Wavelet Classification for Non-Cooperative Non-Orthogonal Signal Communications

Tongyang Xu and Izzat Darwazeh

Department of Electronic and Electrical Engineering, University College London, London, UK

Email: tongyang.xu.11@ucl.ac.uk, i.darwazeh@ucl.ac.uk

Abstract—Non-cooperative communications using non-orthogonal multicarrier signals are challenging since self-created inter carrier interference (ICI) prevents successful signal classification. Deep learning (DL) can deal with the classification task without domain-knowledge at the cost of training complexity. Previous work showed that a tremendously trained convolutional neural network (CNN) classifier can efficiently identify feature-diversity dominant signals while it fails when feature-similarity dominates. Therefore, a pre-processing strategy, which can amplify signal feature diversity is of great importance. This work applies single-level wavelet transform to manually extract time-frequency features from non-orthogonal signals. Composite statistical features are investigated and the wavelet enabled two-dimensional time-frequency feature grid is further simplified into a one-dimensional feature vector via proper statistical transform. The dimensionality reduced features are fed to an error-correcting output codes (ECOC) model, consisting of multiple binary support vector machine (SVM) learners, for multiclass signal classification. Low-cost experiments reveal 100% classification accuracy for feature-diversity dominant signals and 90% for feature-similarity dominant signals, which is nearly 28% accuracy improvement when compared with the CNN classification results.

Index Terms—Signal classification, wavelet, machine learning, SVM, non-cooperative, non-orthogonal, SEFDM, waveform, experiment, software-defined radio.

I. INTRODUCTION

With the commercial deployment of 5G [1] and the recent advancement of 6G [2], [3], enhanced spectral efficiency with better intelligence will be expected. For a reliable signal recovery, traditional communication systems will occupy signalling overhead to send a mutually known side information from a transmitter to notify a receiver the signal format at the cost of wasting spectral efficiency. In addition, wireless channels are time-variant and received signals might be affected by timing/phase/frequency impairments. Compensation has to be operated before extracting accurate side information. Any imperfect operations would cause wrong decisions on signal format.

An intelligent deep learning receiver can automatically identify signal format in a non-cooperative way based on data training without any signal compensation pre-processing. A representative deep learning technique is convolutional neural network (CNN), which employs multiple convolutional layers for automatic feature extractions. CNN has been successfully applied in single carrier modulation classification [4] and multicarrier orthogonal frequency division multiplexing (OFDM) modulation classification [5].

Non-orthogonal signals might be considered for future communications due to their enhanced spectral efficiency. A potential waveform candidate is termed spectrally efficient frequency division multiplexing (SEFDM) [6], which

can achieve either higher data rate or compressed spectral bandwidth. One unique advantage of SEFDM is its similar signal generation method with OFDM. The classification for SEFDM signals has been theoretically and practically investigated in work [7], in which a trained CNN classifier can efficiently identify signals with distinguishable features (i.e. feature-diversity dominant) but it cannot accurately classify signals when features are similar (i.e. feature-similarity dominant). The tremendous fine-tuning for optimal CNN hyperparameters is time consuming. Therefore, manually extracting signal features, based on expert knowledge and traditional machine learning (ML), would be more efficient and convincing.

This work aims to accurately classify the feature-similarity dominant SEFDM signals, such that non-cooperative SEFDM communications will be achievable. Firstly, one-dimensional statistical features are evaluated using support vector machine (SVM). Then, a wavelet transform [8], [9] based time-frequency feature extraction approach is applied in this work. Unlike traditional multilevel structured wavelet decomposition [10], [11], this work focuses on a basic single-level wavelet filtering (WF) strategy. Results indicate that the two-dimensional time-frequency feature with statistical dimensionality reduction can assist SVM to identify feature-similarity dominant signals at high accuracy. Finally, a low-cost experiment is set up to verify the trained classifiers using over-the-air signals.

The main contributions of this work are as the following.

- Statistical features are investigated in SVM for non-orthogonal signal classification.
- Two-dimensional time-frequency features are evaluated via single-level wavelet transform. Various time-frequency feature dimensionality reduction methods are studied to simplify the features and further improve classification accuracy.
- Low-cost over-the-air experiment is designed to verify the robustness of the wavelet classification.

II. SEFDM WAVEFORM

The discrete format of a time-domain SEFDM signal is defined as

$$X_k = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} s_n \exp\left(\frac{j2\pi nk\alpha}{N}\right), \qquad (1)$$

where the expression is very similar to that of OFDM except the bandwidth compression factor $\alpha = \Delta f \cdot T$, in which Δf is the sub-carrier spacing and T is the time period of one SEFDM symbol. The number of sub-carriers
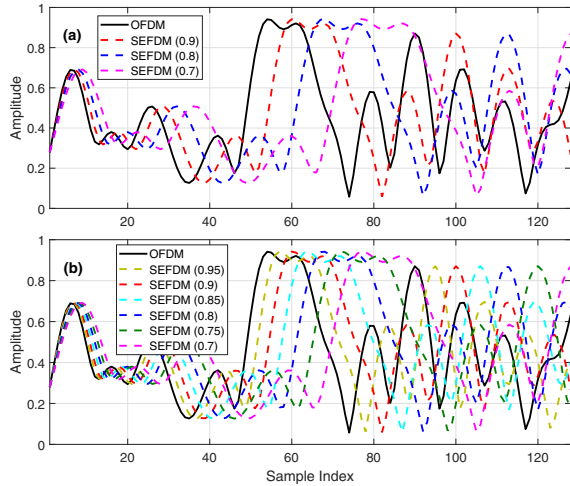
Fig. 1. Time-domain signal visualization by modulating the same QPSK data. Values in the bracket indicate the bandwidth compression factor α. (a) Type-I signal pattern with strong feature diversity. (b) Type-II signal pattern with strong feature similarity.
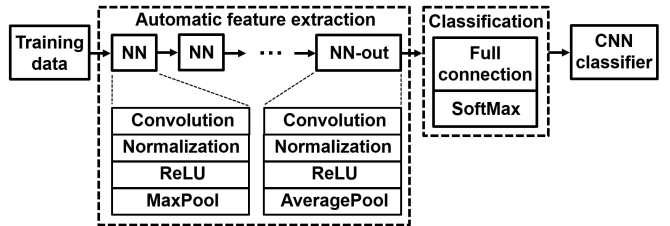

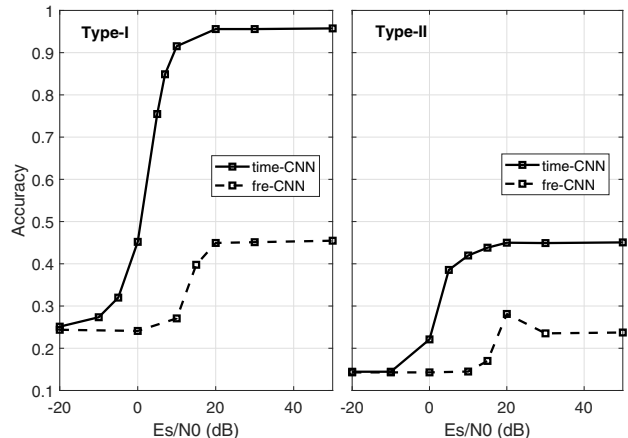Fig. 2. Neural network architecture of the CNN classifier.

Fig. 3. CNN classification accuracy for SEFDM signals considering either time-domain or frequency-domain features. Classifiers are trained by datasets after data augmentation at Es/N0=20 dB.

is determined by N. s_n is the nth single-carrier symbol within one SEFDM symbol and X_k is the kth time sample with $k = 0, 1, ..., N-1$. The signal spectral bandwidth in (1) is compressed when $\alpha < 1$ and is equivalent to that of OFDM when $\alpha = 1$.

Two signal patterns with different α are illustrated in Fig. 1. All of the signals occupy the same T. Therefore, the sub-carrier spacing Δf will be narrower with the reduction of α. Type-I signal pattern packs four signal classes while Type-II signal pattern has more signal options, which can support a wider range of services. However, Type-II will have challenging classification since signal features are similar to each other.

The instantaneous power for one SEFDM symbol is computed in the following

$$
\begin{aligned}
|X_k|^2 &= \frac{1}{N} \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} s_n s_m^* \exp\left(\frac{j2\pi(n-m)k\alpha}{N}\right) \\
&= \frac{1}{N} \sum_{n=0}^{N-1} |s_n|^2 + \\
&\quad \underbrace{\frac{1}{N} \sum_{n=0}^{N-1} \sum_{m \neq n, m=0}^{N-1} s_n s_m^* \exp\left(\frac{j2\pi(n-m)k\alpha}{N}\right)}_{ICI}.
\end{aligned}
$$

$$(2)$$

It is apparent that the inter carrier interference (ICI) term in (2), which is caused by α, determines the possibility of identifying different SEFDM signals. It is inferred that when SEFDM signals have similar values of α, the ICI term will become similar and would complicate signal classification.

III. CLASSIFICATION STRATEGIES

A. CNN Classification

A multi-layer CNN classifier, shown in Fig. 2, is trained in a recent work [7] to automatically extract signal features in either time-domain or frequency-domain. The automatic feature extractor has seven neural network (NN). The first

six NN modules has four sub-layers, namely convolutional layer, normalization layer, ReLU layer and MaxPool layer. For the last NN module, it specifically employs an Average-Pool layer instead of the MaxPool layer in order to obtain smooth features at the end. The classification part packs a full connection layer and a SoftMax layer.

Based on the extracted features, classification results are compared in Fig. 3, in which the time-domain classifier achieves higher accuracy than its frequency-domain counterpart. Classification accuracy can reach 95% when considering limited number of non-orthogonal signal classes in Type-I. However, the accuracy drops greatly when adding more similar signals in Type-II.

B. SVM Classification

The limitation of the previous work [7] is obviously shown in Fig. 3 and the motivation for this work is to accurately classify Type-II signals. The training of an optimal multi-layer CNN classifier is time-consuming since it requires extensive hyperparameter tuning and iterative back propagation optimization. Therefore, it would be more efficient to use traditional machine learning strategies with manual feature extractions. The SVM classifier, based on domain-knowledge dependent features, is applied in this work. Firstly, the training is fast since features are obtained in advance rather than time-consuming data training. Secondly, the methodology of machine learning is deterministic and its working principle can be well explained. Since there are multiple signal classes in Type-I and Type-II, therefore a multiclass error-correcting output codes (ECOC) model
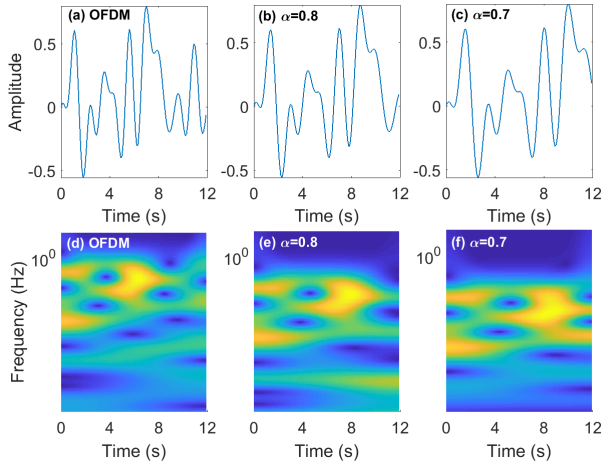
Fig. 4. Spectrogram of OFDM and SEFDM signals after wavelet transform. For the purpose of illustration, the signals are simply generated with N=12 sub-carriers at a data rate R_s=1 bit/s.

Fig. 5. One-dimensional wavelet feature generation based on wavelet filtering and statistical feature dimensionality reduction.

[12] is applied here. A one-versus-one [13] coding strategy is implemented to separate different signal classes, which simplifies the multiclass classification task into multiple binary class classification tasks. Thus, multiple binary SVM learners, with a polynomial kernel of order two, are used for the multiclass classification.

IV. FEATURE SELECTION

A. Statistical Features

The commonly used statistical feature is arithmetic mean, which computes the average value of a dataset. Variance is used to measure the variations of a dataset. Small variance indicates that the values of dataset elements are closer to the arithmetic mean while large variance indicates that the dataset elements are spread out away from the mean. Skewness [14] is a way to measure data distribution characteristics. Negative skewness indicates that a dataset distributes more data to the left side relative to its mean; positive skewness indicates that data is more distributed to the right side of the mean. The ratio between the maximum value and the minimum value is also studied here and the Max-Min ratio can tell the fluctuations of a dataset. Interquartile range (IQR) [15] is a way to measure data dispersion, which equals the difference between the 25th percentile and the 75th percentile.

B. Time-Frequency Features

The previous work [7] revealed that independent time-domain features or frequency-domain features cannot efficiently identify Type-II signals. Therefore, the joint analysis of time-frequency signal features is important since feature diversity would be enhanced by considering two domains. This section applies wavelet transform [8] to manually extract hidden signal features in time-frequency dimensions.

There are two types of wavelet transform for time-frequency analysis, namely continuous wavelet transform (CWT) and discrete wavelet transform (DWT). CWT provides a detailed representation for signals by using fine scale factors. It therefore leads to high-resolution signal analysis and can capture crucia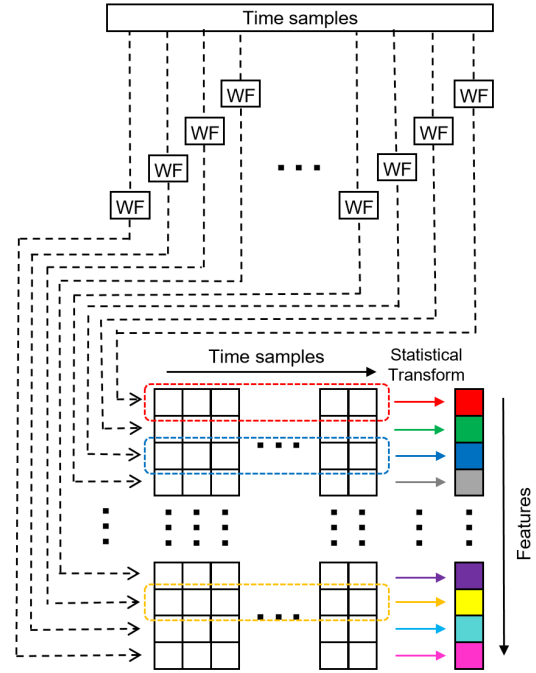l signal features. However, the obvious disadvantage of CWT is its higher computational complexity over DWT. A large time-frequency spectrogram grid would be obtained with the fine representation of scales. In this work, we would like to explore the accurate signal transient localization via detailed time-frequency analysis. Therefore, the high-resolution wavelet transform CWT is used rather than its coarse wavelet transform DWT.

There are several wavelet candidates for wavelet transform. This work employs the widely used Morse wavelet and the effects of different wavelets are not taken into account. The CWT time-frequency analysis for OFDM and SEFDM signals using Morse wavelet is illustrated in Fig. 4. It is clearly shown that with the reduction of α, the frequency scales for SEFDM shrink to show the effect of bandwidth compression while its time scales are stretched to show the time-domain sample characteristics. The typical artificial intelligent solution is to feed the time-frequency grid as an image to a deep learning neural network such as CNN. However, this would cause extra training complexity since the optimal neural network hyperparameters have to be tuned based on iterative attempts. Therefore, pre-processing is required to simplify the two-dimensional time-frequency feature representation into a one-dimensional feature vector as illustrated in Fig. 5. The strategy is to maintain the fine frequency scales of CWT while reducing time samples using the statistical knowledge explained in Section IV-A.

V. CLASSIFIER TRAINING AND TESTING

To have a realistic training scenario, channel/hardware impairments are considered in this work. A three-path wireless channel power delay profile (PDP) with path delay (s) [0 9e-6 1.7e-5] and path relative power (dB) [0 -2 -10] are reused from [4]. The maximum Doppler frequency is 4 Hz considering indoor people walking speed. The K-

Table I: Signal specifications

Parameter	Signal
Sampling frequency (kHz)	200
IFFT sample length	2048
Oversampling factor	8
No. of data sub-carriers	256
Bandwidth compression factor α	1,0.95,0.9,0.85,0.8,0.75,0.7
Modulation scheme	QPSK

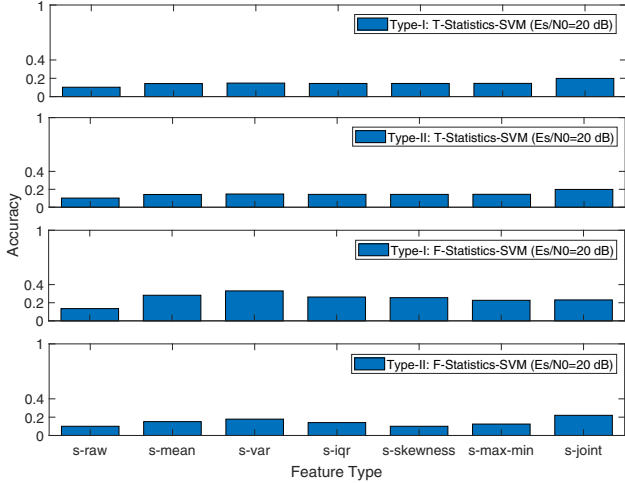Fig. 6. Statistical feature based SVM classification accuracy trained and tested at Es/N0=20 dB.

Fig. 7. Wavelet based SVM classification accuracy trained and tested at Es/N0=20 dB.

factor is 4 and the frequency offset is configured to be 2 parts per million (PPM). It should be noted that this work focuses on classifying OFDM/SEFDN signals and external signal interference and multi-antenna interference are not considered. Signals are generated according to Table I where 2048 time samples are produced at the transmitter for each OFDM/SEFDM symbol. There is no synchronization mechanism between the transmitter and the receiver. Therefore, the receiver would randomly truncate 1024 samples for training. At the training stage, 2,000 OFDM/SEFDM symbols are generated for each class (i.e. each α) following the data augmentation principle in [7]. In this case, there are overall 8,000 symbols for the Type-I signal pattern and 14,000 symbols for the Type-II signal pattern. For testing, there are overall 3,200 OFDM/SEFDM symbols for Type-I and 5,600 symbols for Type-II.

At first, statistical features in either time-domain (i.e. T-Statistics-SVM) or frequency-domain (i.e. F-Statistics-SVM) are compared in Fig. 6. Both the training data and testing data are contaminated by additive white Gaussian noise (AWGN) at a single Es/N0=20 dB. Results reveal that single domain features are not sufficient to classify either Type-I signal pattern or Type-II signal pattern.

The above results naturally lead to the joint time-frequency analysis, which would enhance the feature extraction efficiency. Wavelet transform will create a two-dimensional time-frequency feature grid. The scale range of the Morse wavelet is configured to have 7 octaves and 10 scales per octave. Therefore, considering both real and imaginary part of a signal, there are overall 140 frequency scales. In terms of time scale, following the signal specifi-
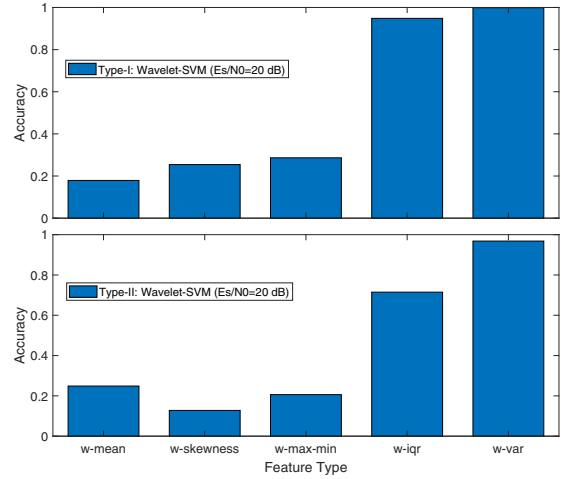
cations in Table I and the 50% random symbol truncation mechanism, 1024 time sample scales will be reserved. Therefore, CWT will generate a two-dimensional 140×1024 time-frequency analysis matrix.

There are many ways [16], [17] to reduce the time-frequency feature dimensionality. Considering computational complexity, this work applies simple statistical methods to reduce the amount of time samples. Thus, the two-dimensional 140×1024 time-frequency grid is simplified into a 140×1 frequency-scale vector following the dimensionality reduction method in Fig. 5. Different statistical transform methods are evaluated at each frequency scale and results are shown in Fig. 7. It is clearly seen that the IQR and variance features enable higher classification accuracy than other features, which can even classify the feature-similarity dominant Type-II signals. The following classifier training will be based on those two statistical features.

A wavelet classifier is firstly trained using data at a fixed Es/N0=20 dB and tested at various Es/N0 with accuracy results shown in Fig. 8(a). It clearly shows that all the curves reach the peak accuracy at 20 dB. However, for other Es/N0 values, accuracy drops significantly. It indicates that training data at a fixed Es/N0 is not robust to train a classifier that can classify signals at a wide range of Es/N0.

To train a robust classifier, a dataset covering different Es/N0 (20, 30, 40 dB) is generated. The classification results are shown in Fig. 8(b), in which better accuracy is reached at high Es/N0 for both Type-I and Type-II signals. However, the accuracy at low Es/N0 still needs improvement.

To enhance the classification sensitivity at low Es/N0, a dataset, covering low Es/N0 (0, 10, 20 dBs), is trained with results shown in Fig. 8(c). All the curves are raised to achieve higher accuracy at low Es/N0. It should be noted that the variance feature enabled wavelet classifier can identify signals even below noise power and it achieves 78% classification accuracy when Es/N0=0 dB. However, its performance drops obviously at high Es/N0, especially those beyond Es/N0=20 dB. For the IQR feature trained classifiers, both Type-I and Type-II curves are stable at
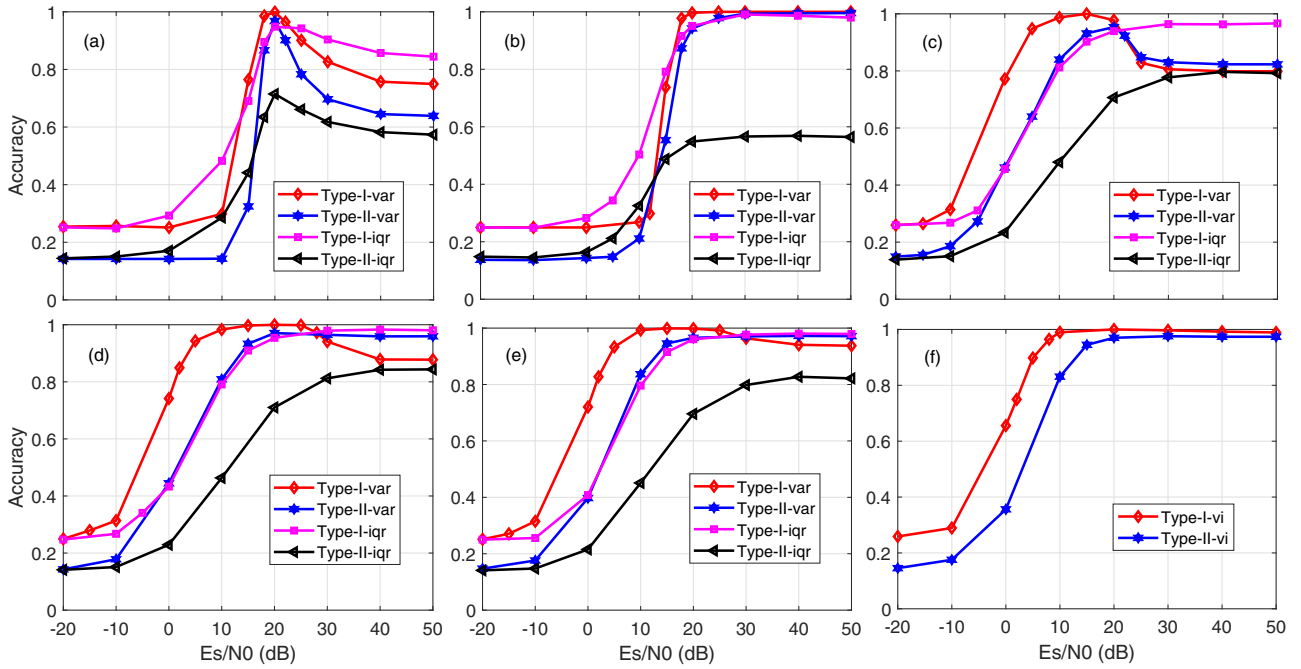
Fig. 8. Wavelet classifier accuracy tested at Es/N0 ranging from -20 dB to 50 dB. The classifier is trained at (a) Es/N0=20 dB. (b) Es/N0=20, 30, 40 dB. (c) Es/N0=0, 10, 20 dB. (d) Es/N0=0, 10, 20, 30, 40 dB. (e) Es/N0=-20, -10, 0, 10, 20, 30, 40, 50 dB. (f) Es/N0=-20, -10, 0, 10, 20, 30, 40, 50 dB. Note: var indicates variance, iqr indicates interquartile range and vi indicates variance-interquartile range.

high Es/N0 and the IQR feature trained Type-I classifier outperforms the variance feature trained model at high Es/N0. It is concluded from the figure that the variance trained model is robust at low Es/N0 while the IQR trained model is robust at high Es/N0.

Based on the above results, it is inferred that classifiers trained at high Es/N0 would enable high testing accuracy merely at high Es/N0 while classifiers trained at low Es/N0 would lead to high testing accuracy at low Es/N0. This indicates that a wider Es/N0 range has to be considered for the training data. In Fig. 8(d), classifiers are trained with data covering an Es/N0 range from 0 dB to 40 dB with an increment step of 10 dB, which basically combines the two Es/N0 ranges in Fig. 8(b) and Fig. 8(c). It clearly shows accuracy improvement for all the curves at both low and high Es/N0. In Fig. 8(e), a wider Es/N0 range between -20 dB and 50 dB is considered. The variance feature trained classifier improves Type-I signal classification accuracy at high Es/N0 while all other curves have no obvious improvement. However, there is still a minor performance gap between the variance trained classifier and the IQR trained classifier at high Es/N0. The robust feature performance of variance at low Es/N0 and IQR at high Es/N0 inspires to combine the two features for a more reliable classifier.

The composite classifiers, trained by joint variance and IQR features, can reach high classification accuracy for both Type-I and Type-II signals at both low and high Es/N0 ranges in Fig. 8(f). Therefore, the composite classifiers will be used in the following over-the-air experiments.

VI. LOW-COST EXPERIMENT AND RESULTS

The experiment is operated in an indoor open space, in which facilities would cause signal reflections and further
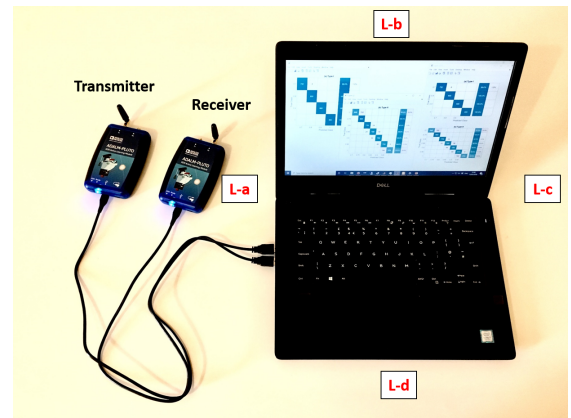

Fig. 9. Low-cost experiment setup for the wavelet classifier training and testing. Four data collection points are labelled as 'L-a', 'L-b', 'L-c' and 'L-d'.

result in frequency selective channel impairments. In addition, people movement in the space would cause Doppler spread and therefore dynamic spectral fluctuations. This work will use a pair of low-cost Analog Devices software-defined radio (SDR) PLUTO [18] to practically transmit and classify over-the-air signals. The signals are designed according to Table I and transmitted at a free-licensed 900 MHz (33-centimeter band) carrier frequency.

The experiment setup, shown in Fig. 9, is low cost since a laptop and two SDR devices are sufficient to realize signal generation, over-the-air transmission, signal reception and classifier training. In order to collect diversified data from an indoor environment, we fix the position of the transmitter side SDR device and place the receiver side
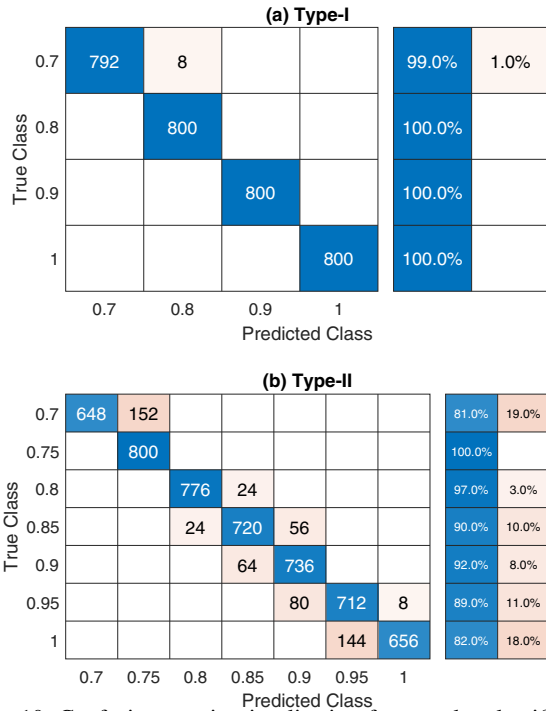
Fig. 10. Confusion matrix visualization for wavelet classification.

converted to a one-dimensional feature vector using statistical transform. Simulation results revealed that the most efficient features are variance and IQR. The combination of variance and IQR, associated with wavelet transform, enables classification accuracy up to 100%. Results also discovered that a wider range of training Es/N0 leading to better classification accuracy. Furthermore, the wavelet classifier can even identify signals when the signal power is below its noise power. Results show that the variance feature enabled wavelet classifier achieves 78% classification accuracy when Es/N0=0 dB. A low-cost experiment is set up using one laptop and two SDR devices. Practical results verified the efficacy of the wavelet enabled time-frequency features. Confusion matrices are obtained to show nearly 100% classification accuracy for the Type-I signal pattern and 90% accuracy for Type-II.

REFERENCES

[1] 3GPP TS 38.212 version 15.2.0, "5G NR; multiplexing and channel coding," Rel. 15, Jul. 2018.
[2] H. Tataria, M. Shafi, A. F. Molisch, M. Dohler, H. Sjland, and F. Tufvesson, "6G wireless systems: Vision, requirements, challenges, insights, and opportunities," 2020.
[3] K. B. Letaief, W. Chen, Y. Shi, J. Zhang, and Y. A. Zhang, "The roadmap to 6G: AI empowered wireless networks," *IEEE Communications Magazine*, vol. 57, no. 8, pp. 84–90, 2019.
[4] 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, Feb. 2018.
[5] S. Hong, Y. Zhang, Y. Wang, H. Gu, G. Gui, and H. Sari, "Deep learning based signal modulation identification in OFDM systems," *IEEE Access*, 2019.
[6] T. Xu and I. Darwazeh, "Transmission experiment of bandwidth compressed carrier aggregation in a realistic fading channel," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 5, pp. 4087–4097, May 2017.
[7] T. Xu and I. Darwazeh, "Deep learning for over-the-air non-orthogonal signal classification," in *2020 IEEE 91st Vehicular Technology Conference (VTC Spring)*, May 2020, pp. 1–5.
[8] S. Mallat, *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way*, 3rd ed. Orlando, FL, USA: Academic Press, Inc., 2008.
[9] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Transactions on Information Theory*, vol. 36, no. 5, pp. 961–1005, Sep. 1990.
[10] Y. Mallet, D. Coomans, J. Kautsky, and O. De Vel, "Classification using adaptive wavelets for feature extraction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 10, pp. 1058–1066, Oct. 1997.
[11] I. Guler and E. D. Ubeyli, "Multiclass support vector machines for EEG-signals classification," *IEEE Transactions on Information Technology in Biomedicine*, vol. 11, no. 2, pp. 117–126, Mar. 2007.
[12] T. G. Dietterich and G. Bakiri, "Solving multiclass learning problems via error-correcting output codes," *Journal of Artificial Intelligence Research*, vol. 2, p. 263286, Jan. 1995. [Online]. Available: http://dx.doi.org/10.1613/jair.105
[13] A. Rocha and S. K. Goldenstein, "Multiclass from binary: Expanding one-versus-all, one-versus-one and ECOC-based approaches," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 2, pp. 289–302, Feb. 2014.
[14] D. P. Doane and L. E. Seward, "Measuring skewness: A forgotten statistic?" *Journal of Statistics Education*, vol. 19, no. 2, pp. 1–18, 2011. [Online]. Available: https://doi.org/10.1080/10691898.2011.11889611
[15] G. Upton and I. Cook, *Understanding Statistics*. Oxford University Press, 1996.
[16] I. Fodor, "A survey of dimension reduction techniques," Lawrence Livermore National Laboratory, CA (US), Tech. Rep., 2002.
[17] L. V. D. Maaten, E. Postma, and J. V. D. Herik, "Dimensionality reduction: A comparative review," Tilburg University, Tech. Rep., 2009.
[18] Analog-Devices, "ADALM-PLUTO SDR active learning module," https://www.analog.com/media/en/news-marketing-collateral/product-highlight/ADALM-PLUTO-Product-Highlight.pdf, technical document, Jul. 2019.

SDR device at different locations. In this case, a number of training datasets, impaired by channel multipath fading, power degradation and Doppler effect, are collected. Unlike the CNN classifier where a large number of training symbols are required for feature extractions, the wavelet classifier can manually extract features based on a limited dataset. Therefore, in this experiment, at each location, 400 symbols are collected for the Type-I signal pattern and 700 symbols for the Type-II signal pattern. There are four data collections considering four different locations of the receiver. Therefore, the overall collected training symbols for Type-I and Type-II are 1,600 and 2,800, respectively. For testing, the same process is repeated with four data collections. To have a fair comparison with the previous work [7], the number of testing symbols per class is fixed at 800.

The collected data will be used to train wavelet classifiers off-line using Matlab. Once a wavelet classifier is trained, the model will be saved. Therefore, SDR devices will reuse the saved model for online signal classification and there is no need to re-train classifiers. Thus, the off-line training is a one-time operation. The confusion matrices are presented in Fig. 10. The classification accuracy for the Type-I signal pattern is nearly 100%. For Type-II signals, the accuracy is 90%, which is much higher than the 70.75% in [7] where a transfer learning enabled CNN classifier is applied.

VII. CONCLUSION

This work aims to explore typical machine learning (ML) algorithms for non-orthogonal signal classification in non-cooperative communications. Multiple statistical approaches are tested for feature extractions in either time-domain or frequency-domain but showing unreliable classification accuracy. Wavelet transform is therefore applied to extract two-dimensional time-frequency features, which are further