, Z. Feng, G. Fang, Y. Xu, and Y. Xu, "Proceed from known to unknown: Jamming pattern recognition under open-set setting," *IEEE Wireless Commun. Lett.*, vol. 11, no. 4, pp. 693–697, Apr. 2022.

[73] J. Xie, J. Fang, C. Liu, and X. Li, "Deep learning-based spectrum sensing in cognitive radio: A CNN-LSTM approach," *IEEE Commun. Lett.*, vol. 24, no. 10, pp. 2196–2200, Oct. 2020.

[74] C. Xiao, S. Yang, Z. Feng, and L. Jiao, "MCLHN: Toward automatic modulation classification via masked contrastive learning with hard negatives," *IEEE Trans. Wireless Commun.*, vol. 23, no. 10, pp. 14304–14319, Oct. 2024.

[75] D. Wu, J. Shi, Z. Li, M. Du, F. Liu, and F. Zeng, "Contrastive semi-supervised learning with pseudo-label for radar signal automatic modulation recognition," *IEEE Sensors J.*, vol. 24, no. 19, pp. 30399–30411, Oct. 2024.

[76] Y. Shi, H. Xu, Y. Zhang, Z. Qi, and D. Wang, "GAF-MAE: A self-supervised automatic modulation classification method based on gramian angular field and masked autoencoder," *IEEE Trans. Cognit. Commun. Netw.*, vol. 10, no. 1, pp. 94–106, Feb. 2024.

[77] Y. Wang, J. Bai, Z. Xiao, Z. Chen, Y. Xiong, H. Jiang, and L. Jiao, "AutoSMC: An automated machine learning framework for signal modulation classification," *IEEE Trans. Inf. Forensics Security*, vol. 19, pp. 6225–6236, 2024.

[78] Y. Lin, Y. Tu, Z. Dou, L. Chen, and S. Mao, "Contour stella image and deep learning for signal recognition in the physical layer," *IEEE Trans. Cognit. Commun. Netw.*, vol. 7, no. 1, pp. 34–46, Mar. 2021.

[79] T. Huynh-The, Q.-V. Pham, T.-V. Nguyen, T. T. Nguyen, R. Ruby, M. Zeng, and D.-S. Kim, "Automatic modulation classification: A deep architecture survey," *IEEE Access*, vol. 9, pp. 142950–142971, 2021.

[80] S. Hu, Y. Pei, P. P. Liang, and Y.-C. Liang, "Deep neural network for robust modulation classification under uncertain noise conditions," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 564–577, Jan. 2020.

[81] S. Zheng, X. Zhou, L. Zhang, P. Qi, K. Qiu, J. Zhu, and X. Yang, "Toward next-generation signal intelligence: A hybrid knowledge and data-driven deep learning framework for radio signal classification," *IEEE Trans. Cognit. Commun. Netw.*, vol. 9, no. 3, pp. 564–579, Jun. 2023.

[82] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Trans. Cognit. Commun. Netw.*, vol. 3, no. 4, pp. 563–575, Dec. 2017.

[83] Y. Wang, G. Gui, H. Gacanin, T. Ohtsuki, H. Sari, and F. Adachi, "Transfer learning for semi-supervised automatic modulation classification in ZF-MIMO systems," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 10, no. 2, pp. 231–239, Jun. 2020.

[84] Y. Dong, X. Jiang, H. Zhou, Y. Lin, and Q. Shi, "SR2CNN: Zero-shot learning for signal recognition," *IEEE Trans. Signal Process.*, vol. 69, pp. 2316–2329, 2021.

[85] T. J. O'Shea, J. Corgan, and T. C. Clancy, "Convolutional radio modulation recognition networks," in *Proc. Int. Conf. Eng. Appl. Neural Netw.* Cham, Switzerland: Springer, 2016, pp. 213–226.

[86] X. Chen, S. Xie, and K. He, "An empirical study of training self-supervised vision transformers," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9640–9649.

[87] T. Chen, S. Kornblith, M. Norouzi, and G. E. Hinton, "A simple framework for contrastive learning of visual representations," in *Proc. Int. Conf. Mach. Learn.*, vol. 1, 2024, pp. 1597–1607.

[88] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 9726–9735.

[89] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," 2018, *arXiv:1803.01271*.

[90] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer. (2017). *Automatic Differentiation in Pytorch*. [Online]. Available: https://api.semanticscholar.org/CorpusID:40027675

[91] T. J. O'Shea, T. Roy, and T. C. Clancy, "Over-the-air deep learning based radio signal classification," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 168–179, Feb. 2018.

[92] J. Wang, Q. Zhu, Z. Lin, J. Chen, G. Ding, Q. Wu, G. Gu, and Q. Gao, "Sparse Bayesian learning-based hierarchical construction for 3D radio environment maps incorporating channel shadowing," *IEEE Trans. Wireless Commun.*, vol. 23, no. 10, pp. 14560–14574, Oct. 2024.

[93] J. Downey, B. Hilburn, T. O'Shea, and N. West, "Machine learning remakes radio," *IEEE Spectr.*, vol. 57, no. 5, pp. 35–39, May 2020.

[94] T. Šolc, C. Fortuna, and M. Mohorčič, "Low-cost testbed development and its applications in cognitive radio prototyping," in *Cognitive Radio and Networking for Heterogeneous Wireless Networks*. Cham, Switzerland: Springer, 2015, pp. 361–405.

[95] *Intelligent Transport Systems (ITS); Access Layer Specification for Intelligent Transport Systems Operating in the 5 GHz Frequency Band; Technical Report en 302 663; Version 1.2. 0; 650 Route Des Lucioles F-06921 Sophia Antipolis Cedex-France*, ETSI, Sophia Antipolis, France, 2012.

[96] M. Girmay, V. Maglogiannis, D. Naudts, M. Aslam, A. Shahid, and I. Moerman, "Technology recognition and traffic characterization for wireless technologies in ITS band," *Veh. Commun.*, vol. 39, Feb. 2023, Art. no. 100563.

[97] T. Pang, H. Zheng, Y. Quan, and H. Ji, "Recorrupted-to-recorrupted: Unsupervised deep learning for image denoising," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 2043–2052.

[98] "Amendment of the commission's rules with regard to commercial operations in the 3550–3650 MHz band," Federal Communications Commission, Washington, DC, USA, GN Docket 12-354, 2015.

[99] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 234–241.

[100] S. D'Oro, M. Polese, L. Bonati, H. Cheng, and T. Melodia, "DApps: Distributed applications for real-time inference and control in O-RAN," *IEEE Commun. Mag.*, vol. 60, no. 11, pp. 52–58, Nov. 2022.

[101] X. Zhang, Y. Huang, M. Lin, Y. Tian, and J. An, "Transmitter identification with contrastive learning in incremental open-set recognition," *IEEE Internet Things J.*, vol. 11, no. 3, pp. 4693–4711, Feb. 2024.

[102] M. Zhan, Y. Li, H. Cui, B. Li, J. Zhang, C. Li, and W. Wang, "MCRFF: A meta-contrastive learning-based RF fingerprinting method," in *Proc. IEEE Mil. Commun. Conf. (MILCOM)*, Oct. 2023, pp. 391–396.

[103] C. Liu, X. Fu, Y. Wang, L. Guo, Y. Liu, Y. Lin, H. Zhao, and G. Gui, "Overcoming data limitations: A few-shot specific emitter identification method using self-supervised learning and adversarial augmentation," *IEEE Trans. Inf. Forensics Security*, vol. 19, pp. 500–513, Oct. 2024, doi: 10.1109/TIFS.2023.3324394.

[104] L. Ding, S. Wang, F. Wang, and W. Zhang, "Specific emitter identification via convolutional neural networks," *IEEE Commun. Lett.*, vol. 22, no. 12, pp. 2591–2594, Dec. 2018.

[105] X. Wen, C. Cao, Y. Li, and Y. Sun, "DRSN with simple parameter-free attention module for specific emitter identification," in *Proc. IEEE Int. Conf. Trust, Secur. Privacy Comput. Commun. (TrustCom)*, Dec. 2022, pp. 192–200.

[106] Z. Xu, G. Han, L. Liu, H. Zhu, and J. Peng, "A lightweight specific emitter identification model for IIoT devices based on adaptive broad learning," *IEEE Trans. Ind. Informat.*, vol. 19, no. 5, pp. 7066–7075, May 2023.

[107] Y. Luo, X. Chen, N. Ge, W. Feng, and J. Lu, "Transformer-based device-type identification in heterogeneous IoT traffic," *IEEE Internet Things J.*, vol. 10, no. 6, pp. 5050–5062, Mar. 2023.

[108] S. Basak, S. Rajendran, S. Pollin, and B. Scheers, "Combined RF-based drone detection and classification," *IEEE Trans. Cognit. Commun. Netw.*, vol. 8, no. 1, pp. 111–120, Mar. 2022.

[109] J. Wang, B. Zhang, J. Zhang, N. Yang, G. Wei, and D. Guo, "Specific emitter identification based on deep adversarial domain adaptation," in *Proc. 4th Int. Conf. Inf. Commun. Signal Process. (ICICSP)*, Sep. 2021, pp. 104–109.

[110] L. J. Wong, W. C. Headley, S. Andrews, R. M. Gerdes, and A. J. Michaels, "Clustering learned CNN features from raw I/Q data for emitter identification," in *Proc. IEEE Mil. Commun. Conf. (MILCOM)*, Oct. 2018, pp. 26–33.

[111] J. Robinson and S. Kuzdeba, "RiftNet: Radio frequency large population classification," in *Proc. IEEE 18th Annu. Consum. Commun. Netw. Conf. (CCNC)*, Jan. 2021, pp. 1–6, doi: 10.1109/CCNC49032.2021.9369455.

[112] A. Al-Shawabka, F. Restuccia, S. D'Oro, T. Jian, B. C. Rendon, N. Soltani, J. Dy, S. Ioannidis, K. Chowdhury, and T. Melodia, "Exposing the fingerprint: Dissecting the impact of the wireless channel on radio fingerprinting," in *Proc. IEEE Conf. Comput. Commun.*, Jul. 2020, pp. 646–655.

[113] S. Hanna, S. Karunaratne, and D. Cabric, "WiSig: A large-scale WiFi signal dataset for receiver and channel agnostic RF fingerprinting," *IEEE Access*, vol. 10, pp. 22808–22818, 2022.

[114] T. Zhao, B. W. Domae, C. Steigerwald, L. B. Paradis, T. Chabuk, and D. Cabric, "Drone RF signal detection and fingerprinting: UAVSig dataset and deep learning approach," in *Proc. IEEE Mil. Commun. Conf. (MILCOM)*, Oct. 2024, pp. 431–436.

[115] S. Mohanti, N. Soltani, K. Sankhe, D. Jaisinghani, M. Di Felice, and K. Chowdhury, "AirID: Injecting a custom RF fingerprint for enhanced UAV identification using deep learning," in *Proc. IEEE Global Commun. Conf.*, Dec. 2020, pp. 1–6.

[116] N. Soltani, G. Reus-Muns, B. Salehi, J. Dy, S. Ioannidis, and K. Chowdhury, "RF fingerprinting unmanned aerial vehicles with non-standard transmitter waveforms," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 15518–15531, Dec. 2020.

[117] B. Hua, L. Han, Q. Zhu, C.-X. Wang, K. Mao, J. Bao, H. Chang, and Z. Tang, "Ultra-wideband nonstationary channel modeling for UAV-to-ground communications," *IEEE Trans. Wireless Commun.*, vol. 24, no. 5, pp. 4190–4204, May 2025.

[118] X. Zhou, J. Xiong, X. Zhang, X. Liu, and J. Wei, "A radio anomaly detection algorithm based on modified generative adversarial network," *IEEE Wireless Commun. Lett.*, vol. 10, no. 7, pp. 1552–1556, Jul. 2021.

[119] E. Ayanoglu, K. Davaslioglu, and Y. E. Sagduyu, "Machine learning in NextG networks via generative adversarial networks," *IEEE Trans. Cognit. Commun. Netw.*, vol. 8, no. 2, pp. 480–501, Jun. 2022.

[120] A. Toma, A. Krayani, L. Marcenaro, Y. Gao, and C. S. Regazzoni, "Deep learning for spectrum anomaly detection in cognitive mmWave radios," in *Proc. IEEE 31st Annu. Int. Symp. Pers., Indoor Mobile Radio Commun.*, Aug. 2020, pp. 1–7.

[121] A. Parmar, K. Shah, K. M. Captain, M. López-Benítez, and J. R. Patel, "Gaussian mixture model-based anomaly detection for defense against Byzantine attack in cooperative spectrum sensing," *IEEE Trans. Cognit. Commun. Netw.*, vol. 10, no. 2, pp. 499–509, Apr. 2024.

[122] H. Wang, L. Muñoz-González, D. Eklund, and S. Raza, "Non-IID data re-balancing at IoT edge with peer-to-peer federated learning for anomaly detection," in *Proc. 14th ACM Conf. Secur. Privacy Wireless Mobile Netw.*, New York, NY, USA, Jun. 2021, pp. 153–163, doi: 10.1145/3448300.3467827.

[123] L. Ruff, R. A. Vandermeulen, N. Goernitz, L. Deecke, S. A. Siddiqui, A. Binder, E. Müller, and M. Kloft, "Deep one-class classification," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 4393–4402.

[124] B. Zong, S. Qi, M. R. Min, W. Cheng, C. Lumezanu, D.-K. Cho, and H. Chen, "Deep autoencoding Gaussian mixture model for unsupervised anomaly detection," in *Proc. Int. Conf. Learn. Represent.*, 2018. [Online]. Available: https://openreview.net/forum?id=BJJLHbb0-

[125] J. Kim, H. Kim, and B. Kim, "Wireless anomaly signal dataset (WASD): An open dataset for wireless cellular spectrum monitoring and anomaly detection," *IEEE Access*, vol. 12, pp. 196240–196248, 2024.

[126] G. Cerar, H. Yetgin, B. Bertalanic, and C. Fortuna, "Learning to detect anomalous wireless links in IoT networks," *IEEE Access*, vol. 8, pp. 212130–212155, 2020.

[127] A. S. Ali, W. T. Lunardi, G. Singh, L. Bariah, M. Baddeley, M. A. Lopez, J.-P. Giacalone, and S. Muhaidat, "RF jamming dataset: A wireless spectral scan approach for malicious interference detection," *IEEE Commun. Mag.*, vol. 62, no. 11, pp. 114–120, Nov. 2024.



LJUPCHO MILOSHESKI (Member, IEEE) is currently pursuing the Ph.D. degree with the International Postgraduate School Jožef Stefan. He is a Research Assistant at SensorLab, as part of the Department of Communication Systems, Jožef Stefan Institute, Ljubljana, Slovenia. He is involved in Horizon Europe projects and has co-authored several peer-reviewed conferences and journal publications. His research interests include spectrum sensing, self-supervised learning, and generative AI for wireless networks.



BLAŽ BERTALANIČ (Member, IEEE) received the Ph.D. degree (Hons.) from the Faculty of Electrical Engineering, University of Ljubljana. He is currently a Researcher with SensorLab, Jožef Stefan Institute. His research interests include the advancement of machine learning and AI algorithms, especially in the context of time series analysis and smart infrastructures. He held several leadership positions in Slovenian Chapter with over 15 IEEE publications.



CAROLINA FORTUNA was a Postdoctoral Researcher with Ghent University, Ghent, Belgium. She was a Visiting Researcher at InfoLab, Stanford University, Stanford, CA, USA. She is currently a Research Associate Professor with Jožef Stefan Institute, where she leads SensorLab. She has led and contributed EU-funded projects, such as H2020 NRG5, eWINE, WISHFUL, FP7 CREW, Planetdata, ACTIVE, and under various positions. She has advised/co-advised more than six M.Sc. and Ph.D. students. She has consulted public and private institutions. She has co-authored over 100 articles, including IEEE Communications Surveys and Tutorials, *IEEE Wireless Communications Magazine*, IEEE Open Journal of the Communications Society, and IEEE Access. Her research interests include developing the next generation of smart infrastructures that surround us and improve the quality of our lives. She contributed to community work as a TPC Member, the Track Chair, and a Reviewer at several IEEE conferences, including Globecom and ICC.



MIHAEL MOHORČIČ (Senior Member, IEEE) is currently the Head of the Department of Communication Systems and a Scientific Counselor with Jožef Stefan Institute, as well as a Full Professor at Jožef Stefan International Postgraduate School. His research interests include advanced wireless communication systems, including mobile, satellite, and stratospheric networks; heterogeneous and ad hoc networks; wireless sensor networks; and the Internet of Things. His current research interests include AI-driven resource management in wireless communications, smart infrastructure connectivity, and intelligent sensing applications. He has contributed to over 25 international research projects in mobile and satellite communications, UAV communication systems, and wireless sensor networks, along with more than 15 national basic and applied research projects. He has co-authored more than 230 journals and conference publications, three books, and one patent. He actively serves on conference organizing committees, including as the general chair and the TPC chair.

Chapter 3

Deep Clustering of Wireless Spectrum Activity Based on Spectrogram Processing

Building on the extensive review of self-supervised approaches to radio signal recognition in Chapter 2, this chapter focuses on investigating UDL with an emphasis on deep clustering for RAT recognition. Given the growing body of work and the demonstrated effectiveness of UDL in related domains such as image and audio processing, these methods offer a compelling framework for learning informative signal representations from unlabeled data. The learned representations can subsequently be exploited in downstream tasks relevant to spectrum sensing, including classification and clustering.

Over the past decade, rapid advances in DL model design and training have produced a rich ecosystem of high-quality open-source libraries, providing access to state-of-the-art architectures such as ResNet, VGG, and Vision Transformers (ViT), as well as pretrained models on large benchmark datasets like ImageNet. This has facilitated rapid prototyping and deployment of DL-based solutions across many application domains, including radio signal processing. However, unlike fields such as image recognition, the wireless communications domain remains comparatively less open, with standardized datasets available only for selected tasks, for example, RadioML for automatic modulation classification. Many research directions rely on application- and environment-specific datasets, which limit reproducibility and provide only a partial view of model performance across diverse deployment scenarios. Furthermore, wireless environments are inherently heterogeneous, dynamic and shaped by location-specific propagation conditions, physical obstructions and interference patterns, rendering labeled data both difficult and costly to obtain. This makes the UDL-based learning a suitable approach for feature learning directly from raw data, and does not require explicit knowledge of data or environment specifics, including class annotations.

Deep clustering represents a particularly promising UDL paradigm for wireless applications, as it enables meaningful representation learning without requiring labeled data, a common constraint in real-world spectrum monitoring. This chapter, therefore, investigates the suitability of deep clustering for RAT recognition using fully unlabeled spectrum data, as typically collected from real operating environments. Specifically, we consider spectrum traces from the unlicensed 868 MHz band, where multiple commercial RATs, such as LoRa and IEEE 802.15.4, coexist.

More specifically, the study evaluates the adoption and validation of widely used DL feature extraction architectures, including ResNet, VGG, and ViT, within a deep clustering pipeline, assessed across two closely related tasks: RAT recognition and spectrum activ-

ity detection. The study employs spectrogram representations in the form of 128 × 128 time–frequency segments, which capture energy bursts across the monitored frequencies over time. While such representations encode structurally distinctive patterns characteristic of individual RATs, their information density is significantly lower than that of the RGB images, for which these architectures were originally designed. This necessitates careful adaptation of feature dimensionality and representation learning strategies.

To address this, a detailed feature-space analysis and performance evaluation are conducted across the considered architectures and use cases. The results demonstrate that CNN-based deep clustering, particularly when employing ResNet, offers the most favorable performance–complexity trade-off. It consistently outperforming both a baseline autoencoder-based approach and classical K-means clustering across multiple feature quality metrics and task-level performance measures. In contrast, more complex architectures such as ViT yield only marginal performance improvements while incurring substantially higher computational costs.

This chapter thus supports contribution **C2** by demonstrating the applicability of SSL techniques to wireless spectrum sensing, introducing domain-aware architectural adaptations, and establishing a comprehensive evaluation methodology that integrates complementary quantitative and qualitative perspectives. Furthermore, the empirical findings corroborate Hypothesis **H1**, showing that *UDL architectures can effectively learn discriminative representations from unlabeled RF spectrum data*, provided that appropriate domain-specific adaptations are employed.

The work presented in this chapter was published as:

- L. Milosheski, G. Cerar, B. Bertalanič, C. Fortuna, and M. Mohorčič, "Self-supervised learning for clustering of wireless spectrum activity," Computer Communications, vol. 212, pp. 353–365, 2023.

Computer Communications 212 (2023) 353–365

Contents lists available at ScienceDirect

Computer Communications

journal homepage: www.elsevier.com/locate/comcom

Self-supervised learning for clustering of wireless spectrum activity

Ljupcho Milosheski [a,b,*], Gregor Cerar [a], Blaž Bertalanič [a], Carolina Fortuna [a], Mihael Mohorčič [a,b]

[a] Jozef Stefan Institute, Department of Communication Technologies, Jamova 39, 1000, Ljubljana, Slovenia
[b] Jozef Stefan International Postgraduate School, Jamova 39, 1000, Ljubljana, Slovenia

ARTICLE INFO

MSC:
90B20

Keywords:
Spectrum
Analysis
Clustering
Self-supervised
Machine learning

ABSTRACT

In recent years, much work has been done on processing of wireless spectrum data involving machine learning techniques in domain-related problems for cognitive radio networks, such as anomaly detection, modulation classification, technology classification and device fingerprinting. Most of the solutions are based on labeled data, created in a controlled manner and processed with supervised learning approaches. However, spectrum data measured in real-world environment is highly nondeterministic, making its labeling a laborious and expensive process, requiring domain expertise, thus being one of the main drawbacks of using supervised learning approaches in this domain. In this paper, we investigate the utilization of self-supervised learning (SSL) for exploring spectrum activities in a real-world unlabeled data. In particular, we assess the performance of SSL models, based on the reference DeepCluster architecture. We carefully consider the current state-of-the-art feature extractors, taking into account the performance and complexity trade-offs. Our findings demonstrate that SSL models achieve superior performance regarding the feature quality and clustering performance compared to baseline feature learning approaches. With SSL models we achieve significant reduction of the feature vectors size by two orders of magnitude, while improving the performance by a factor ranging from 2 to 2.5 across the evaluation metrics, supported by visual assessment. Furthermore, we showcase how adapting the reference SSL architecture to domain-specific data is followed by a substantial reduction in model complexity up to one order of magnitude, without compromising, and in some cases, even improving the clustering performance.

1. Introduction

The number and type of wireless devices connected to the Internet is rapidly increasing with the current affordable personal mobile and Internet of Things (IoT) devices, requiring wireless networks to handle large number of connections and high traffic loads. As a reference, the requirement for the number of connected devices in the fifth-generation (5G) networks is one million devices per square kilometer. The existence of such a number of devices requires complex wireless resource management. Over time, several new approaches to wireless resource sharing, including dynamic spectrum access [1], licensed shared access [2] have been proposed. However, additional technological components, such as spectrum usage databases [3] and radio environment maps [4], had to be developed to enable such sophisticated and dynamic spectrum usage approaches. To be able to correctly inform on spectrum usage, additional knowledge of other devices operating within the range of a wireless device is critical for future smart usage of the spectrum. In this respect, some of the recent efforts were focused on detecting the modulation used [5], technology used [6], anomalous activities [7], etc.

As also discussed in [8] and [9], significant effort is still being invested in the field to develop accurate and scalable deep learning (DL) algorithms able to accurately and automatically manage spectrum resource usage. With respect to the learning approach, these techniques can be divided into (1) *supervised* that require labels to be present for the training data, and (2) *unsupervised* that do not assume any such labels. Applications in wireless spectrum management need to be aware of operating details (i.e., type of technology, transmission parameters, etc.). Development of a DL-based model to support such application typically requires labeled data that is expensive to acquire as it requires complex wireless and computing equipment [10,11] or intense labeling efforts by domain experts [12] that do not always lead to high quality labels due to the nondeterministic nature of wireless operating environments. *Semi-supervised* and *active-learning* emerged as alternative techniques that have the advantage of using a relatively small amount of labeled samples for achieving performance that is comparable to the regular supervised approach.

Given the advent of large datasets which are expensive or practically impossible to label, *self-supervised learning* (SSL) [13], as another

* Corresponding author at: Jozef Stefan Institute, Department of Communication Technologies, Jamova 39, 1000, Ljubljana, Slovenia.
E-mail address: ljupcho.milosheski@ijs.si (L. Milosheski).

https://doi.org/10.1016/j.comcom.2023.10.009
Received 17 May 2023; Received in revised form 23 September 2023; Accepted 13 October 2023
Available online 16 October 2023
0140-3664/© 2023 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

intermediate learning approach, is becoming an important alternative that is particularly suitable to reduce the data labeling cost and leverage the unlabeled data pool. SSL is a representation learning method where a supervised task is created out of the unlabeled data. Using an SSL approach, it is possible to create very similar groups (i.e. clusters) from a large, unlabeled dataset and then label each cluster. By labeling the learnt clusters, it is possible to then use the model as a classifier by assigning new, unseen examples to those clusters and therefore label them as one would do in a typical classification task.

Developing an easy to use, automated and technology agnostic way to explore spectrum activities and group similar activities, eventually enabling automatic rather than manual transmission identification and cataloging as currently done for instance in the Signal Identification Guide,[1] is still an open research topic which motivated this investigation.

Employing DL architectures based on convolutional neural networks (CNN) allows for direct processing and classification of the raw In-phase and Quadrature (I/Q) time series [14]. While for modulation classification [5] and device fingerprinting [15], direct processing on the raw I/Q time series data is desirable, for radio access technology classification it is better to employ spectrograms due to lower complexity and robustness towards low SNR conditions [6]. Additionally, the image-like (2D matrices) format of the spectrograms provide the possibility to utilize some state-of-the-art architectures from the closely related machine vision domain, such as Vision Transformers (ViT) [16] that recently emerged as competitors to the CNN-based solutions. Inter domain adaptations, such as in [17], could additionally stimulate the development of spectrum sensing technologies which are increasingly important for the next generation radio networks.

In this paper, we investigate the suitability of SSL to support automatic spectrum exploration on an example of an unlicensed 868 MHz Short Range Device (SRD) band in an urban environment using 15 days of spectrum sweeps collected in the LOG-a-TEC[2] wireless testbed. Leveraging this data, we propose an SSL architecture adapted for spectrum activity identification and clustering, in which segments of spectrograms containing signal activity are used to train the DL self-supervised network and enable the discovery of the types of transmissions available over the respective period of time. It is based on machine vision and inspired by DeepCluster [18], which is also used in this study as a reference SSL model. In this architecture, we experiment with CNN-based and ViT-based feature extraction and prove that such an architecture is suitable for spectrogram analysis by learning spectrum features and clustering spectrograms based on their content.

The main contributions of this work can be summarized as follows:

- Adaptation of an SSL architecture to discover wireless transmissions in real-world spectrogram data when no prior knowledge (i.e. labels) is available, while also achieving significant reduction of the complexity of the architecture with up to 10 times less trainable parameters compared to the original implementation.
- Proposing dimensionality reduction of the embedding space by principal component analysis (PCA) using a threshold on the amount of explained variance ratio (EVR), achieving features quality and clustering performance improvement by a factor of 2–2.5 across the selected evaluation metrics.
- Development of a methodology for quantitative and qualitative evaluation of the transmissions clustering from raw spectrograms. The methodology consists of features quality assessment in the embedded space and clustering evaluation with selected metrics, verified by data-specific visualizations.
- Experimental evaluation of the adapted SSL architecture for two use cases, activity detection and fine grain transmissions classification.
- Comparison of two feature extraction modules in the adapted SSL architecture, one based on ViT, considered as more powerful in RGB image classification tasks, and CNN, that proved more suitable for low-content data such as spectrograms, taking into account the performance and complexity trade-offs.

The rest of the paper is structured as follows. Section 2 analyzes the related work, Section 3 introduces SSL and baseline system architectures, Section 4 elaborates on the experimental methodology and Section 5 presents the experimental results. Finally, Section 6 concludes the paper.

2. Related work

In recent years, as in many other research areas the use of DL models gained a lot of attention also in the development of algorithms for processing spectrum data. Selected works from the domain which are considered as most relevant and closely related to this paper are listed in Table 1, which summarizes their main characteristics, and briefly elaborate in the following where they are grouped according to the adopted learning approach.

2.1. Supervised learning

In the existing works, *supervised* DL-based models are most widely represented and they achieve significant increase of performance when compared to more traditional ML approaches, as shown in [6,14,20].

In [6], the authors compare the performance and generalization ability of models that use manually extracted expert features with models that use raw spectrum data. They are solving the task of technology classification on a dataset containing transmissions of three different wireless technologies. They prove that using CNN on raw I/Q data or spectrogram images outperforms all other models in terms of accuracy, generalization ability to unseen datasets from different operating environments, and robustness to different noise levels.

In [20], an application of CNN supervised learning for device identification, again using raw I/Q data, is proposed. The dataset contains transmissions from five devices. Device identification is based on the CNN's ability to learn various device-specific impairments in the raw signals. SVM and logistic regression are used as reference models for performance comparison and authors show that the proposed CNN model significantly outperforms the baseline algorithms for the posed task.

In many cases, state-of-the-art accuracy is achieved by adopting and modifying DL-based architectures that are already known and well established in other, closely related signal processing fields, such as image and sound processing [7,19]. In [15], device classification task using real-world I/Q data of transmissions from a large population of nearly 10,000 devices is solved using a new neural network architecture based on dilated causal convolutional layers. The design is motivated by an existing audio signals processing architecture.

2.2. Semi-supervised and transfer learning

Although the supervised models have superior performance judging by their accuracy, the necessity for large labeled training datasets as one of the major downsides of this approach remains. One way of going around this problem is to train feature extractor in unsupervised manner and then tune the classification on a small labeled dataset. This approach is employed in [22] to solve the problem of anomaly detection in wireless communications.

Another approach for the labeling problem is employing transfer learning [17,19]. In [17], object detection "You Only Look Once" (YOLO) model, pretrained on the ImageNet dataset, is tuned to detect and classify different types of transmissions using spectrograms. It is shown that the proposed model performs well in classifying interfering signals on simulated data with different signal-to-noise ratios and

[1] https://www.sigidwiki.com/wiki/Signal_Identification_Guide
[2] https://log-a-tec.eu/datasets.html

Table 1
Related works with main characteristics.

Publi-cation	Problem type	Architec-ture	Data type	Dataset	Conti-nuous sensing (yes/no)	Labels (yes/no)	Approach	Band
[5]	Modulation classification *(multiclass)*	LSTM	I/Q, PSD	RadioML, Electrosense	Yes	Yes	Supervised	174–230 MHZ, 470–862 MHZ, 25–1300 MHZ
[19]	Device fingerprinting *(multiclass)*	DCC	I/Q	DARPA, Synthetic	/	Yes	Supervised	2.4 GHz, 5 GHz, 978 MHZ, 1090 MHZ
[20]	Device fingerprinting *(multiclass)*	CNN	I/Q	Test-bed	/	Yes	Supervised	5 GHz
[6]	Technology classification *(multiclass)*	DT, FNN, CNN	RSSI, I/Q, Spectrograms	Ghent, Dublin	Yes	Yes	Supervised	5540 MHZ, 2412 MHZ, 806 MHZ
[14]	Modulation classification *(multiclass)*	CNN	I/Q	Test-bed, synthetic	/	Yes	Supervised	900 MHZ
[17]	Technology characterization through object detection *(multiclass)*	YOLO CNN	Spectrograms	Test-bed, Ghent	Yes	Yes	Supervised	5540 MHZ, 2412 MHZ, 806 MHZ
[21]	Device fingerprinting with clustering *(multiclass)*	CNN	I/Q	Test-bed	/	Yes	Semi-supervised	0.25–1.25 MHZ, 1.67 MHZ, 2.5 MHZ
[15]	Device fingerprinting *(multiclass)*	DCC + CNN	I/Q	DARPA	/	Yes	Supervised, Unsupervised	2.4 GHz, 5 GHz, 978 MHZ, 1090 MHZ
[7]	Anomaly detection *(binary)*	PredNet-autoencoder	Spectrograms	Synthetic	/	No	Unsupervised*	/
[22]	Anomaly detection *(binary)*	Supervised CNN-based	PSD	Synthetic, HackRF SDR, Electrosense	Yes	Yes	Unsupervised	10 MHZ–3 GHz
[12]	Transmissions detection and classification *(binary)*	Image processing	Spectrograms	Log-A-Tec	Yes	Yes	Unsupervised	868 MHZ
Ours	Transmissions clustering *(multiclass)*	Deep-clustering	Spectrograms	Log-A-Tec	Yes	No	Unsupervised (Self-supervised)	868 MHZ

provides additional information about the position of the transmission events in the spectrum.

Although the semi-supervised and transfer learning are proven to lower the amount of data that is required for the training, labeled data is still needed for the tuning. Spectrum data content is not as universal as the RGB images and it is much more dependent on the operating environment regarding the noise, signal strength, fading, multipath effects, etc., which makes it necessary to provide labels for each specific radio environment.

2.3. Unsupervised learning

Considering the availability of large amounts of unlabeled spectrum data [10,11], another approach to the labeling problem with high potential, which we consider as under-explored in the spectrum data processing domain, is using unsupervised models. We believe that the development of unsupervised approaches could more effectively solve the labeling problems and provide highly automated architectures that can be adapted to different operating environments with minimum expert intervention.

There are existing efforts in this direction, but they only consider the marginal cases of event detection (binary classification). A pipeline for automatic detection of wireless transmissions using classic image processing techniques is proposed in [12]. Another such example is the work in [23] where auto-encoder is utilized as feature learner for anomaly detection and shown to outperform the robust PCA.

Some recent works employ contrastive learning as an SSL representation learning approach, for instance on I/Q samples [24] and motion sensor data [25]. Although they show promising performance, this approach is more suitable for datasets that have clear distinction between positive and negative samples, while also requiring data-specific transformations for training.

In our work, we align with the efforts of developing completely unsupervised models that would provide feature learning in an automated manner. We develop a self-supervised architecture that can provide the functionality of simple activity detection, but also enable fine grain classification of wireless transmissions of different technologies in wireless spectrum data.

3. Architectures of the SSL and baseline systems

In this paper we investigate the suitability of SSL deep clustering models for automatic learning of relevant features and clustering wireless spectrum activities identified from spectrogram segments. These activities represent various transmission events, each characterized by distinct shapes and signal intensities. The learning process relies on an iterative refinement of the feature extraction parameters, in our case in modules based on CNN or ViT, and clustering assignments which are updated in alternating manner. No prior knowledge is assumed about the types and number of different spectrum activities present in the dataset or their specific visual appearance in the spectrograms. Since the spectrograms are 2D data, similar to gray-scale images, we follow the same ideas as in [18]. We conduct a comparative analysis of Self-Supervised Learning (SSL) models against two distinct baseline approaches for feature learning. The first baseline method employs PCA as a fundamental unsupervised technique for representation learning. The second baseline utilizes a CNN-based auto-encoder, which represents a more advanced unsupervised DL approach in our evaluation.

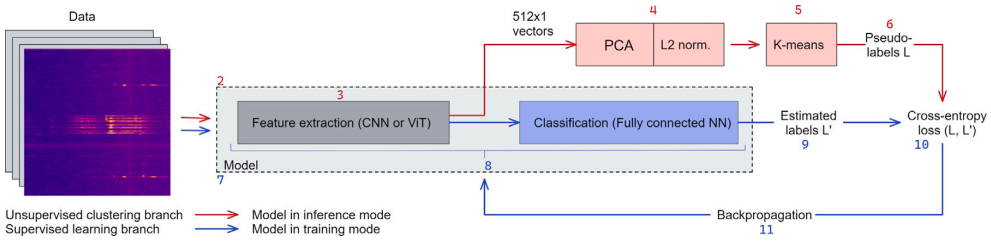


Fig. 1. Architecture of the self-supervised CNN system.

3.1. SSL architectures

As a reference SSL system we adopted the existing self-supervised DeepCluster architecture proposed in [18] for RGB image feature learning based on a VGG (Visual Geometry Group) standard deep Convolutional Neural Network (CNN) architecture with batch normalization that is regularly used in computer vision applications. This system alternates between clustering the image features produced by CNN and updating its weights by predicting the cluster assignments.

Inspired by the approach of this reference system, we developed a SSL system depicted in Fig. 1, which is actually an adaptation of DeepCluster for spectrum activity identification and clustering from spectrograms, characterized by much less content compared to RGB images, thus allowing for significant complexity reduction and performance optimization. As it can be seen from the figure, the system design contains two branches, (1) a so-called unsupervised branch depicted with red arrows and (2) a supervised branch depicted with blue arrows. Both branches contain a feature extractor. In the unsupervised branch it is followed by a dimensionality reduction (PCA) and normalization (L2) preprocessing block and concluded with the K-means clustering block. In the supervised branch, the feature extractor is followed by fully connected layers as a classifier. Note that even if we refer to the second branch as supervised, it does not rely on actual labeled data. It relies on pseudo-labels that are repeatedly generated by K-means in unsupervised branch and refined through feedback, i.e. backpropagation. By working in tandem, the two branches realize self-supervision.

Compared to the original DeepCluster architecture, we propose the following variations:

Alternative CNN. We want to better understand the discriminative power of CNNs which should provide feature vectors in the embedding space that are correlated to the transmission shapes, and utilize clustering algorithm in the embedding space that will provide pseudo labels for CNN's parameters optimization, together resulting in iterative learning and grouping of the samples.

Motivated by the findings in [14], we also select ResNet architecture as a possible feature extraction architecture that can replace VGG in the DeepCluster architecture depicted in Fig. 1 as it was shown to perform better in a supervised task of modulation classification. As we show later, with ResNet we achieve very similar performance as with VGG at significant complexity reduction of roughly 10 times less trainable parameters. From the ResNet family we use ResNet18 (18 showing the number of convolution layers in CNN) in its original form, but with input and output layers customized according to the shape of the images and the number of classes, i.e., a single input channel in the case of spectrogram images compared to a 3-channel input required for RGB images that ResNet18 was originally designed for.

We also added PCA feature space reduction before clustering to adapt to the significantly lower amount of content in spectrograms compared to RGB images.

Transformers. A more recent and very powerful architecture in natural language processing and machine vision is represented by transformers, therefore we also consider ViT in the SSL architecture, as it is proven [26] to be powerful feature extraction module in the machine vision domain.

ViT-based models are devised by simple replacement of the feature extraction and classification modules of the SSL architecture with transformer architecture, which includes projection head (classification module) itself. In such model, features from the ViT embedded space undergo the same process of PCA-transformation and L2 normalization before being clustered by the K-means.

Further in the paper, we used the following template notation AA-FE-X where AA refers to the architectural approach, and can have two values: SSL or B from baseline. FE stands for feature extractor and can have values VGG, RN from ResNet, ViT, AE, PCA while X represents the number of PCA components used for the model instance development. For example, we refer to the VGG-based and ResNet-based self-supervised models with 20 PCA components as SSL-VGG-20 and SSL-RN-20, respectively.

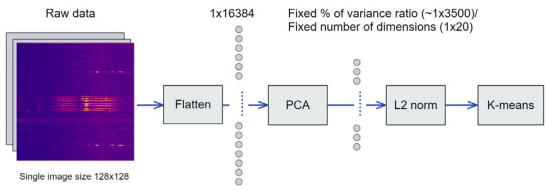
The pseudo-code corresponding to the workflow of the proposed self-supervised system is given in Algorithm 1. For clarity and clear mapping, the blocks of the system in Fig. 1 are marked with the corresponding line numbers from the pseudo-code in Algorithm 1.

3.1.1. Workflow of the unsupervised branch

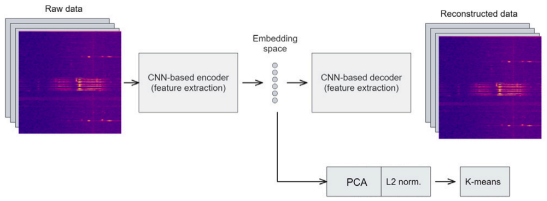
In the initial phase, the CNN/ViT is initialized randomly and the input data, which consists of image-like spectrum segments, is unlabeled. The clustering algorithm (unsupervised branch, marked with red arrows in Fig. 1) is used to cluster the features and provide pseudo-labels (denoted by L in the figure) at the beginning of each training epoch. Initialization of the cluster centers is random. The features are extracted using the feature extraction module, which has the size of 512 × 1 for the CNN and 1024 × 1 for the ViT (lines 2-3 in Algorithm 1). Thus, a descriptor in the form of 512 × 1 or 1024 × 1 vector is obtained for each spectrum segment image. These descriptors are then PCA-reduced, L2-normalized and finally clustered (lines 4-5 in Algorithm 1).

Algorithm 1 Workflow of the architecture

1: **while** $Iteration < NumberOfEpochs$ **do**
   Unsupervised branch:
2: $Model \leftarrow EvaluationMode$
3: $ExtractFeatures$ ▷ CNN or ViT
4: $ProcessFeatures$ ▷ PCA + L2 norm.
5: $ClusterFeatures$ ▷ K-Means
6: $(L)Pseudo-labels \leftarrow ClusterAssignments$
   Supervised branch:
7: $Model \leftarrow TrainingMode$
8: $EstimateLabels$
9: $L' \leftarrow EstimatedLabels$
10: $Cross-entropy(L, L')$
11: $Backpropagation$ ▷ Update classifier weights
12: **end while**



(a) PCA baseline system.



(b) Auto-encoder baseline system.

Fig. 2. Architectures of the baseline systems.

3.1.2. Working of the supervised branch

In this phase, the feature extraction and classification modules are used in training mode, as a supervised classification model. The flow of this pipeline activity is indicated by the blue arrows in Fig. 1. Using the provided cluster assignments from the previous step as pseudo-labels (L) for the input images, model is trained for one epoch. This completes one iteration of the entire pipeline work cycle.

The procedure stops when the predefined number of iterations (training epochs) is reached. In our experiments, we used 200 training epochs. This number was determined empirically by observing the convergence of the loss function.

3.2. Baseline systems

Considering that the most influential part of the architecture is the feature learning component, we take PCA-based representation learning as a simple, yet informative baseline for comparison with the SSL models complemented by a K-means clustering algorithm, as depicted in Fig. 2(a). The system consists of a flattening block, PCA-dimensionality reduction and L2 normalization for input data preprocessing, followed by the K-means clustering block.

The flattening reorders the elements of the input matrix into a single row for each data sample. On the vectors provided by the flattening, the same sequence of operations is applied, as for the vectors provided by the self-supervised system. PCA is applied on the flattened data and as a result of this operation, reduced feature vectors representing the input data are obtained. These vectors are then L2-normalized and finally used as an input to the K-means clustering algorithm. In the rest of the paper we refer to baseline models as B-PCA.

As a second, more comprehensive baseline algorithm for comparison we utilize a symmetric CNN-based auto-encoder (AE) architecture realized by the same ResNet18 module, which was proposed as feature extractor in the SSL architecture. The visualization of the system is depicted in Fig. 2(b). In the same manner as for the SSL models, PCA is being applied on the embedding space feature vectors to provide reduced dimensionality space, which is later L2 normalized and than fed to the clustering algorithm. Thus we provide stronger baseline model for comparison with similar feature extraction module as the proposed SSL. The training, same as for the SSL model, was performed for 200 epochs.

4. Methodology

In view of developing and benchmarking unsupervised spectrum activity exploration we define the methodological approach to prepare the raw data used to train the proposed self-supervised and baseline models, and to develop and evaluate high quality feature extraction, clustering and resulting models.

4.1. Raw training data preparation

The dataset used for the analysis consists of fifteen days of spectrum measurements acquired at a sampling rate of 5 power spectrum density (PSD) measurements per second using 1024 FFT bins in the 868 MHz license-free (shared spectrum) SRD band with a 192 kHz bandwidth. The data is acquired in the LOG-a-TEC testbed. Details of the acquisition process and a subset of the data can be found in [11]. The acquired data has a matrix form of $1024 \times N$, where N is the number of measurements over time. By windowing the data with a window size W, the resulting raw images for training would have $1024 \times W$ dimension which can be computationally untractable for less capable computing platforms. Therefore, in addition to windowing in time, we are also windowing the dataset in frequency (i.e., FFT bins) direction.

The segmentation of the complete data-matrix into non-overlapping square images along time and frequency (FFT bins) is realized for a window size $W = 128$. An example of such segmentation containing 8 square images is shown in Fig. 3, corresponding to image resolution of 25.6 s (128 measurements taken at 5 measurements per second) by 24 kHz. The window size is chosen to be large enough to contain any single type of activity and small enough to avoid having too many activities in a single image while also having in mind computational cost. Dividing the entire dataset of 15 days using $W = 128$ and zero overlapping, produces 423,904 images of 128 × 128 pixels, where the pixel values are scaled to the range between 0 and 1.

From previous work [12] we obtained labels of part of the transmissions in the employed dataset provided by experts. The labels are bounding boxes around transmissions in the same spectrogram data. We adapt these labels to our use case, considering the data is segmented in square-shaped regions. Each label as a rectangular region in the data matrix covers at least one square segment. All of the square segments that are overlapping with the labeled regions are marked as active, i.e., containing transmission(s). Thus we obtained labeled subset of square-shaped spectrograms that contain transmissions, as marked by experts, which have been used for the subsequent evaluation of the compared models.

The VGG-based model also requires adaptation on the dataset itself. In fact, using the described data segmentation in the SSL-VGG model, which was originally designed for input data size of 224 × 224 pixels, leads to converging the model to create only one cluster and zero valued feature vectors. To avoid this, we upscale the 128 × 128 spectrograms to the originally required input size for SSL-VGG. Another issue with SSL-VGG model is the number of training epochs. Running for too many (e.g., more than 40) epochs again leads to the behavior explained above. The number of epochs that provides meaningful results with the SSL-VGG model was experimentally determined to be around 30, which is notably less than the number of epochs required for the SSL-RN model. We assume that the reason for such behavior is the low amount of content in the spectrograms compared to the content-rich RGB images that the SSL-VGG architecture was originally designed for. On the other side, the other two SSL models (SSL-RN and the SSL-ViT) along with the baseline AE model do not require additional data adaptation and consistently improve their performance during training.

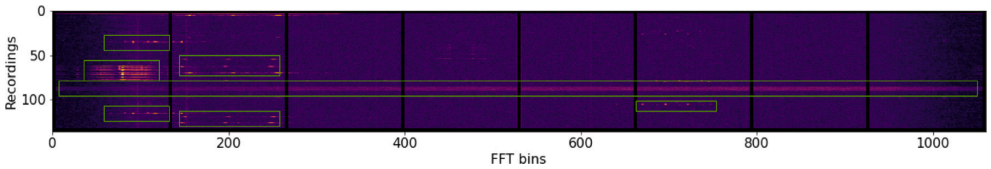


Fig. 3. Sample of 8 spectrogram segments from the data.

4.2. Feature development and evaluation methodology

As shown in Fig. 1,the segmented and normalized raw data is passed through a feature extractor (ResNet18) and PCA/L2 preprocessing, before being actually clustered. These blocks, together, learn to engineer features. The better the quality of the engineered features that train the K-means clustering algorithm, the better the resulting clustering model. Therefore, in the process of developing the proposed self-supervised model, the feature development process needs to be tuned and evaluated.

While the feature extractor is automatically trained, the PCA performing dimensionality reduction for both the proposed self-supervised and the baseline system needs to be configured. We first analyze how many PCA components are needed to keep the most relevant information while discarding as much noise as possible. The number of dimensions D used in the PCA-reduced dimensionality representation is determined by setting a threshold on the amount of Explained Variance Ratio (EVR). We assume that the important information is encoded in much smaller number of dimensions (i.e., between 10 and 20 out of 512) because, contrary to the RGB images from the reference application of SSL model [18], the spectrograms contain much smaller amount of content. Visual analysis of the spectrograms shows that the transmission bursts surface is varying relative to the spectrogram size, but it is not bigger than around 10%, i.e., 1600 pixels of the spectrogram relative to its size of $128 \times 128 = 16384$ pixels. The EVR threshold is determined by exploring the plot of the cumulative sum of EVR of the PCA-transformed space dimensions.

Clusterability of developed features can be evaluated with different metrics such as the Hopkins score [27] and visual assessment of clustering tendency (VAT) [28]. The Hopkins score is a metric that shows the probability that randomly sampled subset of the data comes from a uniform distribution. Resulting values are between 0 and 1, with 0 meaning the data is not uniformly distributed. However, having nonuniform distribution does not guarantee existence of clusters in the data. One such case is if the data has normal distribution, which will show low Hopkins score but will not contain any meaningful clusters. To prevent possible false conclusions derived based on this metric alone, we also use the evaluation with the VAT algorithm, which produces matrix visualization of the dissimilarity of samples based on their pairwise euclidean distances. The value of each element of the visualization matrix is proportional to the pairwise dissimilarity between each of the samples to all of the other samples. Thus, the left diagonal of the matrix is with zero values representing dissimilarity of each sample to itself. The samples are ordered in such a way that groups that are located close in the feature space, according to the distance metric, appear as dark squares along the diagonal of the matrix. Implementation wise, an improved version of the VAT (iVAT) is used, which provides better visualization than the standard one.

In the evaluation process, the Hopkins score and the VAT plots are complementary and will be considered both at once, since VAT does not provide quantification, while the Hopkins score alone only gives an information about the uniformity of data distribution and not on its clusterability.

4.3. Cluster development and evaluation methodology

To develop the cluster model that best finds similar transmissions, the compared architectures have to be trained several times for different values of k in K-means resulting in as many models. These models then need to be evaluated to see which one contains the cleanest clusters, i.e. provides clear data separation. As all the possible transmissions that may occur are not known in advance, we choose $k \in [2, 3, 4, \ldots, 30]$.

The clustering is then evaluated with two standard metrics, the Silhouette score [29] and the Davies–Bouldin index [30], and also manually by cluster analysis and explanation.

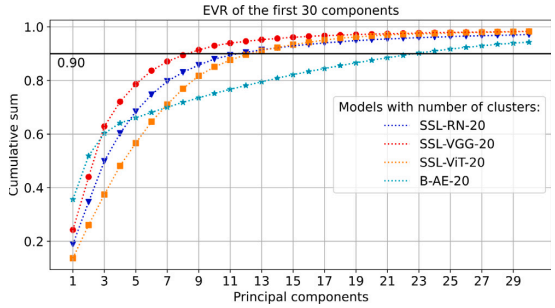
The Silhouette score is a value that measures the quality of the clustering by evaluating how similar is each sample to its assigned cluster. The values it takes are in the range of $[-1, 1]$, where bigger values mean better clustering. Silhouette score of a cluster is the average of the scores of its elements, while Silhouette score of a clustering is the average of the scores of all of the clusters.

The Davies–Bouldin index is a clustering quality metric calculated as average of the similarity value of each cluster to its most similar (closest) cluster. The range of values that this metric takes has only lower bound 0 and smaller values mean better clustering.

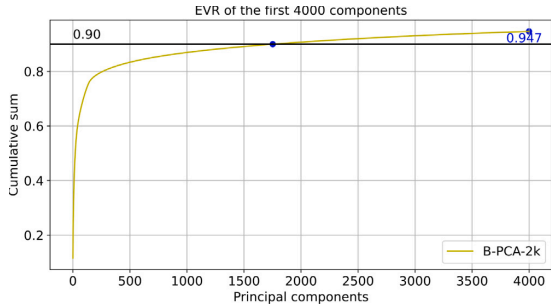
As discussed in Section 3, the clustering algorithm is randomly initialized, so labels are not fixed to any specific type of content. Thus, besides the quantitative evaluation of the quality of the clustering results, manual evaluation is also performed in order to interpret the content that is specific for each cluster together with the frequency that is specific for the content. The motivation for the manual inspection is to verify that the clusters formed as output of the K-means are meaningful subsets of spectrograms containing highly correlated types of spectrum patterns (transmissions). Evaluation is made on plots of averaged spectrograms and histograms of frequency locations specific for the cluster's samples.

4.4. Evaluation with labeled data

After manual evaluation we can assign specific type of transmission to each of the formed clusters. Also, it is justified to expect that there will be clusters with empty/no transmission segments since the dataset is created by uninterrupted measurements from certain time period and it is hardly possible that each spectrogram will contain transmission. This is verified by visualization of sample data. Knowing this, we evaluate the models by utilizing the created labeled data, as explained in Section 4.1, and check how well the models are performing for the task of activity detection. The evaluation is made by running the labeled spectrograms containing transmission through the trained models and calculate the percentage of them that were correctly distributed across the clusters that contain transmissions, marked in the previous step of manual evaluation. Correct classifications would be labeled samples assigned to any of the transmissions-rich clusters and misclassifications would be labeled samples assigned to the empty/no activity clusters.



(a) PCA on CNN extracted features.



(b) Direct PCA on images for the baseline model.

Fig. 4. Cumulative sum of EVR of the PCA-transformed features.

5. Experimental results

We started the performance evaluation with the analysis of the feature space of a few experimentally trained models to provide an insight about the content dynamics of the specific range of PCA components, before we continued with the investigation of the number of clusters in the dataset, and set the number of PCA components to be used in clustering. Using the methodology proposed in previous section, we carried out a number of experiments to compare the performance of SSL and baseline models, and eventually we showcased the suitability of SSL models for two representative use cases.

5.1. Effect of the dimensionality reduction

For configuring the PCA-based dimensionality reduction of the proposed self-supervised and baseline models, we first explored the cumulative sum of the EVR of the features, as shown in Fig. 4, for the different models. Fig. 4(a) plots EVR of the three SSL models and the baseline AE model. It can be seen that 90% of the EVR is contained in at least 17 components across the SSL models and that there is no significant change in the EVR when considering more than 20 components. Therefore we selected the first 20 PCA components as the size of the vectors encoding the input images, which are also input for the K-means algorithm in the unsupervised branch in Fig. 1. The baseline AE model shows comparable performance by encoding the same EVR of 90% in 23 PCA components. In our experimental investigations, we discovered an intriguing result regarding the AE model in conjunction with PCA dimensions. Specifically, we found that there is an almost negligible difference in the evaluation metrics when using either 20 or 25 PCA dimensions. To facilitate more convenient and equitable comparisons in our subsequent experiments, we have opted to employ a 20-dimensional PCA-transformed embedding space also for the AE model, denoted as B-AE-20, aligning it with the SSL models.

The plot in Fig. 4(b) represents the feature space for the simple, PCA baseline model. It can be seen that direct PCA on images requires 1752 components to keep 90% EVR independent of the final number of clusters. Taking this into consideration we used two baseline models, one with 1752 and one with 20 (i.e. the same as with SSL models) PCA components, and we refer to them in the following as B-PCA-2k and B-PCA-20, respectively.

Comparison of results in Figs. 4(a) and 4(b) shows that SSL and AE models are able to learn to encode the relevant information for cluster development in roughly 1% of the components required by the baseline PCA model for the same 90% EVR. Thus, the SSL-based and the AE-based feature extraction provide significant simplification of further processing because of the dimensionality reduction by two orders of magnitude.

5.2. Clustering tendency of the features

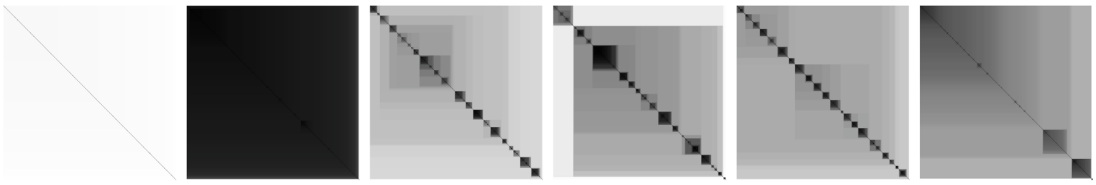
Using the features from the models derived in the previous section, which are the inputs to the K-means algorithm, we perform quantitative evaluation of how suitable are the resulting features for clustering.

The Hopkins scores evaluating the quality of the features using the SSL versus the baseline models are presented in Fig. 6 for 5, 10, 15, 20 , 25 and 30 clusters. We also plot fitted curves for each model to emphasize the general trends. For the baseline models B-PCA-20, B-PCA-2k and B-AE-20, the clusterability is independent of the number of clusters so we expect to have constant values. The small variations of the corresponding values are due to random initialization of the Hopkins calculation procedure. This is not the case for the SSL models since the number of clusters is always the same as the number of classes of the CNN output, thus affecting the overall architecture.

The SSL-VGG-20 model has consistently the best performance across the different number of clusters. The scores of the proposed SSL-RN-20 model are close to the B-PCA-20 model small number of clusters, and they improve as the number of clusters becomes larger than 10 where they are close to the scores of the SSL-ViT-20 model.

For a more comprehensive evaluation, we opted to focus on SSL models configured with 23 clusters, a choice that aligns with the optimal range for SSL architectures. Notably, this cluster count variation had no discernible impact on the baseline models. Subsequently, we conducted an analysis of VAT plots for both the selected SSL and baseline models, as depicted in Fig. 5. Fig. 5(a) corresponding to B-PCA-2k contains no grouping at all, meaning that the feature vectors provided by this model are randomly distributed in the feature space. Fig. 5(b) corresponding to B-PCA-20 shows that there is only a small grouping (outlayers probably) in the lower right corner, which justifies the better performance compared to the B-PCA-2k on the Hopkins evaluation. Another observation is that almost the entire matrix is dark, meaning that all of the samples are closely correlated, which can be explained by the 20 dimensions of the PCA-transformed feature space containing less than 50% of the EVR. The distances between the samples are thus much smaller compared to the SSL models.

Fig. 5(c) shows the clustering tendency of the features obtained with SSL-RN-20. There are around 20 distinguishable squares corresponding to groups of highly correlated samples. This confirms good scores on the Hopkins evaluation achieved by SSL-RN-20 models with 20 to 25 clusters. For the SSL-VGG-20 in Fig. 5(d), the feature space contains three well separated, highly correlated groups of features and comparable size and number of darker squares as for the SSL-RN-20. The three well separated groups are the reason for the consistently better Hopkins scores of the SSL-VGG-20 compared to others, considering the Hopkins's definition. However, the number and size of visually distinguishable dark squares of the SSL-VGG-20 is comparable with the SSL-RN-20, suggesting that similar number of clusters could be existing in both feature spaces. Regarding the SSL-ViT-20 in Fig. 5(e), there are also approximately 20 clusters, but with very similar sizes. This suggests that the ViT may either be able to focus on other specifics



(a) B-PCA-2k (b) B-PCA-20 (c) SSL-RN-20 (d) SSL-VGG-20 (e) SSL-ViT-20 (f) B-AE-20

Fig. 5. VAT plots of the different feature vectors.

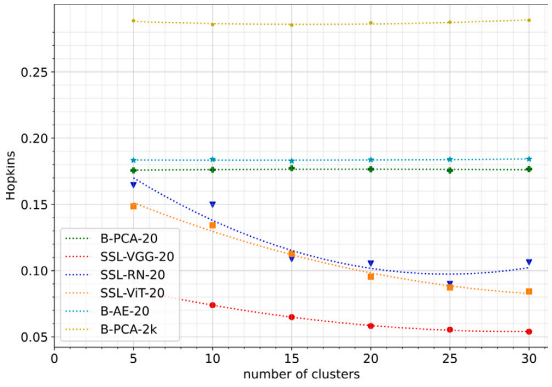
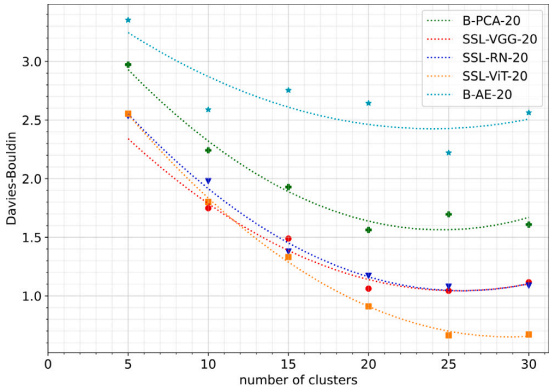
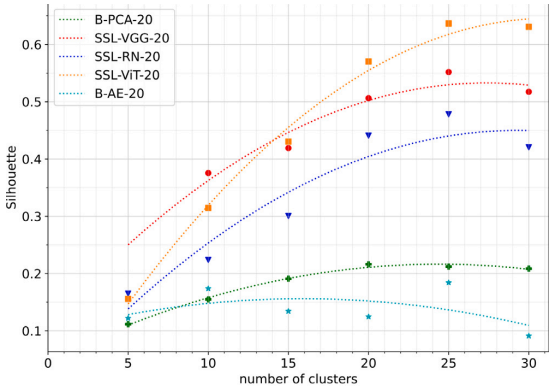


Fig. 6. Hopkins scores of the extracted features.



(a) Davies-Bouldin scores.



(b) Silhouette scores.

Fig. 7. Evaluation curves of metrics for quality of clustering.

of the signals compared to CNNs or could be forcing the separation of the feature space by the defined number of clusters, regardless of the actual distribution.

Finally, B-AE-20 has clustering tendency scores somewhere between the SSL models and the simple PCA baseline models. There are three distinguishable clusters on the VAT plot, which is much lower than in the case of the SSL models. Considering the size of the squares, they seem to be more related to the general background features than the actual transmission patterns.

From this section we can conclude that the PCA-only baseline models are not capable of encoding significant number of groups of correlated features from the input spectrograms. Following this conclusion, we only used the B-PCA-20 for further comparisons, since the B-PCA-2k feature vectors showed no clustering tendency at all, making their further evaluation irrelevant.

5.3. Evaluation of clusters quality

Evaluating the clustering quality is performed according to the methodology described in Section 4.3. For the Davies–Bouldin scores plotted in Fig. 7(a) for 5, 10, 15, 20 , 25 and 30 clusters, it can be seen that the SSL models, i.e. SSL-RN-20, SSL-VGG-20 and SSL-ViT-20 have consistently better scores compared to the baseline model B-PCA-20 and B-AE-20. Considering the SSL models only, the proposed SSL-RN-20 shows comparable performance with the other two models, with SSL-ViT-20 being slightly better above 20 clusters.

Here, again we plot fitted curves to show the trends for both metrics. In general, the Silhouette score evaluation indicates the same pattern as the Davies–Bouldin score. The proposed SSL-RN-20 shows comparable performance with the SSL-VGG-20 and SSL-ViT-20, and significantly better than the baselines. This observation supports our conclusion from the previous Section 5.2, which evaluated the clustering tendency of the features, that all three of the SSL models have comparable groupings in the feature space and are significantly better than both baseline models.

5.4. Analysis of clusters

According to Sections 5.2 and 5.3, models with 20 to 25 clusters are providing the best results on the used dataset. Considering this advantage of using the SSL architectures over the baseline clustering approach, we present two use cases where the learning capabilities of such model are being exploited:

1. Transmission detection, where we group the activity-rich clusters containing transmissions as one class, and the clusters with spectrograms that do not contain transmissions as another class.
2. High granularity clustering, considering all of the existing shapes appearing in the formed clusters.

Table 2
Statistics of the clusters. Legend: yellow-Stripes, red-Dotted, blue-Edges, gray-Idle, green-High intensity, white-Undefined.

B-PCA-20			SSL-VGG-20			SSL-RN-20			SSL-ViT-20			B-AE-20		
ID	Size %	Image mse.	ID	Size %	Image mse	ID	Size %	Image mse	ID	Size %	Image mse	ID	Size %	Image mse
1	30.32	0.00395	4	14.0	0.00007	9	6.28	0.00032	11	6.08	0.00055	1	11.16	0.00012
7	10.2	0.00267	5	7.23	0.00016	10	6.19	0.00034	1	5.78	0.00090	3	8.5	0.00057
18	8.97	0.00420	17	5.66	0.00017	17	6.1	0.00006	20	5.24	0.00029	18	7.98	0.00021
9	3.75	0.00426	3	5.34	0.00016	8	5.77	0.00033	5	4.96	0.00070	6	7.07	0.00011
6	3.66	0.00435	6	5.34	0.00015	1	5.75	0.00033	7	4.85	0.00059	19	6.39	0.00033
22	3.49	0.00447	15	5.12	0.00016	5	5.61	0.00036	0	4.81	0.00077	4	4.44	0.00042
10	3.4	0.00432	12	5.08	0.00016	13	5.48	0.00033	15	4.7	0.00067	20	4.08	0.00049
16	3.19	0.00441	18	4.67	0.00035	4	5.48	0.00010	13	4.67	0.00076	7	4.07	0.00082
12	3.16	0.00452	1	4.57	0.00036	2	5.12	0.00043	2	4.6	0.00072	9	4.02	0.00039
8	3.14	0.00433	13	4.52	0.00055	6	5.1	0.00040	10	4.52	0.00011	16	3.86	0.00213
13	2.88	0.00429	8	4.44	0.00018	14	4.76	0.00037	18	4.5	0.00069	15	3.56	0.00038
14	2.84	0.00451	9	4.38	0.00042	21	4.73	0.00047	8	4.46	0.00091	11	3.51	0.00066
19	2.76	0.00441	0	4.09	0.00008	15	4.57	0.00029	19	4.45	0.00083	12	3.5	0.00062
17	2.7	0.00414	14	3.82	0.00017	19	4.5	0.00057	6	4.35	0.00075	21	3.48	0.00040
21	2.68	0.00463	19	3.75	0.00017	16	3.94	0.00010	9	4.28	0.00108	5	3.36	0.00116
20	2.28	0.00461	10	3.3	0.00040	3	3.7	0.00015	16	4.24	0.00067	17	3.31	0.00060
11	2.04	0.00461	16	3.29	0.00035	20	3.26	0.00036	3	4.18	0.00062	8	3.17	0.00070
15	2.04	0.00463	2	3.22	0.00039	22	3.11	0.00045	12	4.08	0.00076	14	2.89	0.00193
5	1.98	0.00448	11	3.02	0.00108	11	2.34	0.00036	21	4.03	0.00072	13	2.78	0.00031
4	1.87	0.00465	22	2.24	0.00318	12	2.31	0.00041	17	4.01	0.00075	0	2.76	0.00031
0	1.64	0.00428	20	1.28	0.00309	18	2.11	0.00330	14	3.42	0.00065	22	2.6	0.00123
3	0.84	0.00472	21	0.88	0.00041	0	2.07	0.00030	22	3.09	0.00016	10	2.55	0.00032
2	0.18	0.00281	7	0.76	0.00045	7	1.7	0.00006	4	0.7	0.00025	2	0.96	0.00215
Image mse mean		0.00427			0.00046			**0.00044**			0.00065			0.00071
Image mse std		0.00051			0.00074			0.00062			**0.00023**			0.00059

Statistics of the three compared models are summarized in Table 2. Complementary to the table, Fig. 8 provides the average spectrograms and histograms of frequency sub-bands for the samples of each cluster. Cluster statistics in Table 2 contain the cluster sizes in percentages of the entire dataset (*Cluster size %*) and image mean-squared-error similarities between each cluster's average spectrogram and the spectrograms of its assigned samples (*Image mse*). Colors of the cells are related to the clusters' content discussed later in Section 5.5 and depicted in Fig. 8.

Interestingly, the B-PCA-20 has the lowest *overall mean of inter-cluster (IC) distances*, one order of magnitude lower than the SSL models. But, this is a result of having small variance ratio contained in the 20 dimensions of the PCA transformed space of the raw images (baseline PCA approach), and to note again, using bigger amount of components resulted to even worse results in Section 5.2. On the other hand, the IC image distances are one order of magnitude worse compared to the SSL models, which confirms the performance seen in Section 5.3.

SSL-VGG-20 has slightly better *overall mean IC distance* than SSL-RN-20, which corresponds to better distinguishing of the features according to Fig. 6.

Looking at the CNN-based models, SSL-VGG-20 and SSL-RN-20, they show almost identical performance for the *Image mse mean*, 0.00046 and 0.00044, accordingly. At the same time, the overall *Image mse std* is better for the SSL-RN-20. While the SSL-ViT has the lowest *Image mse std*, it has the worst *Image mse mean* among the SSL models. This means that although the encoded features are better grouped by the SSL-VGG-20 and SSL-ViT-20, according to the metrics in Section 5.3, the differences between the clustered spectrograms of these two SSL models are actually larger than the ones resulting from the SSL-RN-20 model, thus showing that the latter has slight advantage in grouping wireless transmissions with similar shapes. The simple baseline model B-PCA-20, expectedly, has the worst performance, with *Image mse mean* of 0.00427 which is one order of magnitude larger than the rest of the models.

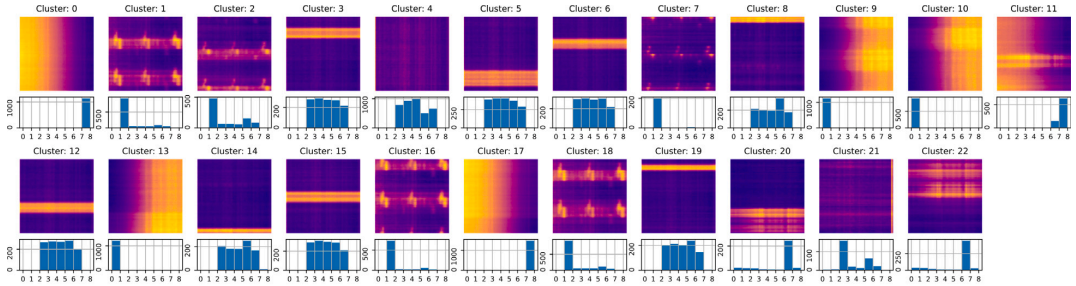
Interestingly, the B-AE-20 has very close performance to the SSL models with *Image mse mean* of 0.00071, contrary to the significant underperforming seen on the feature space evaluations with the other metrics in Figs. 6, 7(a) and 7(b). This means that, while the feature space is not very clearly separated, the clusters are formed by similar types of transmissions. However, it should be noted that the number of clusters for the experiment equal to 23 was selected based on the performance of the SSL models seen using the previous metrics. The evaluations of the B-AE-20 in the embedded domain show similar results almost for the entire range of the number of clusters. So even if the results are good in this case, selection of the right number of clusters for this model is a guesswork, and not a result of methodical reasoning.

5.5. Manual evaluation

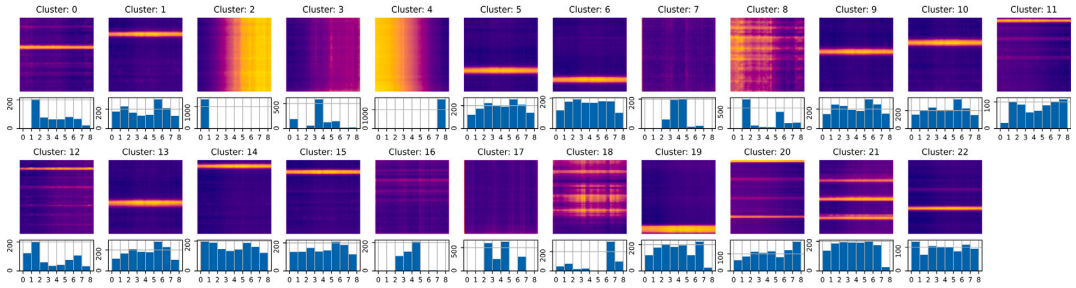
Manual evaluation for the selected models with 23 clusters relies on the distribution of samples per cluster with the relevant cluster statistics in Table 2 and the average spectrogram corresponding to each cluster in Fig. 8. Averaging of spectrogram samples leads to highlighting the position of the most common types of activities that appear in each of the clusters. This visualization reveals which type of spectrum activity (i.e. transmissions of existing wireless technologies) corresponds to each of the clusters. So, instead of labeling large amounts of data, we only need to analyze the content of the formed clusters. Since the B-PCA-20 model has shown weak clustering performance according to all previous metrics, manual evaluation was performed only on the clusters obtained with the SSL models and the B-AE-20.

Based on the activity shapes and frequency histograms, clusters of all the models can be separated in six major groups (with color marking as in Table Table 2): *Idle* clusters (gray), *Edge* clusters (blue), *High intensity* activity (green), *Dotted* activity (red), horizontal *Stripes* activity (yellow), and *Unidentified* (white).

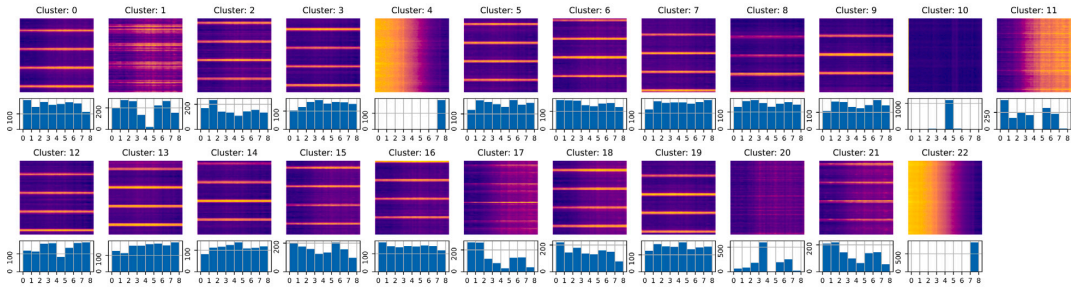
Idle group, is represented by cluster 4 for SSL-VGG-20 and clusters 7, 16 and 17 for the SSL-RN-20, containing 14% and 11.74% of the data, respectively. This means that there is a slight difference in the number of samples assigned to no activity clusters between the models, as discussed later in Section 5.6. The SSL-ViT-20 has two clusters, 20 and 10 with *Idle* samples, totaling 9.76% of the data, which is even less than the other two SSL models. Considering the averaged clusters in Fig. 8(c) we conclude that it is even more sensitive to the low power transmissions, so in Idle clusters are only the samples with containing very low or no transmissions. The B-AE-20 shows completely opposite



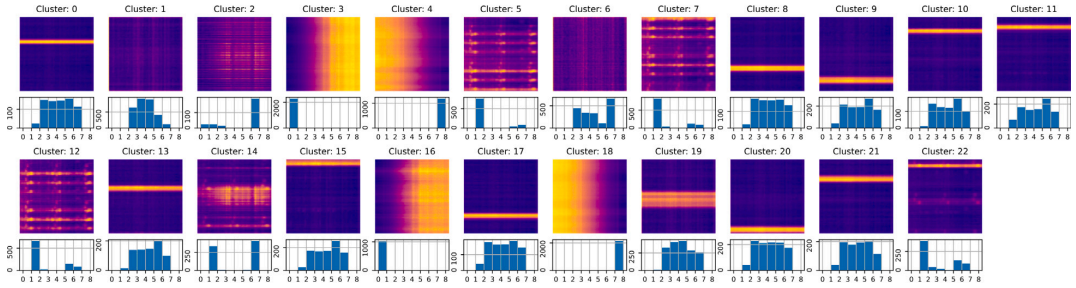
(a) SSL-VGG-20



(b) SSL-RN-20



(c) SSL-ViT-20



(d) B-AE-20

Fig. 8. Average spectrograms and frequency histograms of appearance for each cluster.

behavior, by grouping more than 18% of the data as Idle, contained in the clusters 1 and 6. This means that it is the least sensitive model towards detection of low power transmissions. From practical point of view, the SSL-ViT-20 model would be desirable if there is high criteria for avoiding interference when estimating opportunities for transmission.

The left *Edges* and right *Edges* clusters, marked with blue in Table 2, are mainly formed by samples from the edges (left-most and right-most) of the observed bandwidth (also shown on the histograms under each averaged spectrograms). The varying sensitivity of the sensor towards the edges of the bandwidth appears as highly influential information for all of the models.

The SSL-VGG-20 has clusters 9, 10 and 13 from the left *Edge* of the bandwidth, totaling 12.2% of the whole data which matches nearly perfect with $1/8 = 12.5$ of the data falling in the left-most sub-band (the entire bandwidth being separated in 8 sub-bands). The right *Edge*

cluster formed by clusters 0, 11 and 17, totaling 12.77% of the data, follows the same pattern. Although it is evident that these clusters also contain additional content (e.g. clusters 9, 10, 13), it is clear that the most prominent shape is the varying intensity background noise. This means that for these bandwidth regions, SSL-VGG-20 is focused more on the background noise than the transmission activity. For the SSL-RN-20, the left *edge* and right *Edge* spectrum samples are contained in only two clusters, labeled 2 and 4, with roughly equal size of 5.12% and 5.61% of the total data, respectively. This means that the varying noise is also important for the SSL-RN-20, but more weight is given to the wireless transmission activities.

The SSL-ViT-20 has clean clusters for the right *Edge*, 4 and 22 (totaling 3.79% of the data), while the left *Edge* cluster, 11, is mixed with other types of activities. Meaning, the edge characteristics are not decisive in the clustering process of this model. The baseline B-AE-20 shows similar performance like the SSL-VGG-20 regarding the *Edge* samples, with 24.78% of the data falling into the clusters 3, 4, 16 and 18. Again, almost equal to the total amount of the samples from the edges.

Considering the *High intensity* activity samples, marked with green in Table 2, for SSL-VGG-20 they are distributed in clusters 20 and 22 (totaling 3.52%) while for SSL-RN-20 they are only in cluster 18 (2.11%). For this activity SSL-VGG-20 achieves better clustering. Based on the histogram of cluster 19 which has higher number of samples in the 7th sub-band, exactly where these activities appear according to the histogram of cluster 18, SSL-RN-20 is confusing the high intensity activity with the horizontal *Stripes* activity in cluster 19. The SSL-ViT-20 has no cluster with *High intensity* activities, and the B-AE-20 has clusters 14 and 2 (containing 3.85%), again very close to the numbers of the SSL-VGG-20 model.

Looking at the *Dotted* activity samples, marked with red in Table 2, contained in clusters 1, 2, 7, 16 and 18 for SSL-VGG-20 and in cluster 8 for SSL-RN-20, we observe that SSL-VGG-20 focuses much more on these patterns, ignoring the coexistence of other types of activities in the spectrograms. The same pattern appears also for the B-AE-20 model, considering clusters 5, 7 and 12, which are slightly more mixed with other types of transmissions from the sub-bands 6 and 7. The histograms show that samples from these clusters contain all of the samples of the second sub-band, including those belonging to the Horizontal stripes activity which should obviously appear across the entire bandwidth as indicated in Fig. 3. This is also confirmed by the histograms of the horizontal *Stripes* activity, for both SSL-VGG-20 and B-AE-20, which are expected to be distributed along the entire bandwidth, but they are only in the sub-bands 3-7, confirming misclassifications of the *Dotted* activity samples.

On the other hand, SSL-RN-20 focuses much more on the horizontal *Stripes* activity samples, marked with yellow in Table 2, forming many clusters based only on the vertical location of the activity and having well distributed histograms across all sub-bands, as one would expect. Most of the samples that contain the horizontal *Stripes* and the *Dotted* activity samples are thus classified in the clusters with the horizontal *Stripes* (e.g. clusters 0, 11, 12 and 20). For the SSL-ViT-20, most of the samples (70.6%) fall into the horizontal *Stripes* class, meaning this type of activity is the main focus for this model while only some of the edges and *Idle* samples form separate clusters.

In general, SSL-VGG-20 and SSL-RN-20 show better results on the manual evaluation compared to the SSL-ViT-20. Surprisingly, the B-AE-20 results look even better than the SSL-ViT-20, because of having more variety in the clusters with different types of transmissions, and slightly worse than the SSL-VGG-20.

Both SSL-VGG-20 and SSL-RN-20 architectures show comparable performance, which is more balanced towards different types of transmission activities when compared to the SSL-ViT-20 and the B-AE-20, having similar pros and cons. The difference is that they focus on different shapes of activities, so samples containing multiple transmissions are distributed across different clusters, predominantly the Dotted activity for SSL-VGG-20 and the Horizontal stripes activity for SSL-RN-20. Considering this finding only these two models will be considered in the following use cases.

5.5.1. Use case 1 - transmissions detection

This use case is equivalent to binary classification of the labeled (positive) samples between the two groups of clusters which represent the Idle state and the Occupied state.

Idle state - No/weak spectrum activity. According to Fig. 8, clusters with labels 7, 16, 17 for SSL-RN-20 and 4 for SSL-VGG-20, contain no or very low activity (*idle*). These clusters represent the parts of spectrum where no transmissions are recorded, or the sources of the signals are very distant so the signal intensity is very low. In the context of spectrum availability, they could be considered very similar since in both cases there are no existing patterns of transmissions with significant intensity.

Occupied - presence of transmissions. All other clusters, 20 for the SSL-RN-20 and 22 for the SSL-VGG-20, contain some type of activity, so in the sense of spectrum availability they represent occupied spectrum segments. By considering the counts in Table 2, it can be seen that these amount to 88% (SSL-RN-20) and 86% (SSL-VGG-20) of spectrograms that contain at least one transmission. This represents valuable information about the spectrum occupancy that was extracted in automatic way and all the manual work is reduced to inspection of the averaged clusters' spectrograms.

5.5.2. Use case 2 - Distinguishing types of activities

The cluster contents and the cluster size allow for better understanding of types of activities and frequency of their appearance across the monitored radio spectrum. Considering only the low-activity clusters, we acknowledge the percentage of available slots for transmission during time. Analog to this, there are also percentages for other types of activities. This means that information is provided about what portion of the spectrum is occupied with certain types of activity or combination of overlapping activities. This is useful when more granular separation of the spectrograms is required. The mentioned manual evaluation of the averaged spectrograms has a crucial role in such case. For this use case we only exemplify the SSL-RN-20 model with 23 clusters.

The clusters containing different types of combined activities can be considered as separate classes. In this way, besides the information for the spectrum occupancy, additional information can be derived for the frequency of occurrence and types of overlapping transmissions. In summary, based on manual evaluation of the clusters in Section 5.5, the number of activity types can be reduced to 6, including the *Idle* type of clusters containing 11.74% of the total data.

Occupied - Horizontal stripes. According to the average spectrograms visualization in Fig. 8(b), there are several clusters with mostly horizontal stripes (clusters 1, 5, 6, 9, 10, 13, 14, 15, 19 and 22), which account for 51.35% of the dataset. These samples represent the same type of activity, namely IEEE 802.15.4 transmissions, that occupy the entire bandwidth covered with a single segment, and they last for short time with regards to the time duration of a single sample. The location of the stripe is varying in vertical direction because the segmentation of the data in non-overlapping windows is fixed, as explained previously in Section 4.1, without correlation to the appearance times of any of the activities.

Occupied state - Horizontal stripes + other types of activity. Second type of activities is observed when the same horizontal stripes appear combined with other activities (clusters 0, 11, 12, 20, and 21), representing 14.71% of the samples. Together they represent separate cluster of segments where multiple activities are occurring in the same spectrogram segment. Some of the activities we distinguish correspond to concurrent IEEE 802.15.4, proprietary and LoRA transmissions.

Occupied state - High intensity activity. Cluster 18 (2.11% of the data) contains *High intensity* activity spread along the time axis (vertical) direction. These activities correspond to LoRA transmissions. Many of these transmissions are also coexisting with the IEEE 802.15.4 transmissions, as discussed in Section 5.5.

Table 3
Transmission detection performance and complexity comparison.

Algorithm	SSL-RN-20	SSL-VGG-20	Dilate/erode [12]	TX grouping [12]	SSL-ViT-20	B-AE-20
Precision %	76.0	76.6	**77.7**	68.2	74.1	76.8
Recall %	93.6	94.1	86.3	94.7	**96.5**	90.3
F1 %	83.9	**84.5**	81.8	79.3	83.9	83.0
Num. of parameters	**11.2 M**	114.1 M	/	/	54.6 M	19.9 M

Occupied state - Dotted activity. Cluster 8 (5.77% of the data) contains shapes of dotted activity that are apparent at various locations along the time (vertical) axis. Such activities correspond to LoRa and proprietary transmissions.

Edge sub-bands. It is important to notice that varying background noise is also influential in the clustering. Samples in clusters 2 and 4 contain significant amount of background noise, having weaker values on one side of the spectrogram and gradually increasing in horizontal direction (frequency axis). As a consequence of this, such spectrograms are clustered separately making up 10.6% of the data, which cannot be specified if it is occupied or not.

5.6. Evaluation with labeled data

At last, we evaluate the performance of the models for transmissions detection according to Section 4.4, feeding labeled samples of transmissions to the models and determining how many of these samples were correctly distributed to clusters with present transmissions and clusters with no transmissions. Table 3 summarizes the evaluation. We also compare the proposed solution with existing referenced transmission detection algorithms evaluated on the same dataset in [12]. Surprisingly, the *Dilate/Erode* achieves the highest precision score with 1.1pp (percentage point) margin and the lowest recall score with 6-7pp margin, while the *TX grouping* achieves the second highest recall score and the lowest precision with almost 8pp margin. This means that, although the *Dilate/Erode* has the highest ratio of real transmissions in the detected transmissions, it fails to detect significant amount (roughly 14%) of the total number of transmissions. On the opposite, the *TX grouping* detects the high percentage of the total number of transmissions, but at the same time also classifying many of the *Idle* samples as samples with transmissions, causing poor precision.

Contrary to the *Dilate/Erode* and *TX grouping*, the SSL-based models show more balanced performance, considering the difference between the precision and recall, thus achieving higher F1 scores by margin of more than 2pp. In general, the SSL-based models outperform the referenced ones on the task of transmission detection in continuously sensed spectrum data.

The B-AE-20, employed as a baseline model also outperforms the referenced models on this task, however it is still slightly worse than the SSL-based models.

Now comparing the SSL-based models, all three of them show comparable performance. The (original) SSL-VGG-20 has slight advantage of up to 0.6 pp in F1 score. And the SSL-ViT-20 has more significant 2.9 pp advantage on the Recall. Thus, while CNN-based SSL models are good for fine grained, transmission-specific clustering they may confuse certain transmissions with background noise. On the other hand, the ViT-based model is able to better separate Idle samples, as can also be seen from Figure Fig. 8, thus achieving better Recall performance on the labeled evaluation for activity detection according to Table 3. Anyway, the proposed SSL-RN-20 achieves such performance by utilizing roughly 10 times less trainable parameters compared to the SSL-VGG-20, roughly 5 times less compared to the SSL-ViT-20, which could be a beneficial trade-of regarding the computing requirements for potential applications. The scripts used for these experiments are publicly available as open source software.[3]

[3] https://github.com/sensorlab/self-supervised-spectrum-sensing

6. Conclusions

In the work presented in this paper, we investigated the suitability of SSL for automatic feature learning and clustering of real-world radio spectrum events making use of three different SSL models, based on a similar architecture. SSL-VGG-20 originally developed for RGB image feature learning, its adaptation to the specifics of wireless spectrograms - SSL-RN-20, and SSL-ViT-20, an adaptation based on a vision transformer. We employed these architectures to develop SSL models with no labels and no prior knowledge, and compared their performance against two baseline models, PCA dimensionality reduction as a simple, yet informative representation learning approach and another, CNN-based auto-encoder, both complemented with K-means clustering. The comparison was first made for the evaluation of the clustering tendency of the input data with appropriately proposed metrics for the different models and the quality of the clustering results. The self-supervised models enable compact feature extraction, encoded in significantly lower number of dimensions (reduction is of two orders of magnitude) while preserving the same amount of EVR of the input features. They also outperform the baseline models in the clustering quality by 0.3 according to Silhouette scores, and by 0.6 according to Davies–Bouldin score. The study was supported also by manual evaluation based on visual inspection and evaluation by labeled data. Results show that SSL-based models outperform the existing transmission detection algorithms on continuously sensed data by more than 2pp according to F1 scores.

CRediT authorship contribution statement

Ljupcho Milosheski: Conceptualization, Methodology, Software, Investigation, Writing – original draft, Visualization. **Gregor Cerar:** Software, Resources, Data curation. **Blaž Bertalanič:** Methodology, Resources. **Carolina Fortuna:** Conceptualization, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Mihael Mohorčič:** Conceptualization, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: All coauthors report financial support was provided by Public Research Agency of the Republic of Slovenia. Carolina Fortuna, Gregor Cerar report financial support was provided by Horizon Europe.

Data availability

Data will be made available on request.

Acknowledgments

This work was funded in part by the Slovenian Research Agency under the grant P2-0016. This project has received funding from the European Union's Horizon Europe Framework Programme under grant agreement No 101096456 (NANCY). The project is supported by the Smart Networks and Services Joint Undertaking and its members.

References

[1] Q. Zhao, B.M. Sadler, A survey of dynamic spectrum access, IEEE Signal Process. Mag. 24 (3) (2007) 79–89.

[2] V. Frascolla, A.J. Morgado, A. Gomes, M.M. Butt, N. Marchetti, K. Voulgaris, C.B. Papadias, Dynamic Licensed Shared Access-A new architecture and spectrum allocation techniques, in: 2016 IEEE 84th Vehicular Technology Conference, VTC-Fall, IEEE, 2016, pp. 1–5.

[3] X. Chen, J. Huang, Database-assisted distributed spectrum sharing, IEEE J. Sel. Areas Commun. 31 (11) (2013) 2349–2361.

[4] D.F. Külzer, S. Stańczak, M. Botsov, CDI maps: Dynamic estimation of the radio environment for predictive resource allocation, in: 2021 IEEE 32nd Annual International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC, IEEE, 2021, pp. 892–898.

[5] S. Rajendran, W. Meert, D. Giustiniano, V. Lenders, S. Pollin, Deep learning models for wireless signal classification with distributed low-cost spectrum sensors, IEEE Trans. Cogn. Commun. Netw. 4 (3) (2018) 433–445.

[6] J. Fontaine, E. Fonseca, A. Shahid, M. Kist, L.A. DaSilva, I. Moerman, E. De Poorter, Towards low-complexity wireless technology classification across multiple environments, Ad Hoc Netw. 91 (2019) 101881.

[7] N. Tandiya, A. Jauhar, V. Marojevic, J.H. Reed, Deep predictive coding neural network for rf anomaly detection in wireless networks, in: 2018 IEEE International Conference on Communications Workshops, ICC Workshops, IEEE, 2018, pp. 1–6.

[8] C. Zhang, P. Patras, H. Haddadi, Deep learning in mobile and wireless networking: A survey, IEEE Commun. Surv. Tutor. 21 (3) (2019) 2224–2287.

[9] L.J. Wong, W.H. Clark, B. Flowers, R.M. Buehrer, A.J. Michaels, W.C. Headley, The RFML ecosystem: A look at the unique challenges of applying deep learning to radio frequency applications, 2020, arXiv preprint arXiv:2010.00432.

[10] S. Rajendran, R. Calvo-Palomino, M. Fuchs, B. Van den Bergh, H. Cordobés, D. Giustiniano, S. Pollin, V. Lenders, Electrosense: Open and big spectrum data, IEEE Commun. Mag. 56 (1) (2017) 210–217.

[11] T. Šolc, C. Fortuna, M. Mohorčič, Low-cost testbed development and its applications in cognitive radio prototyping, in: Cognitive Radio and Networking for Heterogeneous Wireless Networks, Springer, 2015, pp. 361–405.

[12] T. Gale, T. Šolc, R.-A. Moşoi, M. Mohorčič, C. Fortuna, Automatic detection of wireless transmissions, IEEE Access 8 (2020) 24370–24384.

[13] L. Jing, Y. Tian, Self-supervised visual feature learning with deep neural networks: A survey, IEEE Trans. Pattern Anal. Mach. Intell. 43 (11) (2020) 4037–4058.

[14] T.J. O'Shea, T. Roy, T.C. Clancy, Over-the-air deep learning based radio signal classification, IEEE J. Sel. Top. Sign. Proces. 12 (1) (2018) 168–179.

[15] J. Robinson, S. Kuzdeba, J. Stankowicz, J.M. Carmack, Dilated causal convolutional model for RF fingerprinting, in: 2020 10th Annual Computing and Communication Workshop and Conference, CCWC, IEEE, 2020, pp. 0157–0162.

[16] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al., An image is worth 16x16 words: Transformers for image recognition at scale, 2020, arXiv preprint arXiv:2010.11929.

[17] E. Fonseca, J.F. Santos, F. Paisana, L.A. DaSilva, Radio access technology characterisation through object detection, Comput. Commun. 168 (2021) 12–19.

[18] M. Caron, P. Bojanowski, A. Joulin, M. Douze, Deep clustering for unsupervised learning of visual features, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 132–149.

[19] S. Kuzdeba, J. Robinson, J. Carmack, Transfer learning with radio frequency signals, in: 2021 IEEE 18th Annual Consumer Communications & Networking Conference, CCNC, IEEE, 2021, pp. 1–9.

[20] S. Riyaz, K. Sankhe, S. Ioannidis, K. Chowdhury, Deep learning convolutional neural networks for radio identification, IEEE Commun. Mag. 56 (9) (2018) 146–152.

[21] L.J. Wong, W.C. Headley, S. Andrews, R.M. Gerdes, A.J. Michaels, Clustering learned cnn features from raw i/q data for emitter identification, in: MILCOM 2018-2018 IEEE Military Communications Conference, MILCOM, IEEE, 2018, pp. 26–33.

[22] S. Rajendran, W. Meert, V. Lenders, S. Pollin, SAIFE: Unsupervised wireless spectrum anomaly detection with interpretable features, in: 2018 IEEE International Symposium on Dynamic Spectrum Access Networks, DySPAN, IEEE, 2018, pp. 1–9.

[23] Q. Feng, Y. Zhang, C. Li, Z. Dou, J. Wang, Anomaly detection of spectrum in wireless communication via deep auto-encoders, J. Supercomput. 73 (7) (2017) 3161–3178.

[24] K. Davaslioglu, S. Boztaş, M.C. Ertem, Y.E. Sagduyu, E. Ayanoglu, Self-supervised RF signal representation learning for nextg signal classification with deep learning, IEEE Wirel. Commun. Lett. 12 (1) (2022) 65–69.

[25] D. Liu, T. Wang, S. Liu, R. Wang, S. Yao, T. Abdelzaher, Contrastive self-supervised representation learning for sensing signals from the time-frequency perspective, in: 2021 International Conference on Computer Communications and Networks, ICCCN, IEEE, 2021, pp. 1–10.

[26] M. Caron, H. Touvron, I. Misra, H. Jégou, J. Mairal, P. Bojanowski, A. Joulin, Emerging properties in self-supervised vision transformers, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 9650–9660.

[27] B. Hopkins, J.G. Skellam, A new method for determining the type of distribution of plant individuals, Ann. Botany 18 (2) (1954) 213–227.

[28] J.C. Bezdek, R.J. Hathaway, VAT: A tool for visual assessment of (cluster) tendency, in: Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02. Vol. 3, Cat. No. 02CH37290, IEEE, 2002, pp. 2225–2230.

[29] P.J. Rousseeuw, Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, J. Comput. Appl. Math. 20 (1987) 53–65.

[30] D.L. Davies, D.W. Bouldin, A cluster separation measure, IEEE Trans. Pattern Anal. Mach. Intell. (2) (1979) 224–227.

Chapter 4

Spectrum Sensing with Deep Clustering for Label-Free Radio Access Technology Recognition Based on FFT Amplitudes

In Chapter 3, it was experimentally demonstrated that existing deep learning architectures originally developed for image processing can be successfully adopted within a deep clustering UDL pipeline for wireless transmission analysis using spectrogram representations. Through systematic feature space adaptation and dimensionality reduction, substantial improvements in computational efficiency were achieved. Nevertheless, the resulting models still remain relatively large, with approximately eleven million trainable parameters, which may pose challenges for deployment in resource-constrained platforms, particularly in distributed spectrum sensing applications operating within O-RAN architectures, where control loop execution times can be as short as ten milliseconds.

In such deployment contexts, lightweight models offer considerably greater practical viability. This is particularly relevant given that future O-RAN systems are expected to execute multiple intelligent applications concurrently while sharing common processing and memory resources. Motivated by these constraints, this chapter investigates the design of a dedicated lightweight deep learning model tailored to the characteristics of wireless transmission signals and integrated within a deep clustering pipeline. In addition, a deployment workflow is formulated that explicitly addresses the inherent limitation of deep clustering, namely, the absence of class assignments. This is achieved by introducing a cluster-to-label mapping step performed by a domain expert, based on data-specific visualizations, thereby reducing the annotation effort from tens of thousands of individual samples to a small number of clusters.

The proposed model is rigorously evaluated against an existing state-of-the-art deep clustering solution and a baseline model across three real-world transmission datasets, two labeled and one unlabeled dataset. This investigation leads to the realization of the contributions **C3** and **C4**, i.e. design of a dedicated deep learning architecture within a deep clustering framework, and its integration as a core component of a practical spectrum sensing workflow.

The evaluation results demonstrate that significant reductions in model size and computational complexity are achievable relative to the existing image processing architectures examined in Chapter 3 while preserving or improving performance on downstream tasks across the individual environments. The diversity of conditions is reflected by datasets

that differ in the number and types of active RATs and were acquired in distinct urban radio environments.

In doing so, this chapter ultimately confirms the hypotheses **H2**, i.e. *adapting existing UDL architectures to the specific characteristics of radio frequency spectrum data enables substantial reductions in model and feature complexity while preserving or improving downstream tasks performance*, and **H3**, i.e. *it is feasible to develop unsupervised deep learning models that remain robust under changing radio environments, including variations in channel conditions, operating frequency bands, and the number of active radio access technologies.*

This work was published as:

- L. Milosheski, M. Mohorčič and C. Fortuna, "Spectrum Sensing With Deep Clustering: Label-Free Radio Access Technology Recognition," in IEEE Open Journal of the Communications Society, vol. 5, pp. 4746-4763, 2024.

IEEE ComSoc IEEE Open Journal of the Communications Society

Received 21 June 2024; accepted 25 July 2024. Date of publication 5 August 2024; date of current version 9 August 2024.

Digital Object Identifier 10.1109/OJCOMS.2024.3436601

Spectrum Sensing With Deep Clustering: Label-Free Radio Access Technology Recognition

LJUPCHO MILOSHESKI [1,2] (Graduate Student Member, IEEE), MIHAEL MOHORČIČ [1,2] (Senior Member, IEEE), AND CAROLINA FORTUNA [1]

[1]Department of Communication Systems, Jožef Stefan Institute, 1000 Ljubljana, Slovenia
[2]Information and Communication Technologies, Jožef Stefan International Postgraduate School, 1000 Ljubljana, Slovenia

CORRESPONDING AUTHOR: L. MILOSHESKI (e-mail: ljupcho.milosheski@ijs.si)

This work was supported in part by the Slovenian Research and Innovation Agency under Grant P2-0016, and in part by the European Union's Horizon Europe Framework Programme under Grant 101096456 (NANCY). The NANCY Project was supported by the Smart Networks and Services Joint Undertaking and its members.

ABSTRACT The growth of the number of connected devices and network densification is driving an increasing demand for radio network resources, particularly Radio Frequency (RF) spectrum. Given the dynamic and complex nature of contemporary wireless environments, characterized by a wide variety of devices and multiple RATs, spectrum sensing is envisioned to become a building component of future 6G, including as a components within O-RAN or digital twins. However, the current SotA research for RAT classification predominantly revolves around supervised Convolutional Neural Network (CNN)-based approach that require extensive labeled dataset. Due to this, it is unclear how existing models behave in environments for which training data is unavailable thus leaving open questions regarding their generalization capabilities. In this paper, we propose a new spectrum sensing workflow in which the model training does not require any prior knowledge of the RATs transmitting in that area (i.e., no labelled data) and the class assignment can be easily done through manual mapping. Furthermore, we adaptat a SSL deep clustering architecture capable of autonomously extracting spectrum features from raw 1D Fast Fourier Transform (FFT) data. We evaluate the proposed architecture on three real-world datasets from three European cities, in the 868 MHz, 2.4 GHz and 5.9 GHz bands containing over 10 RATs and show that the developed model achieves superior performance by up to 35 percentage points with 22% fewer trainable parameters and 50% less floating-point operations per second (FLOPS) compared to an SotA AE-based reference architecture.

INDEX TERMS Analysis, clustering, machine learning, monitoring, self-supervised, radio frequency spectrum.

I. INTRODUCTION

THE EXPONENTIAL growth of the number of connected devices [1] and network densification is driving an increasing demand for radio network resources, particularly Radio Frequency (RF) spectrum. The allocation of new operational frequency bands, such as the 6.425-7.125 GHz range designated by the ITU for licensed mobile communications [2], only partially alleviates the growing demand for RF spectrum resources. Thus, it is crucial to enhance the utilization of the occupied bands through innovative spectrum-sharing strategies that go beyond the existing licensed [3] and license-exempt access schemes [4], meeting the complexity of the fast-evolving wireless networks. Such approaches will increasingly rely on accurate monitoring and understanding of the RF spectrum environment, achieved through spectrum sensing techniques. The process of spectrum sensing involves analyzing radio data for a variety of tasks [5], each playing an important role in the broader context of radio awareness, impacting the efficiency and effectiveness of wireless radio networks. Such tasks and corresponding purposes include Radio Access Technology (RAT) classification [6] for optimizing resources allocation and spectrum utilization, modulation classification [7] for improving data throughput and reducing error rates, anomaly

© 2024 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

IEEE ComSoc IEEE Open Journal of the Communications Society

detection [8] enhancing network security and resilience, interference recognition [9] for identification of sources of interference, and Specific Emitter Identification (SEI) [10] ensuring secure and efficient spectrum sharing by preventing unauthorized access.

Given the dynamic and complex nature of contemporary wireless environments, characterized by a wide variety of devices and multiple RATs, spectrum sensing is envisioned to become a building component of future 6G networks [11]. As such, it will support the progress of Integrated Sensing and Communication (ISAC)-related systems and technologies [12], [13], facilitating their development and integration into the next generation of wireless communication systems. Consequently, spectrum sensing is anticipated to be a key component in the evolution of emerging wireless communication frameworks such as Open Radio Access Network (O-RAN) and Digital Spectrum Twin (DST) [14].

In the emerging O-RAN in which a large proportion of the radio functions are realized in software [15], [16], RAT monitoring plays an important role in optimizing radio and network parameters and spectrum utilization, largely through the employment of AI algorithms encapsulated as xApps and rApps [17]. By continuously assessing the usage and availability of different RATs, O-RAN can utilize intelligent applications to dynamically fine tune network settings and allocate spectrum resources or manage beam forming [18] in real time. Thus, it can enhance the overall efficiency and performance of the network and accommodate growing traffic demands, while better preserving privacy and being less data intensive [19] compared to SEI.

DSTs are digital representations of the real-world characteristics of the RF spectrum in a region based on historical data and measurement updates [20]. They are envisioned as tools to inform, forecast, and enhance spectrum utilization and interference management in wireless networks in a given geographical area. Spectrum sensing is a building component of DST, as outlined in [14]. It provides the measurement input necessary to refine the DST, which is initially developed based on empirical, theoretical, and/or ray-tracing models. Incorporating RAT monitoring into DST could further enhance its capabilities by adding detailed data on the various transmission technologies that are present in the area, such as RAT-specific activity patterns for a given time interval within the operational landscape. This integration could significantly improve the DST's awareness, accuracy and functionality, facilitating more effective spectrum management and optimization strategies.

The current SotA research for RAT classification predominantly revolves around supervised Convolutional Neural Network (CNN)-based approaches [21], [22], [23], [24], which have shown promising results, particularly in diverse signal-to-noise ratio (SNR) settings and various environmental conditions [25]. However, these methods have a number of shortcomings as follows. First, they heavily depend on extensive labeled datasets acquired in controlled settings where the transmission parameters need to be recorded in addition to the received signals requiring a more complex set-up for collection compared to unlabelled data. Second, it is unclear how they transfer to other environments characterized by (i) additional RATs compared to the original training data and (ii) different physical obstructions and noise variations. Third, the majority of the models have been developed and evaluated on a single dataset leaving open questions regarding their *generalization* capabilities.

Self-Supervised Learning (SSL) approaches, extensively explored in other application domains [26], are able to address the first two challenges mentioned above, because they do not require labels for training. As long as a spectrum sensor collecting data is available, that data can be directly used to train such models without requiring recorded transmission parameters. By analogy, this approach should be able to adapt the model in regular retraining periods to new types of RATs that may appear in the area. However, to be able to realize the classification functionality similar to their supervised counterparts, a manual cluster-to-label assignment step is required. Nevertheless, this process is rather efficient as the learnt clusters, together with additional insights, can be presented to a human decision maker for label assignment. The decision maker is also presented with a new cluster including a new technology that emerged from the learning process and can immediately identify it. The adoption of Self-Supervised Learning (SSL) in wireless communications research remains limited. Existing state-of-the-art (SotA) works primarily rely on variations of Autoencoder (AE) architectures [23], [27]. These methods emphasize instance learning, where the CNN part is trained solely or partially on reconstruction loss, focusing on distinguishing individual samples.

Recognizing the need for a more adaptable solution, particularly in scenarios where class information is unknown, we propose[1] the use of a domain-adapted DeepCluster [28] architecture. This approach diverges from the existing methods by basing the learning process of the CNN part on features shared among groups of samples rather than individual instances. This distinction could make the proposed architecture more general and effective in realistic environments, thus addressing the third challenge.

The main contributions of our work are:

- *New spectrum sensing workflow:* We propose a new spectrum sensing workflow in which the model training does not require any prior knowledge of the RATs transmitting in that area (i.e., no labelled data) and the class assignment can be easily done through manual mapping.
- *Development of an SSL Deep Clustering (DC) Architecture:* We introduce an adaptation of SSL deep clustering architecture capable of autonomously extracting spectrum features from raw 1D Fast Fourier Transform (FFT) data. Notably, this architecture is

[1] https://github.com/sensorlab/spire

inherently environment-agnostic and adaptable to varying numbers of operating transmission technologies.
- *Generic Unsupervised Model:* We go beyond the SotA in model development for RAT feature learning and subsequent classification by developing a more *generic* model on three real-world datasets from three European cities, in the 868 MHz, 2.4 GHz and 5.9 GHz bands containing over 10 RATs.
- *Efficiency and Performance Optimization:* The developed model boasts 22% fewer trainable parameters and requires 50% less floating-point operations per second (FLOPS) compared to an identical AE-based reference architecture. Besides the efficiency gains, our model achieves superior performance in label-based clustering evaluations by up to 35 percentage points (ppt) on TCD-L dataset (data of each technology acquired in six different sites), up to 4 ppt on ITS-L dataset (samples of each technology acquired at single site) and better separation of the feature space of the unlabeled UNB-U dataset (continuously sensed data in uncontrolled environment).

The structure of this paper is as follows: Section II reviews related work in the field. Section III describes the reference scenario and related assumption. The problem formulation is detailed in Section IV. The proposed and baseline architectures are presented in Section V. Section VI outlines the evaluation methodology, while Section VII discusses the outcomes of this evaluation. The paper concludes with Section VIII, summarizing the key findings and contributions.

II. RELATED WORK

Considering the disadvantages of the supervised learning approaches for spectrum sensing-related tasks, such as the necessity of labels and expert intervention in the setup of models in specific environments, research work towards employing unsupervised models has attempted to advance alternative approaches that do not necessitate labels.

The performance of pure unsupervised clustering models applied to spectrum data is studied in [29]. Authors evaluate classical unsupervised (clustering) approaches on data transformed by t-distributed Stochastic Neighbor Embedding (t-SNE) and Uniform Manifold Approximation and Projection (UMAP). Three clustering algorithms are subjected to comparison: K-means, Agglomerative Hierarchical Clustering, and Hierarchical Density-Based Spatial Clustering. Although such an approach benefits from being general and simple for high-level spectrum analysis, feature extraction is based on t-SNE and UMAP, which is disadvantageous compared to CNN-based, deep feature learning.

A study of another lightweight and general approach for RAT classification is proposed in [30] as an alternative to models based on CNN. It introduces Dynamic Mode Decomposition (DMD) as a feature extraction technique capable of distinguishing RATs based on their bandwidth. The classification process involves a straightforward threshold method, demonstrating superior performance compared to a baseline CNN-based solution that operates on Gramian Angular Summation Field (GASF) images derived from time series data. Although the solution benefits from its simplicity, it faces the problem of manual adjustment of thresholds.

AEs are neural networks trained to encode input data into a compressed representation and then decode that representation back to an approximation of the original input. They are primarily used for dimensionality reduction or feature learning. However, in their original form, where learning is based on reconstruction loss only, they do not provide clustering-friendly embedded space. This led to research on modifications of this architecture by adding different loss functions, as discussed in [31]. AE with an additional clustering loss ([32], [33], [34]) in the embedded space is an SSL model designed for both dimensionality reduction and clustering tasks. Such a model combines the feature-learning capabilities of AEs with the grouping intuition of clustering algorithms by optimizing for a representation that serves both purposes. In general, the clustering loss in the embedded space could be calculated based on distances between samples or based on their distribution.

In [33], a deep learning, AE-based solution is employed for unsupervised feature learning and clustering of embedded features on radio data for the task of modulation recognition. Alongside the AE-specific reconstruction loss, the authors introduce a "Deep Clustering" loss in the embedded space. The total loss function aims to simultaneously improve reconstruction and clustering in the embedded space. The authors provide a detailed evaluation of their proposed model, considering that the number of clusters is the same as the number of different modulations, which is an optimistic setup because it requires knowledge of the exact number of modulations that exist in the data. In a real-world RAT monitoring scenario, it is much more realistic to assume an unknown number of classes/RATs (modulations in their case).

Deep Convolutional Embedded Clustering (DCEC) is another modification of AE for deep feature learning, initially dedicated to image processing proposed in [34]. In this work, the total loss function is a combination of the typical reconstruction loss and Kullback-Leibler divergence loss in the embedded space. While similar learning CNN modules could be used in both approaches, the different loss calculations may optimize the convolution filters to focus on different features in the input data.

Our previous work [35] involves a high complexity self-supervised model as it operates on waterfall plots/spectrograms (2D image-like matrices) employing off-the-shelf ResNet18 model, initially developed for Red-Green-Blue (RGB) images. This approach provides valuable information about the time-frequency occupancy of the spectrum with different RATs that operate in the observed band. Although a substantial reduction of complexity of the feature vectors is achieved by augmenting the feature

IEEE ComSoc IEEE Open Journal of the Communications Society

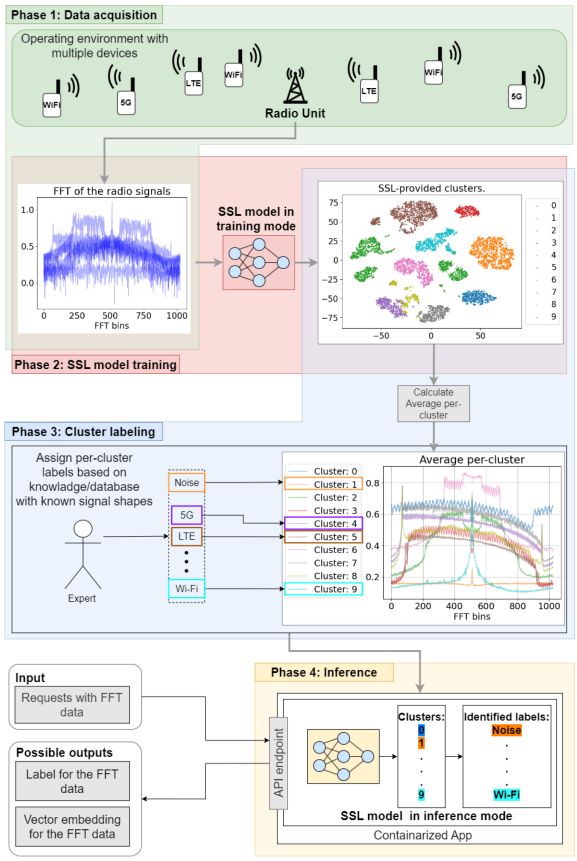


FIGURE 1. SSL enabled spectrum sensing workflow.

processing, the model itself keeps the same (high) complexity as its original implementation for feature learning of RGB images. While this could be advantageous, considering it is off-the-shelf, proven architecture, the high complexity could prevent application in the lower control layers of the O-RAN, for example, as xApp in the Near-real-time (Near-RT) Radio Intelligent Controller (RIC) due to the short control loop times.

In this manuscript, we propose a 1D-CNN-based adaptation of the DeepCluster architecture for feature learning and unsupervised classification of RATs by processing 1D FFT data instead of 2D spectrograms. While the loss of time-frequency dependency is a drawback, this issue can be mitigated through post-processing of the classified 1D swipes.

III. REFERENCE SCENARIO AND PROPOSED SPECTRUM SENSING WORKFLOW

In this paper, we assume a realistic scenario in which terminals, equipped with various RATs as depicted at the top of Figure 1, operate in a given area, possibly also in the same frequency band. For instance, a spectrum sensor may operate in ITS 5.9 GHz band where LTE, Wi-Fi, 5G NR, C-V2X PC5, and ITS-G5 technologies may co-exist [24]. Other examples may refer to monitoring the U.S. 6 GHz band, where Wi-Fi 6E and 5G New Radio Unlicensed (NR-U) share the unlicenseds spectrum, or the LTE/5G and military radar that are sharing the 150 MHz CBRS band at 3.5 GHz [36].

Furthermore, we work under the assumption that we have no prior information on what kind of RATs exist in the area under observation. Without having any prior knowledge on the transmitters in the sensed spectrum, we aim to identify them and recognize them in the future. In order to reach our aim, we propose a four-step workflow as depicted in Figure 1 and elaborated below.

A. PHASE 1: DATA ACQUISITION

The Phase 1, marked with the green area in Figure 1, represents the data acquisition process. In this phase, the Radio Unit collects data from the operating environment in the form of the FFT transformation of the captured radio signals.

B. PHASE 2: SSL MODEL TRAINING

The model training, marked with red in Figure 1, uses the data collected in Phase 1 and learns to group similar FFT shapes together. As a result of the grouping, clusters of similar shapes emerge with each cluster ideally containing samples from a single RAT. As the output of this phase, we obtain the trained model for feature extraction from FFT data and formed clusters.

C. PHASE 3: CLUSTER LABELLING

In this phase, a human expert is able to recognize the existing RATs by looking at the clusters and manually assign a label, i.e., specific RAT such as LTE, WiFi, etc., to each cluster. In this step, hours, days or weeks of spectrum sensing are summarized for the expert end-user to quickly glance at and understand what is happening in the monitored environment. The summary can be presented as an average per cluster sweep such as depicted in Figure 1 or as lower dimensional projections such as T-SNE.

D. PHASE 4: INFERENCE

The model trained on real-world samples in Phase 2 together with the per cluster labels provided in Phase 3 can be packaged as a classification service where the human-assigned labels represent the class of the sample assigned to a cluster. Therefore, the classifier can be encapsulated as a containerized application with an API endpoint, and deployed as a containerized application in a production level system where incoming sweeps would get labels assigned. Furthermore, the trained model could also be exposed as an embedding model, where a compact and low-dimensional representation of the input data is provided to the requester as illustrated in Phase 4 of Figure 1.

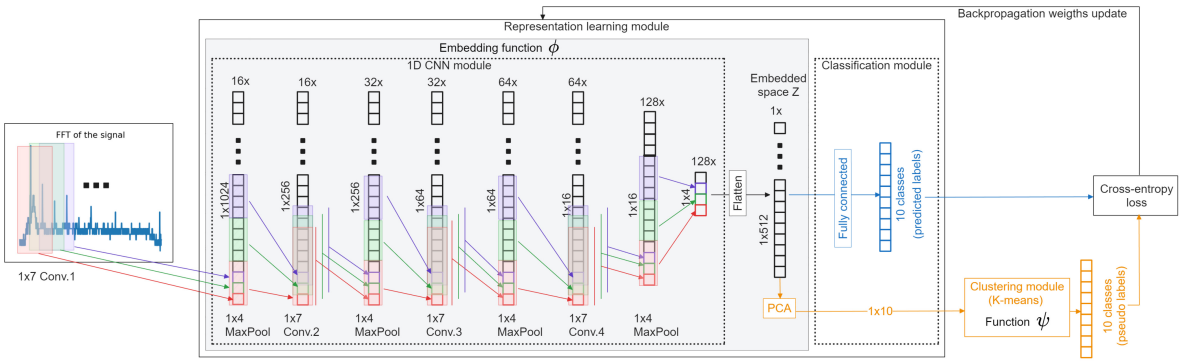


FIGURE 2. Proposed SSDC architecture.

IV. PROBLEM STATEMENT

In this paper, we formulate the spectrum sensing problem as a clustering problem in machine learning that partitions the set of electromagnetic energy measurements collected during individual sweeps into groups (i.e., clusters) that contain sweeps with similar shapes (i.e., energy level envelope). These are later identified and labeled by an expert. This corresponds to Phases 1-3 described in Section III and depicted in Figure 1. We represent a set of electromagnetic energy measurements corresponding to a sweep as $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ and the set of identified RATs as $\mathbf{L} = \{L_1, L_2, \ldots, L_k\}$. To account for realistic environments where the number of types of transmissions is not known apriori, the number of clusters k is assumed to be unknown. This assumption uniquely distinguishes this work from the SotA where the number of classes or clusters are assumed apriori. The objective is to learn a composite function $\Phi : X \rightarrow Y$ that maps raw data points to their identified classes in the clustered space.

$$L = \Phi(X) \tag{1}$$

This function can be realized by decomposition into three sub-functions, the embedding function ϕ, the clustering function ψ, and the mapping function $\mathcal{M}$ and can be expressed as:

$$L = \mathcal{M}(\psi(\phi(X))) \tag{2}$$

A. 1. EMBEDDING FUNCTION

Denoting the ϕ as the embedding function that transforms raw data $\mathbf{X}$ into an embedded space and $Z = \{z_1, z_2, \ldots, z_n\}$ as the set of embedded representations of the data points, formally it is defined as:

$$\phi : X \rightarrow \mathbb{R}^d \tag{3}$$

where $\phi(x_i) = z_i$ and d is the dimensionality of the embedded space.

B. 2. CLUSTERING FUNCTION

Considering $C = \{C_1, C_2, \ldots, C_K\}$ as the set of clusters and K as the number of clusters where $K \leq n$, the clustering function ψ that assigns each embedded data point from $\mathbf{Z}$ to a cluster can be defined as:

$$\psi : \mathbb{R}^d \rightarrow \{1, 2, \ldots, K\} \tag{4}$$

where $\psi(z_i) = k$ indicates that the embedded point z_i is assigned to cluster C_k.

C. 3. MAPPING FUNCTION

Considering $L = \{l_1, l_2, \ldots, l_K\}$ as the set of expert identified labels, the mapping function denoted as $\mathcal{M}$, can be defined as:

$$\mathcal{M} : \{1, 2, \ldots, K\} \rightarrow L \tag{5}$$

where $\mathcal{M}(k) = l_j$ means that cluster C_k is mapped to the label l_j.

D. CONSTRAINTS

The constraint lies in the unknown number of clusters, $\mathbf{K}$, for which the clustering function ψ should be modeled. This is due to the uncontrolled nature of the monitored environment, where there is no prior information about the number of operating RATs. This scenario is general and mirrors a real-world deployment where, regardless of the operating band, new types of signals could appear during the operation of the system, either coming from unauthorized/unknown RATs in the licensed bands or from unknown RATs in the unlicensed bands.

V. PROPOSED SSL ARCHITECTURE

To solve the problem formalized in Section IV, we propose a SSL deep clustering architecture depicted in Figure 2. The architecture iteratively combines deep learning with clustering to learn meaningful data representations without labeled samples. The embedding function from Eq. (3) is realized by the CNN module, part of the representation learning module, while the clustering function from Eq. (4) is realized by the clustering module.

IEEE ComSoc IEEE Open Journal of the Communications Society

A. DESCRIPTION OF ARCHITECTURAL MODULES

1) CLUSTERING MODULE

The clustering module groups similar data points in the embedded space. To realize the clustering module, which in turn realizes the clustering function ψ, we chose the well-known K-means algorithm with the Euclidean distance metric. K-means has been shown to perform very well with embedded spaces [37], is easy to understand, and has low computational complexity. The Euclidean metric used as a loss function for the clustering is also computationally efficient, making it appropriate for large datasets, and it is also effective for lower-dimensionality embedded spaces [38]. It measures the straight-line distance between data points, which allows for a clear interpretation of how data points are assigned to clusters based on their proximity in the feature space.

2) REPRESENTATION LEARNING MODULE

The role of the representation learning module is to learn the mapping from the input raw data to the pseudo-labels generated by the clustering process. The CNN part of the module, together with the flattening and PCA transformation, realizes the embedding function ϕ, which transforms each input data point into a low-dimensional representation in the embedded space.

The CNN module of the SSDC architecture consists of four 1D convolutional layers, each followed by batch normalization, max-pooling, and *ReLU* activation layers, visualized in Figure 2. Batch normalization typically improves convergence speed [39], max-pooling reduces the dimensionality of the data through the layers while focusing on the most prominent features, and ReLU is important for catching the nonlinear dependencies in the data. Regarding the number and size of CNN filters, we followed the design principles discussed in [40] and [41], consisting of a stack of convolution layers followed by fully connected layers. Each consequent convolutional layer consists of double the number of filters of the previous layer and a vector size that is four times smaller than that of its predecessor. We increase the filter size to 7 (compared to 3 in [40]) so its size is large enough to neglect noise influence but small enough to be sensitive to the changes induced by the RAT-specific content in the FFT amplitudes.

Such an architecture totals 128,406 parameters and 0.2 *GFLOPS*, which is significantly lower (up to approx. 100 times, i.e., two orders of magnitude) compared to the 11.7 million parameters and (9 times) 1.81 *GFLOPS* of a model dedicated for spectrograms processing in [35].

B. WORKFLOW OF THE SSDC ARCHITECTURE.

The SSDC follows an iterative workflow consisting of two branches executed successively, mutually optimizing the clustering module and the weights of the representation learning module. The working of the clustering branch (marked with orange color in Figure 2) consists of:

1) The 1D CNN module, which is part of the embedding function ϕ, and is set to evaluation mode, having randomly initialized weights. This module extracts features from the input data, processing the raw FFT through the convolutional layers and providing embedded 512×1 vector representations of the samples at its output. The embedded representation is further transformed to an even lower dimensionality of 1×10 using PCA transformation [42]. The goal is to preserve only the high-variety features of the embedded space before passing it to the clustering module. The CNN module, the flattening layer, and the PCA realize the embedding function defined in Eq. (3).
2) The feature vectors serve as an input to the clustering module, realizing the clustering function ψ in our architecture. The K-means assigns pseudo-labels to the processed samples according to:

$$g(\mathbf{z}_i) = \arg\min_j \|\mathbf{z}_i - \mu_j\|, \tag{6}$$

where g represents the cluster assignment function, $\mathbf{z}_i$ stands for the feature vector, and μ_j is the centroid of cluster C_j. The centroid μ_j is defined as:

$$\mu_j = \frac{1}{|C_j|}\sum_{\mathbf{z}_i\in C_j}\mathbf{z}_i. \tag{7}$$

The K-means clustering forms clusters by minimizing the within-cluster sum of squares, which is given by:

$$\min_{\mathbf{C},k}\sum_{j=1}^{k}\sum_{\mathbf{z}_i\in C_j}\|\mathbf{z}_i-\mu_j\|^2. \tag{8}$$

The output of the K-means algorithm is the partitioned set of feature vectors $\mathbf{Z} = \{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_n\}$ into k clusters. Ideally, each cluster $C_j \subseteq \mathbf{Z}$ would represent a single RAT. The provisioning of the pseudo-labels ends the clustering module optimization as part of a single iteration.

The representation learning module is set to the train mode in the second branch of the iteration, with the corresponding flow in Figure 2 marked with blue. Considering the pseudo-labels provided by the clustering branch described above, the weights of the representation learning module (including the CNN and the FC part) are trained in a classical supervised learning procedure.

1) The representation learning [42] module maps each of the input raw FFT samples to a single class at the output of the classifier, minimizing the difference between the pseudo-labels and the predicted labels using the Cross-entropy loss function defined as:

$$\mathcal{L}_{\mathrm{CE}} = -\sum_{i=1}^{n}\sum_{k=1}^{K}\mathbf{1}_{[g(\mathbf{z}_i)=k]}\log p_{i,k}. \tag{9}$$

2) Parameters θ are updated using gradient descent and backpropagation according to

$$\theta \leftarrow \theta - \eta\nabla_\theta\mathcal{L}_{\mathrm{CE}}, \tag{10}$$

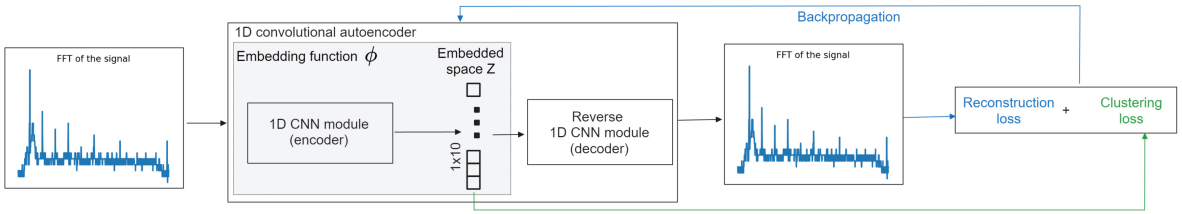


FIGURE 3. SotA autoencoder-based architecture [33].

where η is the learning rate, and $\nabla_\theta \mathcal{L}_{\text{CE}}$ represents the gradient of the loss function with respect to the network parameters θ.

The two separate procedures for optimizing the clustering and the representation learning module complete one full learning iteration. In summary, the clustering algorithm groups input data based on their distances in the embedded domain, providing pseudo-labels. The CNN and classification modules learn this distribution by enhancing the extracted features and predicting the pseudo-labels with each iteration. This iterative process between the CNN-based classification and clustering modules continues for a predefined number of training epochs, starting with randomly initialized CNN weights.

C. REFERENCE AUTOENCODER-BASED ARCHITECTURES

We compare our proposed SSDC architecture with two variants of Autoencoder (AE)-based SSL architecture, depicted in Figure 3. The AE-based architecture is composed of three main modules: encoder, embedding layer, and decoder. Regarding the problem formulation in Section IV, in this architecture, the encoder module represents the embedding function, and the clustering function works on the embedded space representations. Input for the encoder module is the raw data, which is processed by the sequential convolutional layer filters and converted into low-dimensional representation. This low-dimensional representation is flattened and transformed into a 1D vector representation in the embedding layer, which is in the middle of the architecture. The decoder block, which follows the embedding layer, has the same number of layers as the encoder and works in the opposite direction, increasing the dimensionality of the feature vector to the original size of the raw data. Providing the same dimensionality of the data at the output enables reconstruction loss in the learning process. In our work, we purposely use the same CNN module that was used in the SSDC architecture for building the AE-based architectures. This provides equal ground for comparison that will highlight the performance differences which arise from the architecture itself, and not from the variation in the trainable parameters in the representation learning module. However, the AE-based architectures, which are used as baselines, have different loss functions in the embedded space:

- Autoencoder with Modified Loss (AEML) has a custom loss function in the embedded domain based on the relative distances between the samples, as proposed in [33].
- Deep Convolutional Embedded Clustering (DCEC) utilizes the Kullback-Leibler divergence loss in the embedded domain concerning the distribution of the samples instead of their relative distances. DCEC was designed for image processing in [34].

1) WORKFLOW OF THE AE-BASED ARCHITECTURES

The training workflow of the reference AE-based models, AEML and DCEC, consists of two phases, pretraining and joint training.

1) Pre-training: The pretraining phase is the same as for the regular autoencoder. The encoder part takes high-dimensional input data and compresses it into a lower-dimensional embedded space (latent space), and the decoder part aims to reconstruct the original input data from the compressed form obtained by the encoder without considering the clustering loss. The goal is to produce a reconstruction as close as possible to the original input, thereby ensuring that the embedded space captures the essential features of the data.

During the pre-training phase, the autoencoder is trained to minimize the reconstruction error without considering the clustering loss. Thus, the weights update is performed using the reconstruction loss, which is the mean squared error (MSE) between the input x and its reconstruction $\hat{x}$, expressed as:

$$\mathcal{L}_{\text{recon}} = \frac{1}{n}\sum_{i=1}^{n} \|x_i - \hat{x}_i\|^2 \tag{11}$$

where x_i is the input data point, and $\hat{x}_i$ is the reconstructed data point.

2) Joint Training: In this second phase, the model is trained with a combined objective function that includes both the reconstruction loss, calculated as the mean squared error between the input and its reconstruction, and the clustering loss, i.e., a loss that measures how well the clustering assignments match the distribution of the data in the embedded space. This phase further refines the encoder weights to optimize both reconstruction and clusterability based on the combined loss function, expressed as:

IEEE ComSoc IEEE Open Journal of the Communications Society

$$\mathcal{L} = \alpha \mathcal{L}_{\text{recon}} + \beta \mathcal{L}_{\text{cluster}} \tag{12}$$

where α and β are hyperparameters that control the weight of the reconstruction and clustering losses, respectively.

The clustering loss for AEML consists of multiple steps and is detailed in [33]. Here we introduce its general form, defined as:

$$\mathcal{L}_{\text{cluster}}^{\text{AEML}} = \text{CustomLoss}(Z) \tag{13}$$

where Z is the set of embedded representations of the data points and CustomLoss is a function that measures clustering quality based on the relative distances between the samples.

The clustering loss for DCEC, which utilizes the Kullback-Leibler divergence, is detailed in [34]. However, we also briefly introduce it here for completeness. It is calculated as:

$$\mathcal{L}_{\text{cluster}}^{\text{DCEC}} = \text{KL}(P\|Q) = \sum_{i=1}^{n}\sum_{k=1}^{K} p_{ik} \log \frac{p_{ik}}{q_{ik}} \tag{14}$$

where P is the target distribution, Q is the predicted distribution of the clusters, and p_{ik} and q_{ik} are the probabilities that point x_i belongs to cluster k.

VI. EVALUATION METHODOLOGY

In this section, we elaborate on the methodological details of the experiments carried out to assess the performance of the proposed architecture. First, we elaborate on the training data for model development, followed by a summary of the evaluation metrics used, and we end with considerations of the evaluation approach.

A. TRAINING DATA

To evaluate the performance of the proposed architecture we selected three real-world datasets, collected from research testbeds, in particular (i) the Technology Classification Dataset - Labeled (TCD-L) [25] containing DVB-T, LTE and WiFi transmissions, (ii) the ITS-L dataset [24] containing signals from 5 different RAT, i.e., LTE, 5G, WiFi, ITS-G5 and C-V2X, and (iii) the LOG-a-TEC dataset [43] comprised of LoRa, IEEE 802.15.4, SIGFOX and some proprietary technologies. This selection of three very diverse real-world datasets from the 868 MHz, 2.4 GHz and 5.9 GHz frequency bands enables the most extensive evaluation of unsupervised methods to date.

1) TECHNOLOGY CLASSIFICATION LABELED DATASET

We purposely selected the labeled Technology Classification Dataset - Labeled (TCD-L) [25], collected from different neighborhoods located in Ghent, Belgium. During the data collection process, the transmission times and settings were also recorded, therefore this is a real-world labelled dataset. However, we only use the labels for evaluation and not for training. The data captures the influence of the different environments on the different wireless technologies that are operating in the 2.4 GHz band, namely LTE, WiFi, and DVB-T. In the original dataset, samples are collected in different bands specific to each technology, while here, we normalize them in common 1024 FFT bins as if they were coexisting in the same channel. Samples of the captured transmissions by each SSDC are depicted in Figure 4, showing that each of the RATs has its own characteristic shape in the FFT domain.

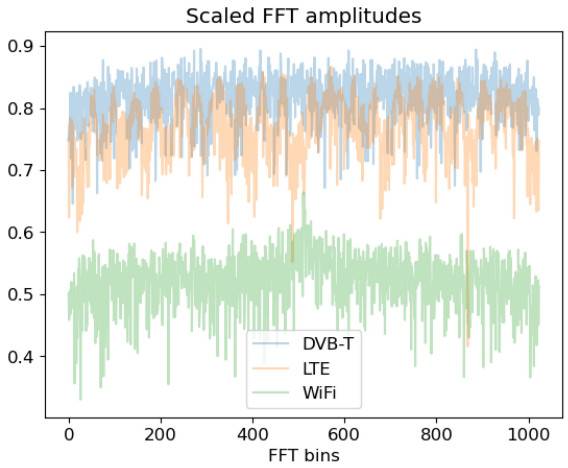


FIGURE 4. Samples for each signal type in the TCD-L dataset.

2) INTELLIGENT TRANSPORTATION SYSTEMS LABELED DATASET

The Intelligent Transportation Systems - Labeled (ITS-L) dataset [24] was collected in the region of Antwerp, Belgium. It contains signals from 5 different RATs (LTE, 5G, WiFi, ITS-G5 and C-V2X). The RAT technologies in the dataset are expected to coexist in the Intelligent Transportation Systems (ITS) 5.9 GHz band. By including the ITS-L, we expect to provide insights about the performance of the models when more types of RAT co-exist (i.e., 5 + noise compared to the 3 in the TCD-L), from which some are significantly different regarding their shape (e.g., WiFi in blue and C-V2X in red in Figure 5), and some are very similar (e.g., LTE in green and 5G in orange in Figure 5).

3) ULTRA NARROW BAND UNLABELED DATASET

Finally, we also included one unlabeled, continuously sensed dataset collected from the LOG-a-TEC testbed [43] in Ljubljana, Slovenia. This dataset consists of real-world spectrum traces in an ultra-narrow bandwidth of 192 kHz inside the European 868 MHz SRD band, which were sensed with a frequency of 5 PSD measurements per second, using 1024 FFT bins. According to an inspection of parts of the data, there are at least 4 technologies appearing in the recordings: LoRa, IEEE 802.15.4, SIGFOX, and some proprietary technology. The goal of employing such data is to check the performance of the evaluated architectures in real-world scenarios, which exposes them to a significantly wider

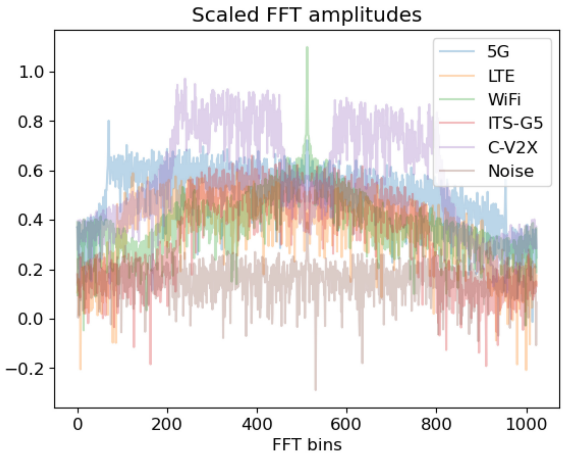


FIGURE 5. Samples for each signal type in the ITS dataset (ITS-L).

variety in the data that is being used, with all the artifacts and interference that occur during the acquisition. This makes the problem of SSDC monitoring much more challenging and closer to the potential deployment environment.

B. TRAINING CONFIGURATIONS

In this subsection, we provide the training configurations of the three architectures used in subsequent performance evaluation, the proposed SSDC architecture and the two AE-based architectures, AEML and DCEC.

1) CONFIGURATION OF THE CLUSTERING MODULE

Evaluations are performed by measuring the K-means clustering performance on the embedded vectors provided by the models created with the proposed architecture SSDC and two reference architectures, i.e., AEML and DCEC, based on multiple metrics described in the following Section VI-C. We approached the experimental evaluation assuming no information about the exact types and quantity of existing RATs in each of the datasets. Thus, for all three datasets, we set the upper number of possible RATs to 10, which is higher compared to the actual number of classes 3 and 6 in the respective labeled datasets TCD-L and ITS-L, and approximately 5 according to [43] in the unlabeled dataset UNB-U. Selection of an approximate number of classes (i.e., operating RATs) is also viable in real-world deployment based on the monitored frequency band, urban/rural area, etc. While such an approach closely imitates the real-world environment, following the problem formulation in Section IV, it also sets fair ground for comparison of the evaluated models. In this way, we avoid potential bias induced by training and evaluation with the actual number of classes, considering that the SSDC employs a clustering algorithm in its training pipeline, as detailed in Section V-A.1, while the AE-based models work without clustering feedback. Furthermore, a higher number of clusters may capture more details and variance within the data, potentially leading to more accurate representations of smaller, nuanced groups. However, this number (10 in our case) should not be too big since too many clusters might also mean overfitting to noise and outliers, making some clusters less meaningful.

2) CONFIGURATION OF THE CNN MODULES

To ensure fair comparisons across the three architectures, all models were constructed with the identical CNN module, as proposed and detailed in Section V-A.2. The implementation was written in Python programming language, using the Pytorch library. The training of the SSDC model was performed with the Adam optimizer with learning rate set to 10^{-3}, and weight decay set to 10^{-5}. The training iterations for the SSDC were set to 250 epochs. The AE-based models AEML and DCEC were configured following the description in Section V-C regarding the loss functions. The weight parameters α and β were set to 1, which gives the same importance for both components of the combined loss function, expressed with Eq. (12). The training was performed with the same configuration of the optimizer as for the SSDC model. Regarding the training iterations, considering the AE-based architectures work in two stages as described in Section V-C, the pre-training ran for 200 epochs, and the second stage of joint-training ran for 50 epochs, totaling 250 epochs, same as for the SSDC model. The embedded space dimensionality for all three models was set to 10.

C. CLUSTER QUALITY EVALUATION METRICS

For better understanding and analyzing the embedded space, we provide three different types of visualizations of the clustering in the embedded space, namely t-distributed Stochastic Neighbor Embedding (t-SNE) [44] of the embedded space, average of samples per cluster, and distribution of ground-truth classes per cluster.

Regarding the quantitative evaluation, we employ multiple existing metrics that are specific to the clustering approaches. According to the way the used metrics are calculated, they can be divided into two groups. The first group is based on labels, mainly evaluating the content/purity of the clusters regarding the different RATs. The second group of metrics is based on the distances in the embedded space, mainly considering the inter-cluster and intra-cluster distances.

1) VISUAL EVALUATION

t-SNE is useful for transforming high-dimensional data into a two-dimensional or three-dimensional space for visualization, thus revealing the structure and patterns in complex datasets. It is particularly suitable for examining the outcomes of the clustering algorithm and visualizing how various data classes are distributed. However, it is important to note that t-SNE visualizations should be taken cautiously due to their nonlinear transformation and dimensionality reduction. They may not necessarily reflect the same data structure that exists in the higher-dimensional embedded

IEEE ComSoc IEEE Open Journal of the Communications Society

space. Thus, the t-SNE visualizations will be considered as informative and complementary views, supporting the cross-interpretation of multiple metrics when drawing conclusions for the performance of such self-supervised models. Due to the large size of the data, the t-SNE embedded space visualization is performed with a randomly sampled subset of $10,000$ points instead of the entire dataset. These visualizations aim to showcase the approximate structure of the clusters created by each model within the feature space.

Calculation and visualization of an average of the samples per cluster will show what the dominant type of RAT for each cluster is and enable the expert mapping of the identified clusters to actual RAT, as described in Section III-D. The averaging will suppress the less common types of RATs and highlight the representative shape for each cluster. This visualization is particularly useful when unlabeled data is considered, which cannot be analyzed by the label-based metrics.

For the evaluation with labeled datasets, we also show the distribution of samples in each cluster regarding the ground-truth labels, providing one additional perspective of the clustering directly related to the label-based metrics described in the following section.

2) LABELS-BASED METRICS

Regarding the labels-based evaluation, we utilize 4 different metrics, namely Normalized Mutual Information (NMI), Adjusted Rand Index (ARI), Clustering homogeneity score (CHS) and Clustering completeness score (CCS).

Normalized Mutual Information is a clustering metric that quantifies the mutual information between true class labels and cluster assignments while accounting for the scale of each. It ranges from 0 to 1, with higher values indicating better agreement between true classes and clusters. Given two cluster assignments U and V, NMI is defined as:

$$\mathrm{NMI}(U, V) = \frac{2 \cdot I(U; V)}{H(U) + H(V)}, \tag{15}$$

where: $I(U; V)$ is the mutual information between U and V and $H(U)$ and $H(V)$ are the entropies of U and V, respectively.

Adjusted Rand Index [45] metric compares the pairwise decisions (whether pairs of elements are in the same or different clusters) in the clustering outcome to the true labels, adjusting for chance grouping. Formally, it is expressed as:

$$\mathrm{ARI} = \frac{\mathrm{RI} - \mathbb{E}[\mathrm{RI}]}{\max(\mathrm{RI}) - \mathbb{E}[\mathrm{RI}]}, \tag{16}$$

where RI is the Rand Index and $\mathbb{E}[\mathrm{RI}]$ is the expected value of the Rand Index. It can range from -1 (indicating completely independent labeling) to 1 (perfectly matching labeling).

The Clustering homogeneity score [46], as a clustering metric, evaluates the quality of a clustering operation, assessing whether each cluster contains only members of a single class. It is defined as:

$$\mathrm{Homogeneity} = 1 - \frac{H(C|K)}{H(C)}, \tag{17}$$

where $H(C|K)$ is the conditional entropy of the classes given the cluster assignments, and $H(C)$ is the entropy of the classes. The CHS has a range of values between 0 and 1. A score of 0 is achieved when the clusters are completely mixed, meaning that each cluster contains an equal proportion of members from different classes. A score of 1 is achieved when each cluster contains members from only one class and no class is spread across multiple clusters, representing perfect homogeneity.

The Clustering completeness score [46] is a metric evaluating the quality of clustering results in terms of how well it groups together elements of the same class. Completeness is computed based on the conditional entropy of the clusters given the class labels. It essentially measures whether all members of a given class are assigned to the same cluster, regardless of how many other classes are also present in that cluster, defined as:

$$\mathrm{Completeness} = 1 - \frac{H(K|C)}{H(K)}, \tag{18}$$

where $H(K|C)$ is the conditional entropy of the cluster assignments given the classes and $H(K)$ is the entropy of the cluster assignments. CCS score ranges from 0 to 1. A score of 0 indicates that the algorithm has dispersed members of a single class across multiple clusters, failing to group them together, and a score of 1 signifies that all members of each class are perfectly grouped within a single cluster.

All four labels-based metrics provide quantitative measures to evaluate the performance of clustering algorithms when true labels are known, which is the case for the TCD-L and ITS-L datasets. Each metric provides a different perspective of the clustering performance, considering the distribution of samples across the clusters relative to the ground-truth labels.

3) DISTANCE-BASED METRICS

For measuring the clustering performance based on the state in the embedded space, we combined metrics used in [35] and [29], i.e., silhouette score, Davies-Bouldin score, and Calinski-Harabasz index.

The silhouette score [47] is a metric used to measure the goodness of a clustering technique. It quantifies how well-separated clusters are in a dataset, ranging from -1 to 1. A higher silhouette score indicates better-defined clusters, with scores closer to 1 indicating more cohesive and separated clusters, while scores near 0 suggest overlapping clusters. A silhouette score that trends toward -1 suggests that many points in the dataset have been placed in inappropriate clusters. Considering a sample i, the average distance between i and all other points in the same cluster $a(i)$, and the minimum average distance between i and points in a

different cluster $b(i)$, minimized over clusters, the silhouette score $s(i)$ is given by:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}. \tag{19}$$

The overall silhouette score for the dataset is the mean silhouette score of all samples.

The Davies-Bouldin [48] score is a clustering evaluation metric that measures the compactness and separation of clusters. Theoretically, this score ranges from 0 upwards, with no fixed upper limit. Formal definition is as follows: considering k clusters, with C_i being the centroid of cluster i and S_i being the average distance of all points in cluster i to the centroid C_i, and similarity measure R_{ij} between clusters i and j defined as:

$$R_{ij} = \frac{S_i + S_j}{\|C_i - C_j\|}, \tag{20}$$

the Davies-Bouldin score DB is the average of the maximum R_{ij} for each cluster:

$$DB = \frac{1}{k} \sum_{i=1}^{k} \max_{j \neq i} R_{ij}. \tag{21}$$

Low scores indicate good clustering quality, where clusters are compact (data points within clusters are close to each other) and well-separated (clusters are far apart from each other).

The Calinski-Harabasz [49] index is a clustering evaluation metric that quantifies the ratio of variance between clusters to variance within clusters. The score can theoretically be as low as zero, indicating extremely poor clustering where the within-cluster variance is as high as the total variance. However, in practice, any non-trivial clustering will result in a score greater than zero. There is no theoretical upper limit to the score. Higher scores indicate better clustering performance, with more significant separation between clusters compared to the variance within clusters. Formal definition is as follows: given k clusters and n data points, B_k as between-group dispersion matrix and W_k the within-group dispersion matrix, the Calinski-Harabasz index CH is defined as:

$$CH = \frac{\text{trace}(B_k)/(k-1)}{\text{trace}(W_k)/(n-k)}, \tag{22}$$

where $\text{trace}(B_k)$ is the trace of the between-group dispersion matrix and $\text{trace}(W_k)$ is the trace of the within-group dispersion matrix.

The Davies-Bouldin, silhouette, and Calinski-Harabasz scores work based on the clustering result and inter-sample distances. Such metrics are crucial to provide a quality assessment of the clustering result when labels are absent, which is the case with the unlabeled dataset UNB-U. Furthermore, evaluation with Davies-Bouldin, silhouette, and Calinski-Harabasz can also provide valuable insights when working with labeled data because, in general, they quantify the quality of the clustering result based on the inter-cluster and intra-cluster distances, independent of the ground-truth labels.

D. TRANSMISSION DETECTION EVALUATION METRIC

Based on the mapping of the clusters to specific RAT classes, performed as described in Section III-C, we can evaluate the monitoring by calculating the standard multiclass classification evaluation metrics *precision*, *recall* and *F1 score*. Since the problem of monitoring at this stage could be interpreted as multiclass classification, the evaluation with the aforementioned metrics is performed per class.

Thus, considering $\mathcal{C} = \{1, 2, \ldots, K\}$ is the set of classes, for each class $k \in \mathcal{C}$:

- Precision P_k is the ratio of true positives to the sum of true positives and false positives for class k:

$$P_k = \frac{\text{TP}_k}{\text{TP}_k + \text{FP}_k}. \tag{23}$$

- Recall R_k is the ratio of true positives to the sum of true positives and false negatives for class k:

$$R_k = \frac{\text{TP}_k}{\text{TP}_k + \text{FN}_k}. \tag{24}$$

- F1 score for class k is the harmonic mean of precision and recall for class k:

$$F1_k = \frac{2 \cdot P_k \cdot R_k}{P_k + R_k}. \tag{25}$$

This evaluation is performed using the models trained with the two labeled datasets, TCD-L and ITS, since these metrics necessitate ground-truth labels. For predicted labels, we use the identified labels for each formed cluster based on the visual evaluation. In the case of having clusters that contain multiple classes, we identify them based on the most dominant type of samples.

E. COMPUTATIONAL PERFORMANCE METRICS

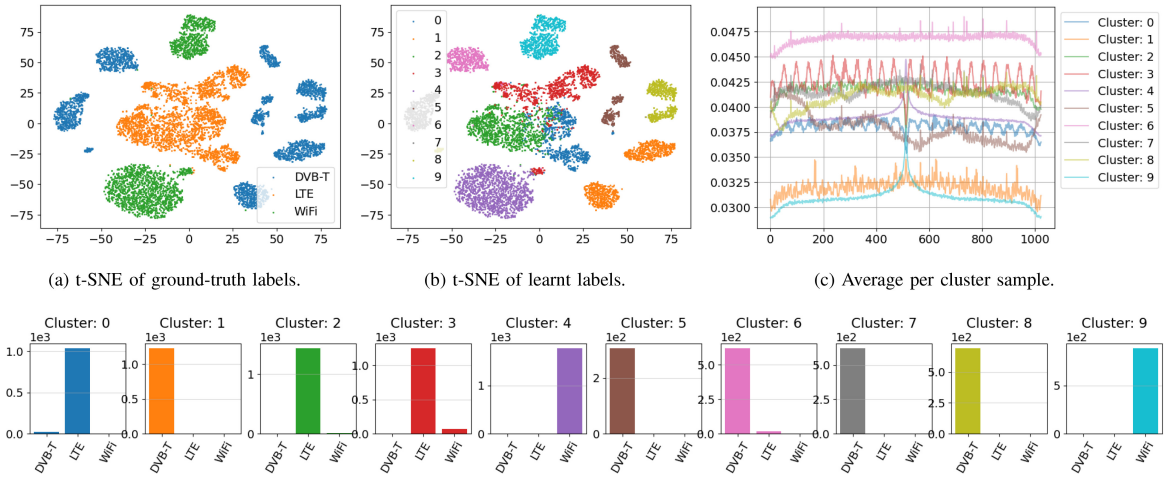
As part of the evaluation, we also assess the potential deployment challenges of the models regarding their structure and size based on the number of trainable parameters and computational requirements based on the calculation of one billion floating-point operations per second (GFLOPS). GFLOPS is a measure of computational performance required or achieved by the model during the training and inference phases. It quantifies the number of floating-point operations (FLOPs) a model needs to execute per second in the billions (giga-). The calculations of the GFLOPS were carried out using the **fvcore**[2] library.

VII. RESULTS

The following results show the evaluation of models obtained from each of the three architectures, the proposed SSDC and the AEML and DCEC reference SotA architectures, with three different datasets, each of them addressing a specific challenge, as described in Section VI-A.

[2]https://github.com/facebookresearch/fvcore

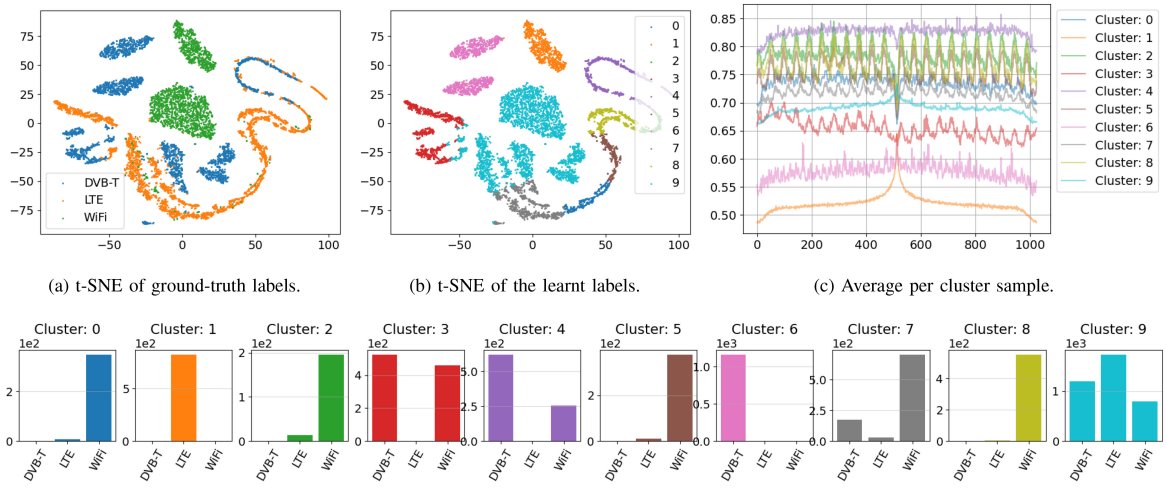
IEEE ComSoc IEEE Open Journal of the Communications Society



(a) t-SNE of ground-truth labels. (b) t-SNE of learnt labels. (c) Average per cluster sample.

(d) Per cluster distribution of ground-truth labels.

FIGURE 6. Evaluation of the model learnt with the SSDC architecture and TCD-L dataset.



(a) t-SNE of ground-truth labels. (b) t-SNE of the learnt labels. (c) Average per cluster sample.

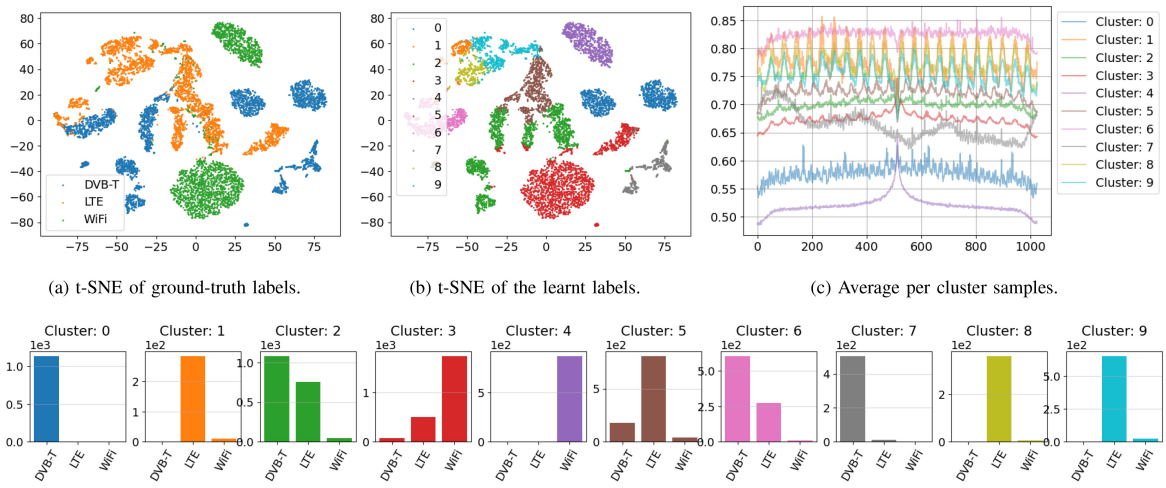
(d) Per cluster distribution of ground-truth labels.

FIGURE 7. Evaluation of the model learnt with the AEML architecture and TCD-L dataset.

A. EVALUATION RESULTS WITH TCD-L

1) VISUAL ANALYSIS

Figures 6(a), 7(a) and 8(a) depict the two-dimensional t-SNE projected features learnt by each of the three models developed. The number of depicted clusters is 10, as selected in Section VI-A, while the colors represent the three types of RAT, namely DVB-T (blue), LTE (orange), and WiFi (green). From Figure 6(a), it can be seen that while the model learns to create 10 clusters due to the clustering parameter selection, these clusters are well separated and homogeneous, with several of them containing the same technology. For instance, there are several blue clusters, all containing DVB-T, one large orange cluster containing LTE, and two large and one small green clusters containing WiFi. Although each model provides the same number of clusters, their shapes vary significantly. As can be seen in Figure 6(a), the SSDC model shows more tendency towards the creation of clusters with elliptic shapes, while the other two, the AEML in Figure 7(a) and DCEC in Figure 8(a) have more variety in the cluster shapes. The distribution of the data input into the clustering methods, i.e., the distribution of the learnt embeddings in our case, is less suitable for well-shaped spheric clusters for the AEML and DCEC models compared to the SSDC. These two characteristics could be considered important advantages for the SSDC model since it provides more "clustering-friendly" feature space.



(a) t-SNE of ground-truth labels. (b) t-SNE of the learnt labels. (c) Average per cluster samples.

(d) Per cluster distribution of ground-truth labels.

FIGURE 8. Evaluation of the model learnt with the DCEC architecture and TCD-L dataset.

In Figures 6(b), 7(b) and 8(b), we show the same feature space embedding, but this time colored with the 10 labels corresponding to the 10 clusters that were learnt by each model. As can be seen in the figures, each of the formed clusters is largely homogeneous, consisting of samples mostly of a single RAT technology. Looking at the largest cluster in Figure 6(b), we know from the ground-truth depiction in Figure 6(a) that it contains only LTE transmissions. It can be seen that it learnt to separate it into three different subclusters (0-blue, 2-green, and 3-red), where each of them contains samples from the same technology. This is also depicted by the clusters' distribution plot in Figure 6(d), with the same coloring as for the K-means result in Figure 6(b). Thus, the averaged clusters' samples in Figure 6(c), also colored with the corresponding colors of the 10 clusters provided by the K-means, are aligned with this, showing the specific shapes for each of the technologies. The additional insight provided by the average plot is that the K-means groups the samples based on the RAT technology, where signal strength has a significant impact on the features, thus leading to the creation of separate clusters of the same technology.

Looking at the results of the reference autoencoder-based models depicted in Figures 7 and 8 for the AEML and DCEC respectively, they seem to have more condensed clusters in the feature space compared to the proposed SSDC model, which means that the samples of each technology in Figures 7(a) and 8(a) are less dispersed in the feature space. However, the fragmented and free forms of the clusters lead to greater confusion/misclassification when performing the K-means clustering. Thus, the resulting K-means clusters, according to Figures 7(b) and 8(b), contain samples from different technologies. For example, cluster 3-red and cluster 9-cyan in Figure 7, which are part of the elongated groups of samples, contain both DVB-T and LTE transmissions. A similar effect of clusters containing a mix of RATs is even more prominent in the visualizations for the DCEC model instance, as can be seen in Figure 8. Clusters learnt by the K-means algorithm in the DCEC provided feature space, visualized in Figure 8(b) with assigned labels 2-green, 5-brown, and 6-pink, contain samples of DVB-T and LTE, and cluster 3-red contains LTE and WiFi, as can also be seen on the per cluster distribution of ground-truth labels in Figure 8(d). This phenomenon is also leading to less distinguishable average samples in Figure 8(c). For example, the shape of cluster 3-red, which has a dominant group of WiFi samples, does not have a smooth characteristic shape, which can be noticed for the average of cluster 4-purple, which has WiFi samples only.

2) QUANTITATIVE ANALYSIS

1) Evaluation with labels-based metrics: As it was described in Section VI, besides the analysis based on visual evaluation of the clustering results, we also performed a quantitative evaluation based on multiple clustering-specific metrics. The results are summarized in Table 1, which lists the evaluation metrics in the first column, followed by the models obtained with the proposed architecture in column two and the reference autoencoder-based architectures in columns three and four, respectively. From the first row of Table 1, it can be seen that NMI of the proposed SSDC is almost double the value of the reference AE-based models indicating stronger mutual dependence between the clustering results and the true labels, considering that 0 NMI means no dependence. The clustering assignments produced by K-means on the SSDC-provided feature-space more successfully identified the inherent groupings within the data that align closely with the ground truth categories, compared to the K-means

IEEE ComSoc IEEE Open Journal of the Communications Society

TABLE 1. Evaluation of clustering by K-means on the different embedded spaces with TCD-L.

Metric \ Model	SSDC	AEML	DCEC
NMI ↑	**0.6209**	0.3266	0.3998
ARI ↑	**0.3874**	0.1362	0.23
CHS ↑	**0.9465**	0.4595	0.5897
CCS ↑	**0.4620**	0.2533	0.3024
Silhouette ↑	0.3174	**0.6050**	0.4325
Davies-Bouldin ↓	1.8727	**0.5286**	1.0396
Calinski-Harabasz ↑	18029	**1259829**	357440

Legend: ↑ - higher the better, ↓ - lower the better.
Labels-based metrics, Distance-based metrics

TABLE 2. Performance of different models on multiclass classification.

Metric	Class	SSDC	AEML	DCEC
Precision	DVB-T	**0.99**	0.76	0.89
	LTE	**0.97**	0.89	0.81
	WiFi	**1.00**	0.57	0.94
Recall	DVB-T	**1.00**	0.63	0.79
	LTE	**1.00**	0.59	0.88
	WiFi	0.96	**0.97**	0.96
F-score	DVB-T	**1.00**	0.69	0.84
	LTE	**0.98**	0.71	0.84
	WiFi	**0.98**	0.72	0.95

Higher is better.

assignments on the feature-space of the reference AE-based models AEML and DCEC.

Considering the ARI metric values in Table 1, all three models have values above 0, meaning that their developed clusters are better than random clustering. Looking at each model's score separately, SSDC leads by a margin of 0.14 to the second DCEC. This indicates that SSDC has the greatest similarity between the clustering assignments and the ground-truth labels. The CHS scores show very high values for SSDC, meaning that the formed clusters contain mostly samples from a single class (RAT), while in the embedded spaces of the reference AE-based models, there is much more variety in the clusters, supporting the conclusions made from the visual analysis. Based on the CCS metric, the proposed SSDC is again much better (by a factor of 1.5 to 1.8). However, the values of all three models are in general low (the highest 0.4578 for SSDC) due to a large number of clusters (10) compared to the number of existing RATs (3) in this dataset.

In general, SSDC is superior according to the labels-based evaluation, DCEC is second, and the worst performance is achieved with the AEML model.

2) Evaluation with distance-based metrics: While all the labels-based metrics have shown a significant advantage for the proposed SSDC model, the distance-based metrics silhouette, Davies-Bouldin and Calinski-Harabasz (last three rows in Table 1) show a clear advantage of AEML. The higher silhouette score of the AEML model indicates better-separated clusters, meaning that there is better cohesion of the inter-cluster samples and better separation between clusters, compared to SSDC and DCEC. The advantage of better separation between the clusters and compactness of each cluster is already seen in Figures 6-8. The reference AE-based models provide more dense clusters, with larger distances between them, while SSDC has more dispersed clusters. Davies-Bouldin and Calinski-Harabasz scores (last two rows in Table 1), which relate to the within-cluster dispersion and between-cluster separation, again indicate the more compact and better-separated clusters of the AEML model compared to SSDC and DCEC.

In general, both the labels-based and distance-based metrics are in line with the observations in the visual analysis of the embedded space (Figures 6, 7 and 8) where better-separated feature space was provided by the AEML. However, the clusters were much cleaner with the proposed SSDC model mostly having single RAT per cluster. This means that the SSDC model successfully learns more representative features for the different RATs compared to the reference AEML and DCEC models, and also exhibits better generalization capabilities, considering that the TCD-L dataset is acquired across multiple environments.

3) Spectrum Monitoring Evaluation: As described in Section VI-D, we evaluate the spectrum monitoring with the different models based on the identified classes, using the standard classification metrics, precision, accuracy, and F1-score. The results are summarized in Table 2. Precision, recall, and F1-score are calculated per class for the three existing classes in the TCD-L dataset. The SSDC model outperforms the AE-based models by a significant margin, up to 40 ppt, in all but one case. There is a negligible difference in the Recall of the WiFi class, where the AEML model has 1 ppt better performance over the other two, SSDC and DCEC. This evaluation again confirms the superior generalization capabilities of the SSDC model compared to the AE-based models, considering that the TCD-L dataset contains data from different environments, as described in Section VI-A, and the potential to use it for training a classifier without labeled data.

B. EVALUATION RESULTS WITH ITS-L

Following the methodology detailed in Section VI, in this paragraph, we analyze the results of the evaluation with the second labeled dataset ITS-L. The evaluation is performed in the same order as for the previous TCD-L dataset, with a focus on the quantitative analysis.

1) EVALUATION WITH LABELS-BASED AND DISTANCE-BASED METRICS

The performance of the models is depicted in Table 3, where again, the first four rows with blue background show the performance measured with the labels-based matrices, and

TABLE 3. Evaluation of clustering by K-means on different embedded spaces, ITS-L.

Metric \ Model	SSDC	AEML	DCEC
NMI ↑	**0.8884**	0.3266	0.8478
ARI ↑	**0.7902**	0.1362	0.7887
CHS ↑	**0.9955**	0.4595	0.9405
CCS ↑	**0.8022**	0.2533	0.7718
Silhouette ↑	0.5892	0.6050	**0.6819**
Davies-Bouldin ↓	0.6953	**0.5286**	1.0396
Calinski-Harabasz ↑	309780	**1259829**	1102317

Legend: ↑ - higher the better, ↓ - lower the better.
Labels-based metrics, Distance-based metrics

TABLE 4. Performance of different models on multiclass classification.

Metric	Class	SSDC	AEML	DCEC
Precision	CV2X	0.94	**1.00**	**1.00**
	5G	**1.00**	0.99	0.95
	ITSG5	**1.00**	0.89	0.99
	LTE	**1.00**	0.93	0.99
	WiFi	**1.00**	0.97	0.98
	Noise	**1.00**	0.88	**1.00**
Recall	CV2X	**1.00**	0.99	0.99
	5G	**1.00**	0.89	0.98
	ITSG5	**0.96**	**0.96**	0.94
	LTE	0.99	0.96	**1.00**
	WiFi	0.98	0.83	**0.99**
	Noise	**1.00**	**1.00**	**1.00**
F1-score	CV2X	0.97	**0.99**	**0.99**
	5G	**1.00**	0.94	0.97
	ITSG5	**0.98**	0.92	0.96
	LTE	**1.00**	0.94	0.99
	WiFi	**1.00**	0.90	0.99
	Noise	**1.00**	0.94	**1.00**

Higher is better.

the last three rows with orange background show the values of the three distance-based metrics.

Regardless of the larger number (6) of different types of signals (5 RATs + noise) in this dataset, the performance of all three models follows a similar pattern as in the evaluation with the TCD-L dataset with 3 RATs. Regarding the labels-based metrics, the SSDC model outperforms the other two across all four metrics. DCEC is very close to the SSDC on the labels-based metrics, whereas AEML performs quite badly, which was not the case for the evaluation with the TCD-L dataset. The relatively high performance of the DCEC means that the AE structure is capable of capturing the different shapes of the signals. However, different loss functions used in the two AE-based architectures have led to the creation of different structures in the embedded domain. The deep clustering component of the loss function of the AEML model is more dominant compared to the reconstruction loss, leading to very compact clusters in the embedded domain while minimizing the influence of the shape of the samples in the learning process. This is confirmed by the high performance on the distance-based metrics of the AEML model, providing compact and well-distanced clusters containing samples from different RATs. The SSDC model shows the most balanced performance, with comparably distinguishable clusters as the AE-based models, which are also homogeneous, containing mostly single RAT samples. Overall, the SSDC successfully captures the RAT-specific features when there is a larger number of RATs, outperforming the AE-based models.

1) Spectrum Monitoring Evaluation: Spectrum monitoring performance of the three models on the ITS dataset, regarding the precision, recall, and F1-score, is summarized in Table 4. The highest scores are bolded in the table. While the SSDC model achieves the highest score in most cases across the evaluation with the three metrics across the six classes, it is evident that it has a very similar performance with the DCEC model. In general, we can notice a similar performance pattern as in the previous labels-based evaluation, demonstrated in Table 3. This means that the SSDC and DCEC can provide high-performance classifiers using completely unlabeled data, with only providing identification of the clusters, as part of the Phase 3, described in Section III-C.

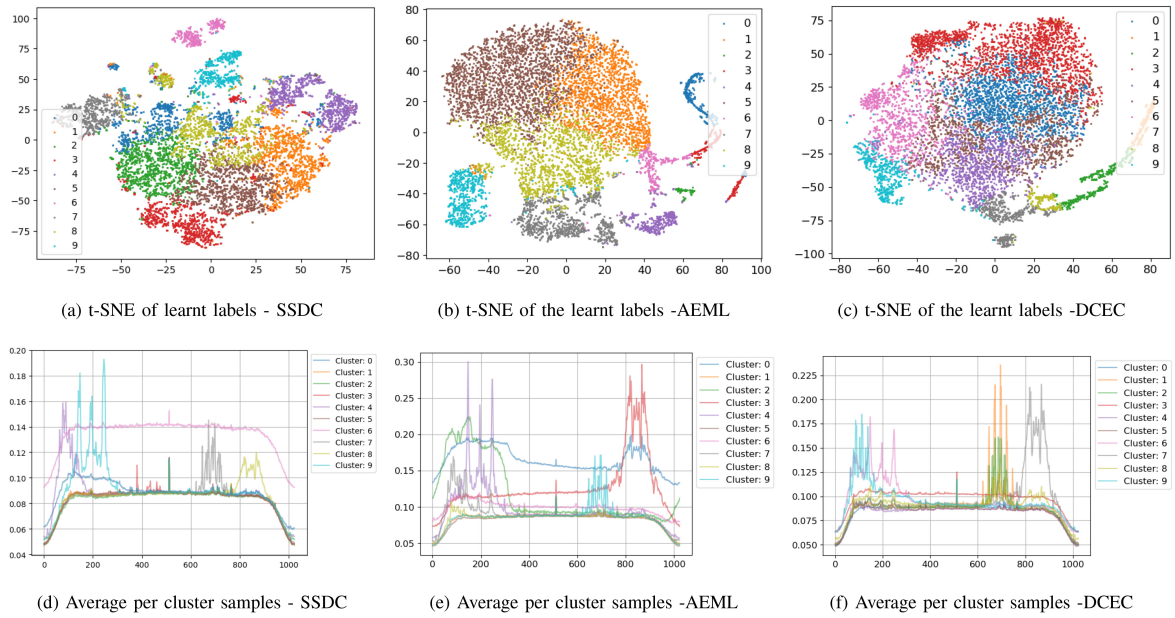
C. EVALUATION RESULTS WITH UNB-U

1) VISUAL EVALUATION

Figure 9 depicts the visual evaluation of the three models with the unlabeled dataset. In this case, since the ground-truth labels are not available, we can only analyze as part of the visual evaluation the t-SNE projections (Figures 9(a), 9(b) and 9(c)) and the cluster averages (Figures 9(d), 9(e) and 9(f)). Considering that this dataset is acquired by continuous sensing, we expect it to have many noisy samples without clear distinction boundaries between them. Also, if there are significant amounts of samples with specific shapes, which would result from some specific RAT operating in the monitored band, we expect them to be grouped together in a separate cluster in the feature space.

Comparing the t-SNE visualizations, the feature space of the SSDC model appears to be separated into more distinctive and better-rounded clusters as shown in Figure 9(a) compared to AE-based models shown in Figures 9(b) and 9(c). This makes the SSDC-provided feature space advantageous for further analysis since it is easier to isolate each specific type of RAT, regardless of the type of clustering algorithm being used. Besides, the AE-based models contain some well-separated clusters, such as the cluster 9-cyan in Figure 9(b). There are also the elongated shapes that result in merged clusters that contain different RATs, such as the clusters 3-red and 0-blue in Figure 9(b) and clusters 2-green and 8-yellow in Figure 9(c). The same effect was also noticed in the evaluation with the TCD-L labeled dataset in Section VII-A.

IEEE ComSoc IEEE Open Journal of the Communications Society



(a) t-SNE of learnt labels - SSDC (b) t-SNE of the learnt labels -AEML (c) t-SNE of the learnt labels -DCEC

(d) Average per cluster samples - SSDC (e) Average per cluster samples -AEML (f) Average per cluster samples -DCEC

FIGURE 9. Feature-space and average clusters of the different models.

These observations are also confirmed with the per-cluster averages depicted in Figures 9(d), 9(e) and 9(f). The well-distinguished clusters are easily recognizable shapes that are specific for different RATs. This allows for further fine-grain analysis of the smaller clusters, which would not be possible for the other two, AEML and DCEC, which show only 3-5 distinguishable clusters. Looking at the averages of the clusters of each model, all of them appear to have formed clusters for mostly similar transmission types. However, the merged clusters which were notable in the feature space visualization in Figures 9(b) and 9(c) are also influencing the per-cluster averages in Figures 9(e) (e.g., cluster 0-blue and cluster 3-red) and 9(f) (e.g., cluster 6-pink and cluster 9-cyan).

Based on analysis of the same dataset (UNB-U) in [43], we can tell which RATs represent some of the formed clusters based on the shape of the per-cluster average. In Figure 9(d), cluster 6-pink are IEEE 802.15.4 transmissions, clusters 7-gray and 9-cyan are two different proprietary transmissions, and cluster 4-purple are LoRA transmissions.

2) QUANTITATIVE EVALUATION

Quantitative evaluation of the three models with UNB-U dataset with the distance-based metrics is summarized in Table 5. Results support the previous conclusions based on the visual analysis of the feature space and the per-cluster averages. The existence of some well-separated clusters has led the SSDC model to significantly outperform the other two AE-based reference models according to the silhouette score, i.e., AEML by almost 40% and DCEC by 150%, which was not the case in the labeled datasets. It also shows

TABLE 5. Evaluation of clustering by K-means on the different embedded spaces with UNB-U.

Metric \ Model	SSDC	AEML	DCEC
Silhouette ↑	**0.2056**	0.1484	0.0825
Davies-Bouldin ↓	1.5218	**1.3183**	1.631
Calinski-Harabasz ↑	57037	71520	**363927**

Legend: ↑ - higher the better, ↓ - lower the better..

comparable performance based on the Davies-Bouldin score by ranking second. The relatively good performance of AE-based models results from the larger density of the formed clusters, which was also noticed in the evaluations with the labeled datasets. It is interesting to notice how the DCEC model performs significantly better on the Calinski-Harabasz score, which results from the good separation and density of some of the clusters on the margins of the feature space (i.e., cluster 1-orange, 9(c)).

Overall, judging by the distance-based metrics, all three models show comparable performance with the UNB-U unlabeled dataset. However, based on the visual evaluation, the clusters of the SSDC model appear to form a larger number of recognizable transmission types; thus, the SSDC model has more potential for further analysis because of the better segmentation of the feature space.

D. POTENTIAL DEPLOYMENT CHALLENGES

Regarding the architectural specifics, the SSDC model could be considered easier to configure (number of training epochs

TABLE 6. Computational complexity of the models.

Metric \ Model	SSDC	AEML / DCEC
Num. params.	**128406**	162827
GFLOPS	**0.196672**	0.38731776

and stopping criteria) compared to the reference AE-based models because of the following main characteristics.

1) SINGLE LOSS FUNCTION

Having two separate components that build the global loss function for the AEML and DCEC could lead to conflicting situations in the learning process. For example, lowering the deep clustering part of the loss in the embedded domain (increasing the distance between samples) could lead to worsening the reconstruction loss. Additionally, in AE-based models, there is no direct feedback from the actual classification process, such as the K-means for the SSDC model, which means that actual classification performance could be even degraded by creating unnecessary clusters and fragmenting the feature space while improving the deep clustering and reconstruction loss.

2) INHERENTLY DEVELOPED CLASSIFIER

The training of the SSDC inherently develops a classifier in parallel with the feature learning, which follows the assignments of the K-means algorithm. On the contrary, the AE-based models require another classification algorithm to work on the extracted features since they only provide encoding of the input data without class information.

3) LOWER COMPUTATIONAL COMPLEXITY

As we described in Section VI, the reference AE-based and the proposed SSDC architectures utilize the same CNN module. Consequently, during the inference mode (deployment scenario), all of the models will have a similar number of parameters. However, in the training mode, the SSDC models have around 22% fewer parameters than AE-based models because of the encoder-decoder structure, which requires symmetrical CNN module for decoding. Considering the GFLOPS calculation, as specified in Section VI, the proposed SSDC architecture requires approximately half of the amount of GFLOPS compared to the AE-based architecture. This means that tuning or retraining of the SSDC model will require significantly less computational resources, as summarized in Table 6. Smaller model sizes and lower computational requirements of the proposed SSDC model could significantly benefit the potential deployment in more restricted devices closer to the edge of the network, thus extending the range of potential applications.

VIII. CONCLUSION

In this paper, we investigated label-free RAT classification for spectrum sensing. We proposed a new spectrum sensing workflow, adequate for realistic set-ups that do not collect labels, based on a novel deep Self-Supervised Learning (SSL), that can be easily ported to various environments. This workflow enables the development of RAT classification models while avoiding the necessity of large labeled datasets. The SSL architecture is capable of autonomously learning low-dimensional features from raw FFT and enables the identification of different RATs without prior knowledge of the specific RATs in the monitored environment. The model is trained in a self-supervised manner, followed by manual cluster labeling by a human expert.

We evaluated the proposed architecture against state-of-the-art Autoencoder (AE)-based architectures using three real-world datasets acquired from three different frequency bands, 2.4 GHz, 5.9 GHz, and 868 MHz, containing signals from over ten different RATs. The trained model using the proposed architecture consistently outperformed the AE-based models by up to 31 ppt F1 score according to spectrum monitoring evaluation, while at the same time requiring 22% fewer trainable parameters and 50% fewer floating-point operations per second (FLOPS).

REFERENCES

[1] (Ericsson, Stockholm, Sweden). *Mobility Report*. Nov. 2023. Accessed: Feb. 20, 2025. [Online]. Available: https://ericsson.com/mobility-report

[2] W. K. Alsaedi, H. Ahmadi, Z. Khan, and D. Grace, "Spectrum options and allocations for 6G: A regulatory and standardization review," *IEEE Open J. Commun. Soc.*, vol. 4, pp. 1787–1812, 2023.

[3] R. K. Saha and J. M. Cioffi, "Dynamic spectrum sharing for 5G NR and 4G LTE coexistence-a comprehensive review," *IEEE Open J. Commun. Soc.*, vol. 5, pp. 795–835, 2024.

[4] M. Parvini et al., "Spectrum sharing schemes from 4G to 5G and beyond: Protocol flow, regulation, ecosystem, economic," *IEEE Open J. Commun. Soc.*, vol. 4, pp. 464–517, 2023.

[5] L. J. Wong, W. H. Clark, B. Flowers, R. M. Buehrer, W. C. Headley, and A. J. Michaels, "An RFML ecosystem: Considerations for the application of deep learning to spectrum situational awareness," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 2243–2264, 2021.

[6] Z. Ke and H. Vikalo, "Real-time radio technology and modulation classification via an LSTM auto-encoder," *IEEE Trans. Wireless Commun.*, vol. 21, no. 1, pp. 370–382, Jan. 2022.

[7] S. Rajendran, W. Meert, D. Giustiniano, V. Lenders, and S. Pollin, "Deep learning models for wireless signal classification with distributed low-cost spectrum sensors," *IEEE Trans. Cogn. Commun. Netw.*, vol. 4, no. 3, pp. 433–445, Sep. 2018.

[8] I. G. A. Poornima and B. Paramasivan, "Anomaly detection in wireless sensor network using machine learning algorithm," *Comput. Commun.*, vol. 151, pp. 331–337, Feb. 2020.

[9] Y. Ghanney and W. Ajib, "Radio frequency interference detection using deep learning," in *Proc. IEEE 91st Veh. Technol. Conf.*, 2020, pp. 1–5.

[10] J. Robinson, S. Kuzdeba, J. Stankowicz, and J. M. Carmack, "Dilated causal convolutional model for RF fingerprinting," in *Proc. 10th Annu. Comput. Commun. Workshop Conf. (CCWC)*, 2020, pp. 157–162.

[11] U. Demirhan and A. Alkhateeb, "Integrated sensing and communication for 6G: Ten key machine learning roles," *IEEE Commun. Mag.*, vol. 61, no. 5, pp. 113–119, May 2023.

[12] C. Ouyang, Y. Liu, H. Yang, and N. Al-Dhahir, "Integrated sensing and communications: A mutual information-based framework," *IEEE Commun. Mag.*, vol. 61, no. 5, pp. 26–32, May 2023.

[13] Z. Wei et al., "Integrated sensing and communication signals towards 5G-A and 6G: A survey," *IEEE Internet Things J.*, vol. 10, no. 13, pp. 11068–11092, Jul. 2023.

[14] S. Tadik et al., "Digital spectrum twins for enhanced spectrum sharing and other radio applications," presented at the IEEE Int. Conf. Digit. Twins Parallel Intell., Jan. 2023.

[15] A. S. Abdalla, P. S. Upadhyaya, V. K. Shah, and V. Marojevic, "Toward next generation open radio access networks: What O-RAN can and cannot do!" *IEEE Netw.*, vol. 36, no. 6, pp. 206–213, Nov./Dec. 2022.
[16] C. Coletti et al., "O-RAN: Towards an open and smart RAN," O-RAN ALLIANCE, Alfter, Germany, White Paper, 2018.
[17] L. Bonati, M. Polese, S. D'Oro, S. Basagni, and T. Melodia, "Open, programmable, and virtualized 5G networks: State-of-the-art and the road ahead," *Comput. Netw.*, vol. 182, Dec. 2020, Art. no. 107516.
[18] M. Hoffmann and P. Kryszkiewicz, "Beam management driven by radio environment maps in O-RAN architecture," 2023, *arXiv:2303.11742*.
[19] S. D'Oro, L. Bonati, M. Polese, and T. Melodia, "OrchestRAN: Network automation through orchestrated intelligence in the open ran," in *Proc. IEEE Conf. Comput. Commun.*, 2022, pp. 270–279.
[20] G. D. Durgin, M. A. Varner, N. Patwari, S. K. Kasera, and J. Van der Merwe, "Digital spectrum twinning for next-generation spectrum management and metering," in *Proc. IEEE 2nd Int. Conf. Digit. Twins Parallel Intell. (DTPI)*, 2022, pp. 1–6.
[21] W. Zhang, M. Feng, M. Krunz, and A. H. Y. Abyaneh, "Signal detection and classification in shared spectrum: A deep learning approach," in *Proc. IEEE Conf. Comput. Commun.*, 2021, pp. 1–10.
[22] A. Scalingi, D. Giustiniano, R. Calvo-Palomino, N. Apostolakis, and G. Bovet, "A framework for wireless technology classification using crowdsensing platforms," in *Proc. IEEE Conf. Comput. Commun.*, 2023, pp. 1–10.
[23] H. Han, W. Li, Z. Feng, G. Fang, Y. Xu, and Y. Xu, "Proceed from known to unknown: Jamming pattern recognition under open-set setting," *IEEE wireless Commun. Lett.*, vol. 11, no. 4, pp. 693–697, Apr. 2022.
[24] M. Girmay, V. Maglogiannis, D. Naudts, M. Aslam, A. Shahid, and I. Moerman, "Technology recognition and traffic characterization for wireless technologies in ITS band," *Veh. Commun.*, vol. 39, Feb. 2023, Art. no. 100563.
[25] J. Fontaine et al., "Towards low-complexity wireless technology classification across multiple environments," *Ad Hoc Netw.*, vol. 91, Aug. 2019, Art. no. 101881.
[26] Y. Ren et al., "Deep clustering: A comprehensive survey," 2022, *arXiv:2210.04142*.
[27] N. Zhang, J. Shen, Y. Shi, and Y. Li, "CNN-Zero: A zero-shot learning framework for jamming identification," in *Proc. IEEE 22nd Int. Conf. Commun. Technol. (ICCT)*, 2022, pp. 1126–1131.
[28] M. Caron, P. Bojanowski, A. Joulin, and M. Douze, "Deep clustering for unsupervised learning of visual features," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 132–149.
[29] G. Cerar, B. Bertalanič, M. Mohorčič, and C. Fortuna, "Learning to detect wireless spectrum occupancy using clustering approaches," in *Proc. 19th Int. Conf. Wireless Mobile Comput., Netw. Commun. (WiMob)*, 2023, pp. 143–148.
[30] A. Elsebaay and H. H. Refai, "Wireless technology identification employing dynamic mode decomposition spectrum," in *Proc. Int. Telecommun. Conf. (ITC-Egypt)*, 2023, pp. 134–139.
[31] Q. Meng, H. Qian, Y. Liu, Y. Xu, Z. Shen, and L. Cui, "Unsupervised representation learning for time series: A review," 2023, *arXiv:2308.01578*.
[32] Q. Ma, J. Zheng, S. Li, and G. W. Cottrell, "Learning representations for time series clustering," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 1–11.
[33] H. Zhou, J. Bai, Y. Wang, J. Ren, X. Yang, and L. Jiao, "Deep radio signal clustering with interpretability analysis based on saliency map," *Digit. Commun. Netw.*, to be published.
[34] X. Guo, X. Liu, E. Zhu, and J. Yin, "Deep clustering with convolutional autoencoders," in *Proc. Int. Conf. Neural Inf. Process.*, 2017, pp. 373–382.
[35] L. Milosheski, G. Cerar, B. Bertalanič, C. Fortuna, and M. Mohorčič, "Self-supervised learning for clustering of wireless spectrum activity," *Comput. Commun.*, vol. 212, pp. 353–365, Dec. 2023.
[36] Z. Luo et al., "Spectrum sensing everywhere: Wide-band spectrum sensing with low-cost UWB nodes," *IEEE/ACM Trans. Netw.*, vol. 32, no. 3, pp. 2112–2127, Jun. 2024.
[37] J. MacQueen et al., "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. Math. Statist. Probab.*, vol. 1, 1967, pp. 281–297.
[38] C. C. Aggarwal, A. Hinneburg, and D. A. Keim, "On the surprising Behavior of distance metrics in high dimensional space," in *Proc. 8th Int. Conf. Database Theory (ICDT)*, 2001, pp. 420–434.
[39] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.
[40] T. J. O'Shea, T. Roy, and T. C. Clancy, "Over-the-air deep learning based radio signal classification," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 168–179, Feb. 2018.
[41] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
[42] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.
[43] T. Gale, T. Šolc, R.-A. Moşoi, M. Mohorčič, and C. Fortuna, "Automatic detection of wireless transmissions," *IEEE Access*, vol. 8, pp. 24370–24384, 2020.
[44] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 2579–2605, 2008.
[45] L. Hubert and P. Arabie, "Comparing partitions," *J. Classif.*, vol. 2, pp. 193–218, Dec. 1985.
[46] A. Rosenberg and J. Hirschberg, "V-measure: A conditional entropy-based external cluster evaluation measure," in *Proc. Joint Conf. Empir. Methods Nat. Lang. Process. Comput. Nat. Lang. Learn. (EMNLP-CoNLL)*, 2007, pp. 410–420.
[47] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, pp. 53–65, Nov. 1987.
[48] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Trans. pattern Anal. Mach. Intell.*, no. 2, pp. 224–227, Apr. 1979.
[49] T. Caliński and J. Harabasz, "A dendrite method for cluster analysis," *Commun. Statist.-Theory Methods*, vol. 3, no. 1, pp. 1–27, 1974.

Chapter 5

Conclusions and Future Work Directions

In this dissertation, we set out to advance spectrum sensing through the investigation and development of unsupervised deep learning approaches, with particular emphasis on deep clustering, for label-free radio signal recognition in realistic wireless environments. To this end, three hypotheses were formulated and investigated, and the findings are summarised in the conclusions and future work directions presented in this chapter.

5.1 Conclusions

Pursuing the stated hypotheses, the research work in this dissertation resulted in four major contributions that collectively demonstrate deep clustering as an effective, scalable, and practically deployable paradigm for RAT detection and classification without reliance on labeled RF data.

UDL approaches for wireless signal recognition: We carried out a systematic and critical survey of unsupervised deep learning for wireless signal recognition, providing a novel dual-axis mapping of recognition challenges to UDL families. Per-challenge analysis of validation approaches exposes critical limitations, including fragmented evaluation practices, limited dataset diversity, and absent benchmarking standards, and derives concrete directions for future research, directly motivating the contributions pursued in this dissertation.

Spectrogram-based deep clustering architecture: We designed a spectrogram-based deep clustering architecture for RAT recognition through systematic adaptation of CNN and transformer-based modules to the spectral domain and carried out their comparative benchmarking. With domain-specific refinements we achieved reduction of model complexity by one order of magnitude, while preserving or improving clustering performance by up to 2.5 times.

Label-free spectrum sensing workflow: We designed a label-free spectrum sensing workflow requiring no prior knowledge of active RATs or operating conditions during training. It supports post-hoc class assignment through cluster-to-class mapping based on domain-specific visualizations, enabling practical deployment in environments where labeled data is unavailable.

Lightweight deep clustering architecture: We designed a lightweight, CNN-based deep clustering architecture tailored for FFT amplitude processing. This architecture achieved a parameter count reduction by two orders of magnitude compared to the validated spectrogram-based approach, and 22% fewer parameters compared to the state of the art. The proposed architecture further reduces computational cost by 50% in terms of

FLOPs relative to existing alternatives, while improving feature separability and clustering quality by 35 percentage points.

The experimental investigations across Chapters 3 and 4 provide consistent and strong support for all three hypotheses. Deep clustering architectures reliably learn compact, discriminative representations from unlabeled RF data, reducing the annotation burden to coarse cluster-level labeling rather than individual sample annotation. Domain-specific architectural adaptation yields meaningful complexity reduction without sacrificing, and in several cases actively improving, clustering performance. Most importantly, the proposed models demonstrate consistent performance across diverse real-world radio environments spanning multiple frequency bands and varying RAT compositions, demonstrating the robust feature-learning capability of the proposed approach.

Taken together, these results establish that label-free operation, architectural efficiency, and consistent performance across diverse real-world datasets are mutually achievable within a single deep clustering framework, delivering the key contributions of this dissertation.

5.2 Future Work

Building on the demonstrated capabilities and validated contributions of this dissertation, four main directions are identified for future research aiming to further advance sensing- and learning-enabled wireless systems by unsupervised deep learning.

Further study of UDL models for radio signals

The survey presented in Chapter 2 reveals evident and growing efforts toward the design of dedicated models and data augmentations informed by the radio signal specifics, and this research direction should be strongly encouraged. The results presented in Chapters 3 and 4 further reinforce this finding, demonstrating that domain- and task-specific architectural design, rather than adoption of the generic existing models, yields substantial gains in both performance and computational efficiency, which increases the potential and flexibility for actual deployment. These findings collectively motivate continued research towards model design and processing pipelines tailored to the radio signals and task specifics, based on the UDL approaches.

Going beyond task-specific advances, several wireless signal recognition challenges, including specific emitter identification (SEI), automatic modulation classification (AMC), and anomaly detection, often rely on similarly structured I/Q (In-phase/Quadrature) signal representations, suggesting that UDL models trained on any of these tasks may learn transferable signal structure across domains. Systematic investigation of such cross-task shared representation learning, therefore, constitutes a well-motivated research direction, with the prospect of improving robustness and data efficiency across the broader wireless signal recognition landscape, ultimately supporting the development of wireless foundational models.

Robustness characterization and representation trade-offs

The evaluation carried out in Chapters 3 and 4 relied on real-world measurements that inherently contain mixed noise and interference, providing evidence of practical viability under realistic conditions. A natural extension is a more fine-grained characterization of model behavior under varying channel conditions, which should include a controlled sensitivity analysis across graded signal-to-noise ratio levels. Such an analysis would sys-

tematically map how recognition performance evolves as channel quality varies and identify the operating range over which reliable recognition is maintained.

A further direction concerns the input representation. The frequency-domain representations adopted in this work are well-suited to RAT recognition, where technologies are distinguishable by their spectral footprint, and they offer substantial data transfer and storage savings. Raw I/Q signals, in contrast, retain phase information that may benefit cases involving spectrally similar technologies, and a direct comparison of FFT amplitude and raw I/Q inputs within the proposed pipeline would quantify this trade-off.

Finally, while the proposed models perform consistently across several independent real-world datasets, the broader question of cross-dataset generalization, where a model trained in one environment is evaluated in a different, previously unseen one, remains open. A dedicated study of this setting would establish how well the learned representations transfer across deployments without retraining.

Applicative use-cases

The spectrum awareness capability enabled by deep clustering-based signal recognition supports two significant further developments. The first is the integration of monitoring feedback into closed-loop control systems for adaptive network parameter management, where RAT recognition outputs can directly inform decisions on channel allocation, transmission power control, beam management, interference mitigation, and handover triggering. Concrete future work in this direction includes the definition of appropriate monitoring granularity and update rates for different control loop timescales, the characterization of model inference time against practical real-time constraints such as the slot timings of 5G-NR, and the validation of end-to-end sensing-to-control pipelines in realistic network deployments.

The second is the development of advanced Radio Environment Maps (REMs) that go beyond conventional RSSI/pathloss estimation based on propagation modeling, environmental geometry, and antenna features. By integrating multi-layer sensing information, including RAT-specific activity patterns, uplink and downlink characteristics, and interference, it becomes possible to construct accurate and semantically rich Digital Twins of the wireless network and its environment, supporting intelligent, context-aware planning and dynamic management of both indoor and outdoor networks.

Open and standard benchmarking

Finally, the survey also highlights a notable lack of uniformity in the validation approaches adopted across the existing literature, including inconsistencies in dataset sizes, evaluation metrics, and few-shot evaluation protocols. This directly motivates future work toward standardization in the evaluation of UDL-based wireless signal recognition methods. This includes the creation of open, well-curated datasets spanning diverse radio environments and configurations, the definition of standard evaluation metrics suited to the label-free settings, and the establishment of reproducible processing pipelines and reference models. Standardized benchmarking is essential for fair and transparent comparison of methods and will play an important role in accelerating progress in this application-driven research field.

References

[1] Ericsson, "Mobility report," Ericsson, Tech. Rep., Nov. 2023.

[2] W. K. Alsaedi, H. Ahmadi, Z. Khan, and D. Grace, "Spectrum options and allocations for 6g: A regulatory and standardization review," *IEEE Open Journal of the Communications Society*, 2023.

[3] M. G. Cotton, "3.45–3.65 ghz spectrum occupancy from long-term measurements in 2018 and 2019 at four coastal sites," Institute for Telecommunication Sciences, Tech. Rep., 2020.

[4] Z. Ke and H. Vikalo, "Real-time radio technology and modulation classification via an lstm auto-encoder," *IEEE Transactions on Wireless Communications*, vol. 21, no. 1, pp. 370–382, 2021.

[5] S. Rajendran, W. Meert, D. Giustiniano, V. Lenders, and S. Pollin, "Deep learning models for wireless signal classification with distributed low-cost spectrum sensors," *IEEE Transactions on Cognitive Communications and Networking*, vol. 4, no. 3, pp. 433–445, 2018.

[6] I. G. A. Poornima and B. Paramasivan, "Anomaly detection in wireless sensor network using machine learning algorithm," *Computer Communications*, vol. 151, pp. 331–337, 2020.

[7] Y. Ghanney and W. Ajib, "Radio frequency interference detection using deep learning," in *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*, IEEE, 2020, pp. 1–5.

[8] J. Robinson, S. Kuzdeba, J. Stankowicz, and J. M. Carmack, "Dilated causal convolutional model for rf fingerprinting," in *2020 10th Annual Computing and Communication Workshop and Conference (CCWC)*, IEEE, 2020, pp. 157–162.

[9] S. Atapattu, C. Tellambura, and H. Jiang, *Energy Detection for Spectrum Sensing in Cognitive Radio*, 1st. New York, NY: Springer New York, NY, 2014. DOI: 10.1007/978-1-4939-0494-5.

[10] T. Gale, T. Šolc, R.-A. Moşoi, M. Mohorčič, and C. Fortuna, "Automatic detection of wireless transmissions," *IEEE Access*, vol. 8, pp. 24 370–24 384, 2020.

[11] K. Kim, I. A. Akbar, K. K. Bae, J.-S. Um, C. M. Spooner, and J. H. Reed, "Cyclostationary approaches to signal detection and classification in cognitive radio," in *2007 2nd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks*, IEEE, 2007, pp. 212–215. DOI: 10.1109/DYSPAN.2007.35.

[12] F. Paisana, N. Prasad, A. Rodrigues, and R. Prasad, "An alternative implementation of a cyclostationary detector," in *The 15th International Symposium on Wireless Personal Multimedia Communications*, 2012, pp. 45–49.

[13] J. Fontaine, A. Shahid, J. Vanhie-Van Gerwen, E. De Poorter, and I. Moerman, "Towards low-complexity wireless technology classification across multiple environments," *Ad Hoc Networks*, vol. 91, p. 101 881, 2019.

[14] T. J. O'Shea, T. Roy, and T. C. Clancy, "Over-the-air deep learning based radio signal classification," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 1, pp. 168–179, 2018.

[15] S. Riyaz, K. Sankhe, S. Ioannidis, and K. Chowdhury, "Deep learning convolutional neural networks for radio identification," *IEEE Communications Magazine*, vol. 56, no. 9, pp. 146–152, 2018.

[16] M. Caron, P. Bojanowski, A. Joulin, and M. Douze, "Deep clustering for unsupervised learning of visual features," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 132–149.

[17] L. Milosheski, G. Cerar, B. Bertalanič, C. Fortuna, and M. Mohorčič, "Self-supervised learning for clustering of wireless spectrum activity," *Computer Communications*, vol. 212, pp. 353–365, 2023.

[18] L. Milosheski, M. Mohorčič, and C. Fortuna, "Spectrum sensing with deep clustering: Label-free radio access technology recognition," *IEEE Open Journal of the Communications Society*, 2024.

[19] C. Liu *et al.*, "Overcoming data limitations: A few-shot specific emitter identification method using self-supervised learning and adversarial augmentation," *IEEE Transactions on Information Forensics and Security*, 2023.

[20] U. Demirhan and A. Alkhateeb, "Integrated sensing and communication for 6g: Ten key machine learning roles," *IEEE Communications Magazine*, 2023.

[21] C. Ouyang, Y. Liu, H. Yang, and N. Al-Dhahir, "Integrated sensing and communications: A mutual information-based framework," *IEEE Communications Magazine*, vol. 61, no. 5, pp. 26–32, 2023.

[22] Z. Wei *et al.*, "Integrated sensing and communication signals towards 5g-a and 6g: A survey," *IEEE Internet of Things Journal*, 2023.

[23] L. Bonati, M. Polese, S. D'Oro, S. Basagni, and T. Melodia, "Open, programmable, and virtualized 5g networks: State-of-the-art and the road ahead," *Computer Networks*, vol. 182, p. 107 516, 2020.

[24] S. Tadik *et al.*, "Digital spectrum twins for enhanced spectrum sharing and other radio applications," *IEEE Journal of Radio Frequency Identification*, 2023.

[25] A. S. Abdalla, P. S. Upadhyaya, V. K. Shah, and V. Marojevic, "Toward next generation open radio access networks: What o-ran can and cannot do!" *IEEE Network*, vol. 36, no. 6, pp. 206–213, 2022.

[26] C. Coletti *et al.*, *O-ran: Towards an open and smart ran*, O-RAN ALLIANCE, 2018.

[27] G. D. Durgin, M. A. Varner, N. Patwari, S. K. Kasera, and J. der Merwe, "Digital spectrum twinning for next-generation spectrum management and metering," in *2022 IEEE 2nd International Conference on Digital Twins and Parallel Intelligence (DTPI)*, IEEE, 2022, pp. 1–6.

[28] W. Zhang, M. Feng, M. Krunz, and A. H. Y. Abyaneh, "Signal detection and classification in shared spectrum: A deep learning approach," in *IEEE INFOCOM 2021-IEEE Conference on Computer Communications*, IEEE, 2021, pp. 1–10.

[29] A. Scalingi, D. Giustiniano, R. Calvo-Palomino, N. Apostolakis, and G. Bovet, "A framework for wireless technology classification using crowdsensing platforms," in *IEEE INFOCOM 2023-IEEE Conference on Computer Communications*, IEEE, 2023, pp. 1–10.

[30] M. Girmay, V. Maglogiannis, D. Naudts, M. Aslam, A. Shahid, and I. Moerman, "Technology recognition and traffic characterization for wireless technologies in its band," *Vehicular Communications*, vol. 39, p. 100 563, 2023.

[31] Y. Ren *et al.*, "Deep clustering: A comprehensive survey," *arXiv preprint arXiv:2210.04142*, 2022.

[32] H. Kim, A. Shahid, J. Fontaine, E. De Poorter, I. Moerman, and H. Nam, "Iot technology recognition using deep clustering," *IEEE Internet of Things Journal*, vol. 11, no. 8, pp. 13 372–13 382, 2024.

[33] M. Cheraghinia, E. De Poorter, J. Fontaine, M. Debbah, and A. Shahid, "A foundation model for wireless technology recognition and localization tasks," *IEEE Open Journal of the Communications Society*, vol. 6, pp. 9879–9896, 2025.

[34] H. Han, W. Li, Z. Feng, G. Fang, Y. Xu, and Y. Xu, "Proceed from known to unknown: Jamming pattern recognition under open-set setting," *IEEE Wireless Communications Letters*, vol. 11, no. 4, pp. 693–697, 2022.

[35] N. Zhang, J. Shen, Y. Shi, and Y. Li, "Cnn-zero: A zero-shot learning framework for jamming identification," in *2022 IEEE 22nd International Conference on Communication Technology (ICCT)*, IEEE, 2022, pp. 1126–1131.

Bibliography

Publications Related to the Thesis

Journal Articles

L. Milosheski, B. Bertalanič, C. Fortuna, and M. Mohorčič, "Radio signals recognition with unsupervised deep learning: A survey," *IEEE Access*, vol. 13, pp. 217 769–217 798, 2025.

L. Milosheski, G. Cerar, B. Bertalanič, C. Fortuna, and M. Mohorčič, "Self-supervised learning for clustering of wireless spectrum activity," *Computer Communications*, vol. 212, pp. 353–365, 2023.

L. Milosheski, M. Mohorčič, and C. Fortuna, "Spectrum sensing with deep clustering: Label-free radio access technology recognition," *IEEE Open Journal of the Communications Society*, vol. 5, pp. 4746–4763, 2024.

Conference Papers

L. Milosheski, G. Cerar, B. Bertalanić, C. Fortuna, and M. Mohorčič, "Xai for self-supervised clustering of wireless spectrum activity," in *2023 International Balkan Conference on Communications and Networking (BalkanCom)*, 2023, pp. 1–6.

L. Milosheski, G. Cerar, B. Bertalanič, C. Fortuna, and M. Mohorčič, "Deep feature learning for wireless spectrum data," in *2023 IEEE International Mediterranean Conference on Communications and Networking (MeditCom)*, 2023, pp. 121–126.

Other Publications

L. Milosheski, L. Mokrovič, B. Bertalanič, M. Mohorčič, C. Fortuna, and J. Hribar, "Self-supervised visual exploration of age of information process in iot," in *2024 IEEE 10th World Forum on Internet of Things (WF-IoT)*, 2024, pp. 753–758.

M. Krasnov, L. Milosheski, M. Mohorčič, and C. Fortuna, "Novel devices identification with deep clustering," in *2025 IEEE International Conference on Machine Learning for Communication and Networking (ICMLCN)*, 2025, pp. 1–6.

Biography

Ljupcho Milosheski received his Bachelor's and Master's degrees in Electrical Engineering and Information Technologies from the Ss. Cyril and Methodius University in Skopje, Macedonia, in 2017 and 2020, respectively. The final projects of the Bachelor's and the Master's degree were dedicated to a continuous research effort in system design and algorithms for surface anomaly inspection based on image processing for composite material production machines. This work was carried out in DIPTeam, and in collaboration with Mikrosam, a private enterprise, where the developed system was implemented and integrated in an industrial setting. Following the completion of his Master's studies, he was employed at Pathkeeper Surgical in Skopje from 2020 to 2021, where he worked on stereo 3D reconstruction for minimally invasive spine surgeries as part of a Horizon EU project.

In 2021, he joined the Jožef Stefan Institute in Ljubljana, Slovenia, first as a visiting researcher (April-July 2021), where he focused on algorithms for wireless transmission analysis using spectrograms and deep unsupervised machine learning. Since January 2022, he has been pursuing a PhD in Information and Communication Technologies at the Jožef Stefan International Postgraduate School, with his doctoral research centred on spectrum sensing using deep unsupervised machine learning approaches. His PhD work was carried out in SensorLab at the Department of Communication Systems and funded by the Slovenian Research Agency under the Young Researcher scheme as well as the HE NANCY and EnertTEF projects.

His current research interests include self-supervised deep learning for automatic feature learning and signal clustering in wireless communications, radio map estimation (REM) using generative models and 3D environmental data, lightweight deep learning models for spectrum sensing, explainable AI for self-supervised models, and Radio Access Technology classification based on spectrogram processing. He is actively involved in all phases of the research process, from conceptualisation and experimental design to critical analysis, scientific writing, and the supervision of junior researchers.

He serves as a peer reviewer for scientific journals and conferences in the field of wireless communications and machine learning, contributing to the quality and integrity of scientific discourse in his area of expertise. He is also an active participant in the broader research community, having attended and engaged in several summer schools, online events, and professional training programmes related to the wireless communications field.