Cognitive radar antenna selection via deep learning

Ahmet M. Elbir[1], Kumar Vijay Mishra[2] ✉, Yonina C. Eldar[2]

[1]Department of Electrical and Electronics Engineering, Duzce University, Duzce, Turkey
[2]Andrew and Erna Viterbi Faculty of Electrical Engineering, Technion-Israel Institute of Technology, Haifa, Israel
✉ E-mail: mishra@ee.technion.ac.il

Abstract: Direction-of-arrival (DoA) estimation of targets improves with the number of elements employed by a phased array radar antenna. Since larger arrays have high associated cost, area and computational load, there is a recent interest in thinning the antenna arrays without loss of far-field DoA accuracy. In this context, a cognitive radar may deploy a full array and then select an optimal subarray to transmit and receive the signals in response to changes in the target environment. Prior works have used optimisation and greedy search methods to pick the best subarrays cognitively. In this study, deep learning is leveraged to address the antenna selection problem. Specifically, they construct a convolutional neural network (CNN) as a multi-class classification framework, where each class designates a different subarray. The proposed network determines a new array every time data is received by the radar, thereby making antenna selection a cognitive operation. Their numerical experiments show that the proposed CNN structure provides 22% better classification performance than a support vector machine and the resulting subarrays yield 72% more accurate DoA estimates than random array selections.

1 Introduction

Cognitive radar has gained much attention in the last decade due to its ability to adapt both the transmitter and receiver to changes in the environment and provide flexibility for different scenarios as compared with conventional radar systems [1–3]. Several applications have been considered for radar cognition such as waveform design [4–7], target detection and tracking [8, 9] and spectrum sensing and sharing [10–14]. Cognitive radar design requires reconfigurable circuitry for many subsystems such as power amplifiers, waveform generator and antenna arrays [15]. In this paper, we focus on this latter aspect of antenna array design in cognitive radar.

For a given wavelength, good angular resolution is achieved by a wide array aperture resulting in a large number of array elements, physical area and the cost associated with the array circuitry [16–18]. Hence, general approaches have been proposed to effectively use the array output with a minimal number of antenna elements. For example, non-uniform array structures [19, 20] are used to virtually increase the array aperture for direction-of-arrival (DoA) estimation. Given a full Nyquist antenna array, one could also randomly choose a few antenna elements to transmit/receive (Tx–Rx) and then employ efficient recovery algorithms, so that the spatial resolution does not degrade [7, 16, 21–23]. However, such approaches are agnostic to information about the received signal. A cognitive approach may be to select these Tx–Rx antennas based on the current target scenario encoded in the received signal [15] connecting antenna selection with contemporary interest in cognitive radar. The key idea is to exploit available data from the current radar scan to choose an optimal subarray for the next scan since target locations change little during consecutive scans.

Recent research [24–26] has proposed a reconfigurable array structure for a cognitive radar, which obtains an adaptive switching matrix after a combinatorial search for an optimal subarray that minimises a lower bound on the DoA estimation error. Related work in [27] proposed a greedy search algorithm to find a subarray that maximises the mutual information between the collected measurements and the far-field array pattern. Very recently, a semi-definite programme proposed in [28] selects a Tx–Rx antenna pair for a multiple-input–multiple-output (MIMO) radar that maximises the separation between desired and parasitic DoAs. Similar

problems have also been investigated in communications especially in the context of massive MIMO [29] to achieve energy and cost-efficient antenna designs and beamforming [30–32]. More generally, in the context of sensor selection, Joshi and Boyd [33] and Roy *et al.* [34] solve convex optimisation problems to obtain optimal antenna subarrays for DoA estimation. Similarly, Godrich *et al.* [35] selects the sensors for a distributed multiple-radar scenario through a greedy search with the Cramér–Rao lower bound (CRB) as a performance metric.

Nearly, all of these formulations solve a mathematical optimisation problem or use a greedy search algorithm. A few other works explore supervised machine learning (ML) to estimate DoA in the context of radar [36] and communications [37, 38]. Specifically, Joung [39] employs support vector machines (SVMs) for antenna selection in wireless communications. As a class of ML methods, deep learning (DL) has gained much interest recently for the solution of many challenging problems such as speech recognition, visual object recognition and language processing [40, 41]. DL has several advantages such as low computational complexity when solving optimisation-based or combinatorial search problems and the ability to extrapolate new features from a limited set of features contained in a training set [37, 40]. In the context of radar, DL has found applications in waveform recognition [42], image classification [43, 44], range–Doppler signature detection [45] and rainfall estimation [46].

In this paper, we introduce a DL-based approach for antenna selection in a cognitive radar. DL techniques directly fit our setting because the antenna selection problem can be considered as a classification problem, where each subarray designates a class. Among prior studies, the closest to our work is [39] where SVM is fed with the channel state information to select subarrays for the best MIMO communication performance. However, SVM is not as powerful as DL for extracting feature information inherits in the input data [40]. Furthermore, Joung [39] considers only small array sizes. On the other hand, the optimisation methods suggested in [34, 35] assume a priori knowledge of the target location/DoA angle to compute the CRB. Compared to these studies, we leverage DL to consider a relatively large scale of the selection problem, wherein the feature maps can be extracted to train the network for different array geometries. The proposed approach avoids solving a difficult optimisation problem [33]. Unlike random array thinning

where a fixed subarray is used for all scans, we select a new subarray based on the received data. In contrast to Roy *et al.* [34] and Godrich *et al.* [35], we also assume that the target DoA angle is unknown while choosing the array elements. To the best of our knowledge, this is the first work that addresses the radar antenna selection problem using DL.

In particular, we construct a convolutional neural network (CNN) for our problem. The input data to our CNN are the covariance samples of the received array signal. Previous radar DL applications [42–45] have used image-like inputs such as synthetic aperture radar signatures and time–frequency spectrograms. Our proposed CNN models the selection of K best antennas out of M as a classification problem, wherein each class denotes an antenna subarray. To create the training data, we choose those subarrays which estimate DoA with the lowest minimal bound on the mean-squared error (MSE). We consider minimisation of CRB as the performance benchmark in generating training sets for one-dimensional (1D) uniform linear arrays (ULAs) and 2D geometries such as uniform circular arrays (UCAs) and randomly deployed arrays (RDAs). For ULAs, we also train the network with data obtained by minimising Bayesian bounds such as the Bobrovsky–Zakai bound (BZB) and Weiss–Weinstein bound (WWB) on DoA [47] because these bounds provide better estimates of MSE at low signal-to-noise ratios (SNRs). In particular BZB-based selection has been shown to have the ability to control the sidelobes and avoid ambiguity in DoA estimation [24, 25]. Our results show that the proposed CNN classification performance is 22% better than the SVM. The DoA estimation accuracy using the subarrays obtained from our CNN network is 32 and 72% higher than the SVM and random array selections, respectively.

The rest of this paper is organised as follows. In the next section, we describe the system model and formulate the antenna selection problem. In Section 3, we introduce the proposed CNN and provide details on the training data. We evaluate the performance of our DL method in Section 4 through several numerical experiments.

Throughout this paper, we reserve boldface lowercase and uppercase letters for vectors and matrices, respectively. The ith element of a vector $\boldsymbol{y}$ is $y(i)$, whereas the (i, j)th entry of the matrix $\boldsymbol{Y}$ is $[\boldsymbol{Y}]_{i, j}$. We denote the transpose and Hermitian by $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\mathrm{H}}$, respectively. The functions $\angle\{\cdot\}$, $\mathbb{Re}\{\cdot\}$ and $\mathbb{Im}\{\cdot\}$ designate the phase, real and imaginary parts of a complex argument, respectively. The combination of selecting K terms out of M is denoted by $(M/K) = (M!/K!(M-K)!)$. The calligraphic letters $\mathscr{S}$ and $\mathscr{L}$ denote the position sets of all and selected subarrays, respectively. The Hadamard (point wise) product is written as $\odot$. The functions $E\{\cdot\}$ and max give the statistical expectation and maximum value of the argument, respectively. The notation $x \sim u\{[u_l, u_u]\}$ means a random variable drawn from the uniform distribution over $[u_l, u_u]$ and $x \sim \mathcal{CN}(\mu_x, \sigma_x^2)$ represents the complex normal distribution with mean μ_x and variance σ_x^2.

2 System model and problem formulation

Consider a phased array antenna with M elements, where each element transmits a pulsed waveform $s(t)$ toward a Swerling Case 1 point target. Since we are interested only in DoA recovery, the range and Doppler measurements are not considered and the target's complex reflectivity is set to unity. The assumption of a Swerling I model implies that the target parameters remain constant for the duration of the scan. We characterise the target through its DoA $\Theta = (\theta, \phi)$, where θ and ϕ denote, respectively, the elevation and the azimuth angles with respect to the radar. The radar's pulse repetition interval (PRI) and operating wavelength are, respectively, T_s and $\lambda = c_0/f_0$, where $c_0 = 3 \times 10^8 \, \mathrm{ms}^{-1}$ is the speed of light and $f_0 = \omega_0/2\pi$ is the carrier frequency.

To further simplify the geometries, we suppose that the targets are far enough from the radar so that the received signal wavefronts are effectively planar over the array. The array receives a narrowband signal reflected from a target located in the far field of the array at Θ. We denote the position vector of the mth receive

antenna by $\boldsymbol{p}_m = [p_{x_m}, p_{y_m}, p_{z_m}]^{\mathrm{T}}$ in a Cartesian coordinate system and assume that the antennas are identical and well-calibrated. Let $s(t)$ and $y_m(t)$ denote the source signal and the output signal at the mth sensor of the array, respectively. The baseband continuous-time received signal at the mth antenna is then

$$y_m(t) = a_m(\Theta)s(t) + n_m(t), \quad 0 \le t \le T_s, \tag{1}$$

where $n_m(t)$ is temporally and spatially white zero-mean Gaussian noise with variance σ_n^2 and

$$a_m(\Theta) = \exp\left\{\mathrm{j}\frac{2\pi}{\lambda}c_0\tau_m\right\}, \tag{2}$$

is the mth element of the steering vector

$$\boldsymbol{a}(\Theta) = [a_1(\Theta), a_2(\Theta), \ldots, a_M(\Theta)]^{\mathrm{T}}. \tag{3}$$

Here, τ_m is the time delay from the target to the mth antenna with respect to the reference antenna in the array, and is given by $\tau_m = -(1/c_0)\boldsymbol{p}_m^{\mathrm{T}}\boldsymbol{r}(\Theta)$ where $\boldsymbol{r}(\Theta)$ is the 2D *DoA parameter*

$$\boldsymbol{r}(\Theta) = [\cos(\phi)\sin(\theta), \sin(\phi)\sin(\theta), \cos(\theta)]^{\mathrm{T}}. \tag{4}$$

The radar acquires the signal for the lth snapshot over a PRI T_a. For a given snapshot l, we define an $M \times 1$ received signal vector $\boldsymbol{y}(lT_a) = [y_1(lT_a), \ldots, y_M(lT_a)]^{\mathrm{T}}$. For all L snapshots, omitting T_a from the indices for notational simplicity, we can express the received signal in matrix form as

$$\boldsymbol{Y} = \boldsymbol{a}(\Theta)\boldsymbol{s}^{\mathrm{T}} + \boldsymbol{N}, \tag{5}$$

where $\boldsymbol{Y}$ is the $M \times L$ matrix given as $\boldsymbol{Y} = [\boldsymbol{y}(1), \ldots, \boldsymbol{y}(L)]$, $\boldsymbol{s} = [s(1), \ldots, s(L)]^{\mathrm{T}}$ and $\boldsymbol{N} = [\boldsymbol{n}(1), \ldots, \boldsymbol{n}(L)]$ with $\boldsymbol{n}(l) = [n_1(l), \ldots, n_M(l)]^{\mathrm{T}}$ denoting the noise term.

Expanding the inner product $\boldsymbol{p}_m^{\mathrm{T}}\boldsymbol{r}(\Theta)$ in the array steering vector gives $a_m(\Theta) = \exp\left\{-\mathrm{j}\frac{2\pi}{\lambda}(p_{x_m}\mu + p_{y_m}\nu + p_{z_m}\xi)\right\}$, where $\mu = \cos(\phi)\sin(\theta)$, $\nu = \sin(\phi)\sin(\theta)$ and $\xi = \cos(\theta)$. Evidently, $a_m(\Theta)$ is a multi-dimensional harmonic. Once the frequencies μ, ν and ξ in different directions are estimated, the DoA angles are obtained using the relations

$$\theta = \tan^{-1}\left(\frac{\nu}{\mu}\right), \quad \phi = \cos^{-1}\left(\frac{\xi}{\sqrt{\mu^2 + \nu^2 + \xi^2}}\right), \tag{6}$$

with the usual ambiguity in $[0, 2\pi]$. In the case of a linear array, there is only one parameter in the steering vector whereas two parameters are involved in planar and 3D arrays.

For instance, consider a planar array so that $p_{z_m} = 0$ and there is only one incoming wave. Then, a minimal configuration to find the two frequencies consists of at least three elements in an L-shape configuration to estimate frequencies in the $x-$ and $y-$ directions. More sensors are needed if the incoming signal is a superposition of P wavefronts. Many theoretical works have investigated the uniqueness of 2D harmonic retrieval (see, e.g. [48–50]). For example, in the case of a uniform rectangular array of size $M_1 \times M_2$, classically $P \le M_1 M_2 - \min(M_1, M_2)$ specifies the minimum required number of sensors in the absence of noise. This can be relaxed by obtaining several snapshots or using co-prime sampling if the sources are uncorrelated [19, 20] or spatially compressed sensing [21, 22].

In the antenna selection scenario, the radar has M antennas but desires to use only $K < M$ elements to save computational cost, energy and sharing aperture momentarily to look at other directions. In practise, the signal is corrupted by noise and the antenna elements are not ideally isotropic. Therefore, K should be larger than the minimum number of elements predicted by classical

results. Removal of elements from an array rises the sidelobe levels and introduces ambiguity in resolving DoAs. The exact choice of K depends on the estimation algorithm employed by the receiver processor. For example, Mishra *et al.* [21], Rossi *et al.* [22] and Nion and Sidiropoulos [50] provide different guarantees for the minimum K depending on the array configuration and algorithm used for extracting the DoA.

In general, the target's position changes little during consecutive scans while a phased array can switch very fast from one antenna configuration to the other. Here, we consider the following scan strategy for the radar: at the beginning (the very first scan), all M antennas are active and the received signal from this scan is fed to the network. Our goal is to find an optimal antenna array for the next scan, in which only K antennas will be used. The radar continues to use this subarray for a few subsequent scans. After surveying the target scene with this optimal subarray for a predetermined number of scans, the radar switches back to the full array for a single scan. The received signal from this full array scan is then used to find a new, optimal subarray for the subsequent few scans. The frequency of choosing a new subarray can be decided off-line based on the nature of the target and analysis of previous observations. This switching of elements between different scans is a cognitive operation because a new array is determined in every few scans based on received echoes from the target scene. In the next section, we describe our DL approach to cognitive antenna selection.

3 Deep NN for antenna selection

Assume that an antenna subarray composed of K antennas is to be selected from an M-element antenna array. There are $Q = (M/K)$ possible choices. This can be viewed as a classification problem with Q classes each of which represents a different subarray. Let $\mathscr{P}_k^q = \{p_{x_k}^q, p_{y_k}^q, p_{z_k}^q\}$, $k = 1, \ldots, K$, be the set of the kth antenna coordinates in the qth subarray. Then, the qth class consisting of the positions of all elements in the qth subarray is

$$\mathscr{S}_q = \{\mathscr{P}_1^q, \ldots, \mathscr{P}_K^q\}, \tag{7}$$

and all classes are given by the set

$$\mathscr{S} = \{\mathscr{S}_1, \mathscr{S}_2, \ldots, \mathscr{S}_Q\}. \tag{8}$$

In Section 3.2, we propose a CNN to solve this classification problem by selecting the positions of the antenna subarray that provides the best DoA estimation performance. We discuss in Section 3.1 our procedure to create training data relying on the CRB. Note that, for an operational radar, generation of an artificial or simulated dataset to train the DL network is not necessary. Instead, the network can train itself with the data acquired by the radar during previous scans. During the test phase, for which we present results in Section 4, the DoA angles are unknown to the network. The CNN accepts the features from the estimated covariance matrix and outputs a new array. This stage is, therefore, cognitive because the radar is adapting the antenna array in response to the received signal. In our simulations, we observe that the proposed network can provide robust antenna selection performance with up to two degrees DoA mismatch between the data of consecutive scans, where an UCA is considered for $M = 20$, $K = 5$ and SNR = 15 dB.

3.1 Training data design

The training samples are the input for the CNN. To generate the training data, we select a set of target DoA locations and analyse all possible array configurations. We then generate class labels for those arrays that minimise a lower bound on the DoA estimation error. We choose CRB to label the training samples because this bound leads to simplified expressions. While other bounds such as BZB and WWB can also be considered [24, 25, 39, 51]; existing literature provides their expressions only for ULAs. Hence, we use CRB for UCA and RDA geometries in this work.

Consider L statistically independent observations of the qth subarray with K elements

$$\boldsymbol{y}_q(l) = \boldsymbol{a}_q(\Theta)s(l) + \boldsymbol{n}_q(l), \tag{9}$$

where $\boldsymbol{a}_q(\Theta)$ and $\boldsymbol{n}_q(l)$ denote the $K \times 1$ elements of $\boldsymbol{a}(\Theta)$ and $\boldsymbol{n}(l)$ corresponding to the qth subarray position set $\mathscr{S}_q$. The signal and noise are assumed to be stationary and ergodic over the observation period. The covariance matrix of the observations for the qth subarray is

$$\boldsymbol{R}_q = \mathrm{E}\{\boldsymbol{y}_q \boldsymbol{y}_q^{\mathrm{H}}\} = \boldsymbol{a}_q(\Theta)E\{s[l]s^{\mathrm{H}}[l]\}\boldsymbol{a}_q^{\mathrm{H}}(\Theta) + \sigma_n^2 \boldsymbol{I}_K, \tag{10}$$

where $\boldsymbol{I}_K$ is the identity matrix of dimension K. To simplify the CRB expressions, we represent the $K \times 1$ steering vector $\boldsymbol{a}_q(\Theta)$ as $\boldsymbol{a}_q$, and assume that $E\{s[l]s^{\mathrm{H}}[l]\} = \sigma_s^2$, where σ_s^2 and σ_n^2 are known. Let $\sigma_s^2 = 1$ for simplicity and define SNR as $10\log_{10}(\sigma_s^2/\sigma_n^2)$ dB. For this model, the CRBs for jointly estimating the target DoA coordinates θ and ϕ are, respectively, [52–54]

$$\mathrm{CRB}_\theta = \frac{\sigma_n^2}{2L\mathbb{R}\mathrm{e}\{(\dot{\boldsymbol{a}}_{q\theta}^{\mathrm{H}}[\boldsymbol{I}_K - \boldsymbol{a}_q \boldsymbol{a}_q^{\mathrm{H}}/K]\dot{\boldsymbol{a}}_{q\phi}) \odot (\sigma_s^4 \boldsymbol{a}_q^{\mathrm{H}} \boldsymbol{R}_q^{-1} \boldsymbol{a}_q)\}}, \tag{11}$$

and

$$\mathrm{CRB}_\phi = \frac{\sigma_n^2}{2L\mathbb{R}\mathrm{e}\{(\dot{\boldsymbol{a}}_{q\theta}^{\mathrm{H}}[\boldsymbol{I}_K - \boldsymbol{a}_q \boldsymbol{a}_q^{\mathrm{H}}/K]\dot{\boldsymbol{a}}_{q\theta}) \odot (\sigma_s^4 \boldsymbol{a}_q^{\mathrm{H}} \boldsymbol{R}_q^{-1} \boldsymbol{a}_q)\}}, \tag{12}$$

The partial derivatives $\dot{\boldsymbol{a}}_{q\theta} = (\partial \boldsymbol{a}_q/\partial\theta)$ and $\dot{\boldsymbol{a}}_{q\phi} = (\partial \boldsymbol{a}_q/\partial\phi)$ are computed using the expressions in (2) and (3). The absolute CRB [54] for the 2D DoA $\Theta = \{\theta, \phi\}$ using subarray $\mathscr{S}_q$ is

$$\eta(\Theta, \mathscr{S}_q) = \frac{1}{\sqrt{2}}(\mathrm{CRB}_\theta^2 + \mathrm{CRB}_\phi^2)^{1/2}. \tag{13}$$

The classification problem for antenna selection poses a large number of classes especially for large arrays since Q increases significantly with $O(M!)$. To alleviate this issue, one can collect the classes randomly to reduce the complexity according to a computation/performance trade-off [55]. Owing to the direction of the target and the array geometry, $\bar{Q}$, the number of classes that provide best subarrays is much smaller than Q which allows us to reduce the number of classes.

To label the training samples, we first compute the sample covariance matrix from L snapshots of noisy observations. We then obtain the CRB $\eta(\Theta, \mathscr{S}_q)$ for each target direction Θ in the training set with all subarrays $q = 1, \ldots, Q$. The class labels for the input data indicate the *best array*, i.e. the array which minimises the CRB in a given scenario. Let us define $\bar{Q}$ as the number of subarrays that provide the best DoA estimation performance for different directions. Then, $\bar{Q}$ is generally much smaller than Q because of the direction of the target and the aperture of the subarrays. For an illustrative comparison of Q and $\bar{Q}$, we refer the reader to Table 1 which lists the number of these classes for an UCA. Hence, we construct a new set $\mathscr{L}$ which includes only those classes that represent the selected subarrays for different directions

$$\mathscr{L} = \{l_1, l_2, \ldots, l_{\bar{Q}}\}, \tag{14}$$

where $\bar{Q}$ is the reduced number of classes: $l_{\bar{q}}$ is the subarray class that provides the lowest CRB for direction Θ, namely

$$l_{\bar{q}} = \underset{q = 1, \ldots, Q}{\arg\min} \, \eta(\Theta, \mathscr{S}_q), \tag{15}$$

for $\bar{q} = 1, \ldots, \bar{Q}$. Once the label set $\mathscr{L}$ is obtained, the input–output data pairs are constructed as $(\boldsymbol{X}, z)$, where $\boldsymbol{X}$ is an $M \times M \times 3$ real-valued input data obtained from the covariance matrix as defined in

Table 1 Number of classes Q and the reduced number of classes $\bar{Q}$ for the UCA geometry with $M = 10$ and 16 antennas. Here, elevation angle is fixed at $\theta = 90°$ and the number of azimuthal grid points $P_\phi = 100$ for uniformly gridded azimuth plane in $[0°, 360°)$

	$K = 3$	$K = 4$	$K = 5$	$K = 6$	$K = 7$	$K = 8$
UCA with $M = 10$						
Q	120	210	252	210	120	45
$\bar{Q}$	10	10	10	10	10	9
UCA with $M = 16$						
Q	560	1820	4368	8008	11,440	12,870
$\bar{Q}$	16	10	16	11	16	16

Input: Number of given antenna elements M, number of desired elements K, number of snapshots L, number of different DoA angles considered P, number of signal and noise realisations T, $\text{SNR}_{\text{TRAIN}}$ and $\sigma_s^2 = 1$.

Output: *Training data*: Input–output pairs consisting of sample covariances $\hat{\boldsymbol{R}}^{(i, p)}$ and output labels $z_p^{(i)}$ for $p = 1, \ldots, P$ and $i = 1, \ldots, T$.

1. Select P DoA angles $\Theta_p = (\theta_p, \phi_p)$ for $p = 1, \ldots, P$.

2. Generate T different realisations of the array output $\left\{\boldsymbol{Y}_p^{(i)}\right\}_{i=1}^{T}$ for $p = 1, \ldots, P$ as

$$\boldsymbol{Y}_p^{(i)} = \left[\boldsymbol{y}_p^{(i)}(1), \ldots, \boldsymbol{y}_p^{(i)}(L)\right],$$

where $\boldsymbol{y}_p^{(i)}(l) = \boldsymbol{a}(\Theta_p)s^{(i)}(l) + \boldsymbol{n}^{(i)}(l)$, $s^{(i)}(l) \sim \mathscr{CN}(0, \sigma_s^2)$ and $\boldsymbol{n}^{(i)}(l) \sim \mathscr{CN}(0, \sigma_n^2 \boldsymbol{I})$.

3. Construct the sample covariance matrix $\hat{\boldsymbol{R}}$ and the $K \times K$ covariance matrices $\hat{\boldsymbol{R}}_q^{(i, p)}$ for $q = 1, \ldots, Q$.

4. Compute the CRB values $\eta(\Theta_p, \mathscr{S}_q)$ following (13) and obtain the class set $\mathscr{L}$ representing the best subarrays using (15).

5. Generate the input–output pairs as $\left(\hat{\boldsymbol{R}}^{(i, p)}, z_p^{(i)}\right)$ for $p = 1, \ldots, P$ and $i = 1, \ldots, T$.

6. Construct training data by concatenating the input–output pairs

$$\mathscr{D}_{\text{train}} = \left\{(\hat{\boldsymbol{R}}^{(1, 1)}, z_1^{(1)}), (\hat{\boldsymbol{R}}^{(2, 1)}, z_1^{(2)}), \ldots, (\hat{\boldsymbol{R}}^{(T, 1)}, z_1^{(T)}), \right.$$
$$\left. (\hat{\boldsymbol{R}}^{(1, 2)}, z_2^{(1)}), \ldots, (\hat{\boldsymbol{R}}^{(T, P)}, z_P^{(T)})\right\},$$

where the size of the training dataset is $J = TP$.

Fig. 1 *Algorithm 1: Training data generation*

Section 3.2 and $z \in \mathscr{L}$ is the output label which represents the best subarray index for the covariance input $\boldsymbol{R}$.

We summarise the steps for generating the training data in Algorithm 1 (see Fig. 1). In step 4 of the algorithm, the class $\mathscr{L}$ is chosen from the full combination Q. For large size arrays, one could train by choosing less correlated subarrays or even randomly dropping out some of the subarrays. The covariance matrix used in the computation of the CRB in step 4 is the sample data covariance $\hat{\boldsymbol{R}}_q = (1/L)\sum_{l=1}^{L} \boldsymbol{y}_q(l)\boldsymbol{y}_q^{\text{H}}(l)$ generated with $\text{SNR}_{\text{TRAIN}}$. Even though an analytical expression for $\boldsymbol{R}_q$ is available, we use the sample data covariance here because it is closer to a practical radar operation where, in general, $\boldsymbol{R}_q$ is estimated.

3.2 Network structure and training

Using the labelled training dataset, we build a trained CNN classifier. The input of this learning system is the data covariance and the output is the index of the selected antenna set.

Given the $M \times L$ output $\boldsymbol{Y}$ of the antenna array, the corresponding sample covariance is a complex-valued $M \times M$ matrix $\hat{\boldsymbol{R}}$. The first step toward efficient classification is to define a set of real-valued features that capture the distinguishing aspects of the output. The features we consider in this work are the angle, real and imaginary parts of $\hat{\boldsymbol{R}}$. One could also consider magnitude here

but we did not find much difference in the results when this feature was included. We construct three $M \times M$ real-valued matrices $\{\boldsymbol{X}_c\}_{c=1}^{3}$, whose (i, j)th entry contains, respectively, the phase, real and imaginary parts of the signal covariance matrix $\hat{\boldsymbol{R}}$: $[\boldsymbol{X}_1]_{i, j} = \angle\left[\hat{\boldsymbol{R}}\right]_{i, j}$; $[\boldsymbol{X}_2]_{i, j} = \mathbb{Re}\left\{\left[\hat{\boldsymbol{R}}\right]_{i, j}\right\}$; and $[\boldsymbol{X}_3]_{i, j} = \mathbb{Im}\left\{\left[\hat{\boldsymbol{R}}\right]_{i, j}\right\}$.

Fig. 2 depicts the DL CNN structure that we used. The proposed network consists of nine layers. In the first layer, the CNN accepts the 2D inputs $\{\boldsymbol{X}_c\}_{c=1}^{3}$ in three real-valued channels. The second, fourth and sixth layers are convolutional layers with 64 filters of size 2×2. The third and fifth layers are max-pooling to reduce the dimension by 2. The seventh and eighth layers are fully connected with 1024 units, whose 50% are randomly dropped out to reduce overfitting in training [56]. There are rectified linear units after each convolutional and fully connected layers, where the $\text{ReLU}(x) = \max(x, 0)$. At the output layer, there are Q units wherein the network classifies the given input data using a softmax function and reports the probability distribution of the classes to provide the best subarray.

To train the proposed CNN, we sample the target space for P directions and collect the data for several realisations. We realised the proposed network in MATLAB on a personal computer with 768-core graphics processing unit. During the training process, 90 and 10% of all data generated are selected as the training and validation datasets, respectively. Validation aids in hyperparameter tuning during the training phase to avoid the network simply memorising the training data rather than learning general features for accurate prediction with new (test) data. We used the stochastic gradient descent algorithm with momentum [57] for updating the network parameters with learning rate 0.05 and mini-batch size of 500 samples for 50 epochs. As a loss function, we use the negative log-likelihood or cross-entropy loss. Another useful metric we consider for evaluating the network is the accuracy

$$\text{accuracy}(\%) = \frac{\zeta}{J} \times 100, \tag{16}$$

where J is the total number of input datasets, in which the model identified the best subarrays correctly ζ times. This metric is available for training, validation and test phases.

4 Numerical experiments

In this section, we present numerical experiments to train and test the proposed CNN structure shown in Fig. 2 for different antenna geometries. In the following, we append the subscripts TEST and TRAIN to indicate parameter values used for training and testing modes, respectively. The training data is obtained by sampling the DoA space with P directions, whereas the DoA angles in the test data are uniform randomly selected.

4.1 Uniform linear array

We first analyse the effect of the performance metrics on the antenna selection and DoA estimation accuracy by employing the simplest and most common geometry of an ULA. For creating the training data, we employed three bounds: CRB, BZB and WWB [47]. The network was trained for $M = 10$, $K = 4$, $L_{\text{TRAIN}} = 100$ snapshots, $T_{\text{TRAIN}} = 100$ signal and noise realisations and
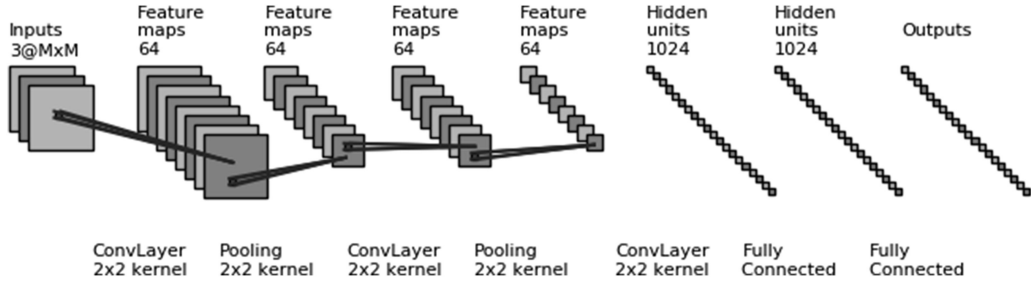
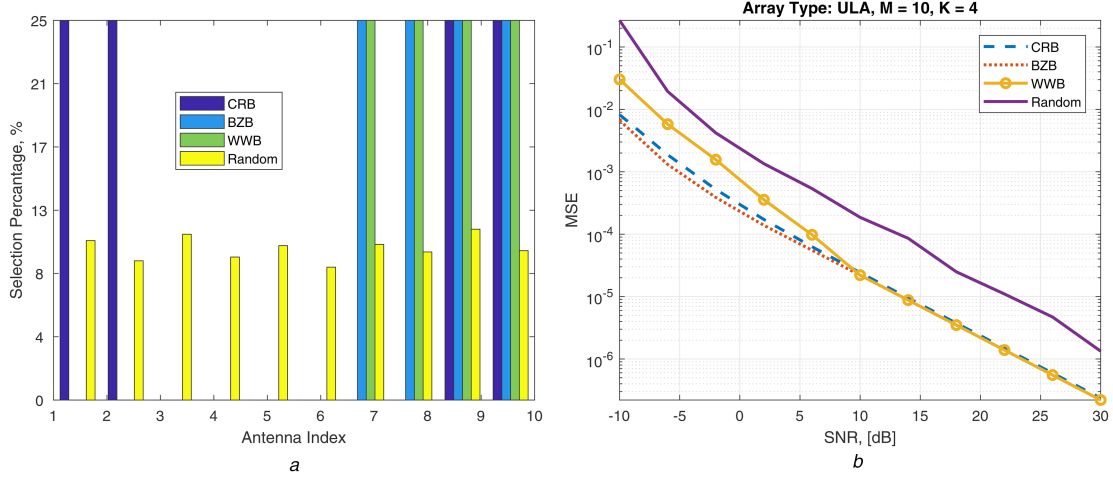**Fig. 2** *CNN structure for antenna selection*

Fig. 3 *Left: antenna selection percentage over $J_{TEST} = 10,000$ trials. Right: MSE of DoA for selected subarrays. The array geometry is an ULA with $M = 10$ and $K = 4$*
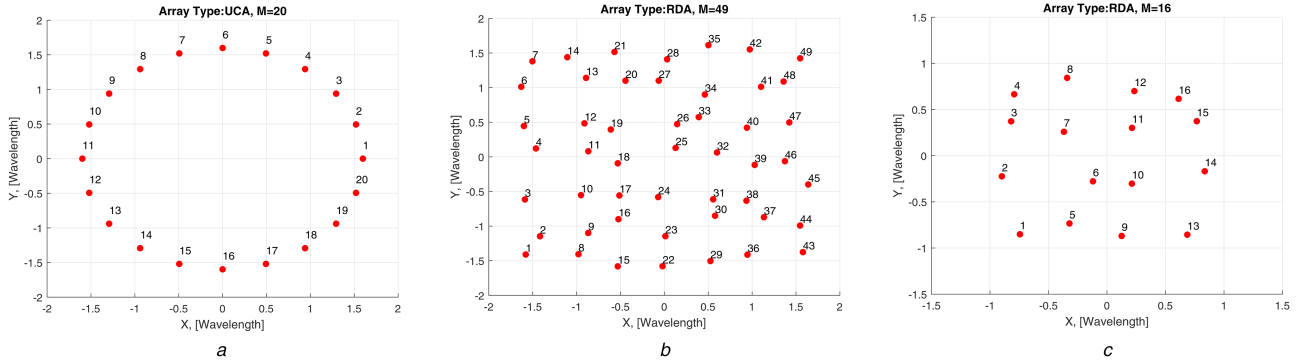

Fig. 4 *Placement of antennas for*
(a) UCA with $M = 20$ elements, *(b)* RDA with $M = 49$, *(c)* RDA with $M = 16$

$P_{\text{TRAIN}} = 100$ DoA angles. The number of uniformly spaced azimuthal grid points is set to $P_\phi = 100$.

For test mode, we fed the network with data corresponding to $P_{\text{TEST}} = 100$ DoA angles different than the ones used in the training phase but keeping the values of M, K, L and T same as in the training. The top plot of Fig. 3 shows the percentage of times a particular antenna index appears as part of the optimal array in the output over $J_{\text{TEST}} = T_{\text{TEST}} P_{\text{TEST}}$ trials with different performance metrics used during training. As seen here, when the CNN is trained with data created from the CRB, the classifier output arrays usually consist of the elements at the extremities. However, the network trained on BZB and WWB usually selects arrays with elements close to each other leading to low sidelobe levels. Also shown here is the random selection, wherein each element is chosen with ~10% selection rate. We provide the DoA estimation performance of the antenna subarrays selected by the network for different values of test data SNRs in the bottom plot of Fig. 3. We observe that, compared to our DL approach, the random thinning results in inferior DoA estimation due to small array aperture. Among various bounds, the MSE is somewhat similar at high SNR regimes with the BZB faring better than CRB at low SNRs.

4.2 2D arrays

We now investigate more complicated array geometries such as UCAs and RDAs. In Table 1, the computed values of Q and the reduced number of classes $\bar{Q}$ are shown for UCAs with $M = 10$ and 16 elements, where $P_\phi = 100$ are uniformly spaced grid points in $[0°, 360°)$ and $\theta = 90°$. We remark that the size of the best subarray classes, $\bar{Q}$, is much less than Q.

4.2.1 Experiment #1: 1D scenario:
We assume that the target and the antenna array are placed in the same plane (i.e. $\theta = 90°$). We consider different array geometries such as UCAs and RDAs as shown in Figs. 4*a* and *b*, respectively. The UCAs consist of $M = 20$ and 45 elements, where each antenna is placed a half wavelength apart from each other. To generate the RDA geometry, we first take a uniform rectangular array of size 7×7, and then perturb the antenna positions as $\{p_{x_m} + \delta_x, p_{y_m} + \delta_y\}_{m=1}^M$, where $\delta_x, \delta_y \sim u\{[-0.1\lambda, 0.1\lambda]\}$.

The training set is constructed with $P_{\text{TRAIN}} = 100$ DoA angles. Note that $P_{\text{TRAIN}} = 100$ is sufficient to train the network for antenna selection in the whole azimuth space. As an illustration of
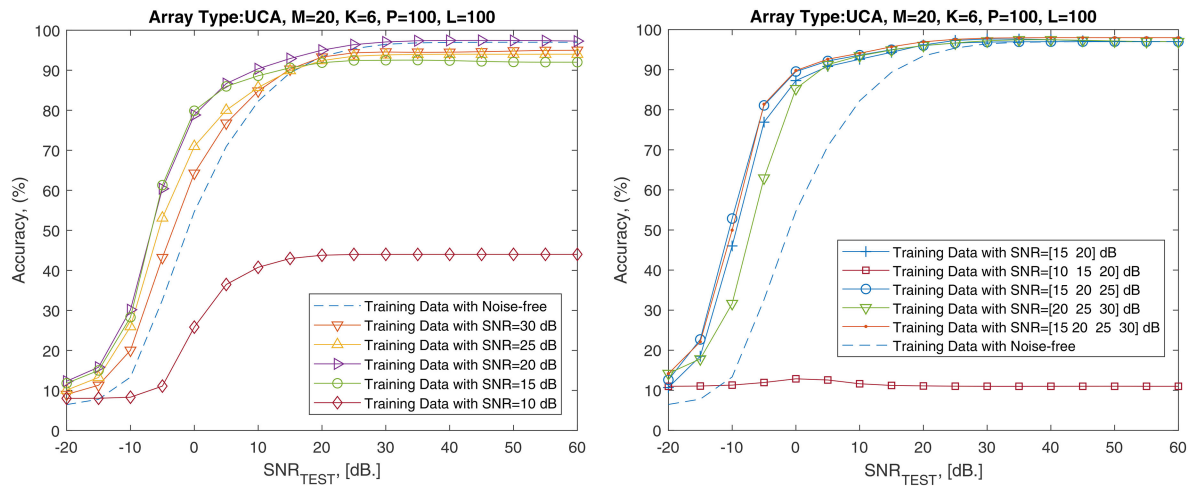
Table 2 Selected antenna indices for UCA with $M = 16$ and $K \in \{3, 4, 5, 6, 7\}$, $\phi \in \{21.42°, 74.28°, 127.14°, 232.85°, 285.71°\}$

	21.42°	74.28°	127.14°	232.85°	285.71°
$k = 3$	{7, 17, 18}	{10, 11, 20}	{3, 13, 14}	{9, 18, 19}	{2, 11, 12}
$k = 4$	{7, 8, 17, 18}	{1, 10, 11, 20}	{3, 4, 13, 14}	{8, 9, 18, 19}	{1, 2, 11, 12}
$k = 5$	{7, 8, 16, 17, 18}	{1, 10, 11, 19, 20}	{3, 4, 12, 13, 14}	{8, 9, 10, 18, 19}	{1, 2, 11, 12, 13}
$k = 6$	{6, 7, 8, 16, 17, 18}	{1, 9, 10, 11, 19, 20}	{2, 3, 4, 12, 13, 14}	{8, 9, 10, 18, 19, 20}	{1, 2, 3, 11, 12, 13}
$k = 7$	{6, 7, 8, 16, 17, 18, 19}	{1, 9, 10, 11, 12, 19, 20}	{2, 3, 4, 5, 12, 13, 14}	{8, 9, 10, 17, 18, 19, 20}	{1, 2, 3, 11, 12, 13, 20}

Number of snapshots $L = 100$ and $\mathrm{SNR_{TRAIN}} = 20$ dB. The antennas are indexed counterclockwise with the first antenna placed as shown in Fig. 4a.

Table 3 Accuracy percentages for training and validation datasets in 1D and 2D scenarios

$\mathrm{SNR_{TRAIN}}$, dB	1D, UCA with $M = 20$, $K = 6$		1D, UCA with $M = 45$, $K = 5$		1D, RDA with $M = 49$, $K = 5$		2D, RDA with $M = 16$, $K = 6$	
	Training, %	Validation, %	Training, %	Validation, %	Training, %	Validation, %	Training, %	Validation, %
10	65.2	68.7	98.7	97.8	97.8	95.7	8.1	10.7
15	98.1	98.5	99.0	98.7	99.9	98.6	60.1	63.2
20	99.2	99.5	100	99.7	97.5	98.1	80.8	80.6
25	99.4	99.8	100	100	100	100	88.9	89.2
30	100	100	100	100	100	100	82.6	83.2
inf	100	100	100	100	100	100	85.0	83.9

Fig. 5 *Performance of test dataset using CNN with respect to $\mathrm{SNR_{TEST}}$. The antenna geometry is an UCA with $M = 20$ and $K = 6$. Left shows the performance when a single SNR value was used during the training phase. Right plot shows the same when multiple SNR values are used for training the network*

the training set generated, Table 2 lists the indices of a few UCA antennas that yield the best CRB for $M = 16$ and different target directions as K varies. We note that the subarrays that yield the largest aperture for the given target direction are selected due to the structure of the UCA. Moreover, note that the same antenna subarrays are selected for the symmetric angles due to the symmetric geometry. The training samples are prepared for different SNR values (i.e. $\mathrm{SNR_{TRAIN}}$) and the accuracy of the training and validation phases is shown in Table 3.

To evaluate the classification performance of the proposed CNN structure, we fed the trained network with the test data generated with $L_{\mathrm{TEST}} = 100$, $T_{\mathrm{TEST}} = 100$ and $P_{\mathrm{TEST}} = 100$ with $\phi_{\mathrm{TEST}} \sim u\{[0°, 360°]\}$. Fig. 5 shows the classification performance of the CNN for $J_{\mathrm{TEST}} = 100$ Monte Carlo trials. The results are given for both the training generated with a single $\mathrm{SNR_{TRAIN}}$ (top) and the multiple $\mathrm{SNR_{TRAIN}}$ (bottom). Fig. 5 also shows the performance of the noisy test data when the network is trained with noise-free dataset; its performance degrades especially at low SNR levels. These observations imply that noisy training datasets should be used for robust classification performance with the test data. On the other hand, when the training data is corrupted with strong noise content (e.g. $\mathrm{SNR_{TRAIN}} \leq 10$ dB), then despite using the noisy training data, the proposed CNN does not recover from poor performance at low $\mathrm{SNR_{TEST}}$ regimes. Similar observations can be made for multiple $\mathrm{SNR_{TRAIN}}$ scenarios, i.e. the network has poor performance if the training data includes the data prepared with

$\mathrm{SNR_{TRAIN}} = 10$ dB. This leads to the conclusion that the training data should not include too much noise. While there is a slight difference comparing the single and multiple $\mathrm{SNR_{TRAIN}}$ cases for high $\mathrm{SNR_{TEST}}$ regimes, CNN performs better in low SNR (i.e. $\mathrm{SNR_{TEST}} \leq 10$ dB) if multiple SNRs are used in the training data. Specifically, the training data generated with $\mathrm{SNR_{TRAIN}} = [15, 20, 25, 30]$ dB provides the best result for a large range of $\mathrm{SNR_{TEST}}$.

The performance at low SNRs can be improved when the size of the array increases and, as a result, the input data is huge and the SNR is enhanced due to large M. As an example, Fig. 6 illustrates the performance of the network for UCA with $M = 45$ and $K = 5$, where the network provides high accuracy for a wide range of $\mathrm{SNR_{TEST}}$ compared with the scenario in Fig. 5.

We also compared CNN with the SVM technique (as in [39]), where we used identical parameters for the data generation and identical data covariance input to the SVM. The performance of SVM is shown in Fig. 7. We observe from Figs. 5–7 that CNN is more than 90% accurate for $\mathrm{SNR_{TEST}} \geq 10$ dB when the network is trained by datasets with $\mathrm{SNR_{TRAIN}} \geq 15$ dB. In comparison, SVM performs poorly being unable to extract the features as efficiently as CNN.

Similar experimental results for an RDA (Fig. 4b) with $M = 49$ and $K = 5$ shown in Fig. 8. The training dataset is prepared with the same parameters as in the previous experiment with UCA. For some selected cases, the accuracies of training and validation data
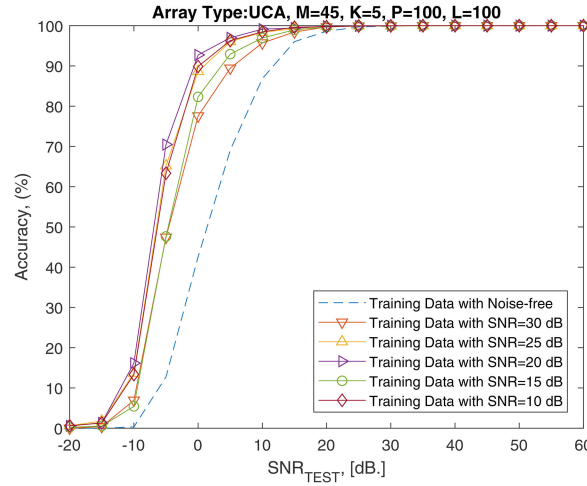
Fig. 6 *Performance of test data using CNN with respect to SNR when the training data is prepared with $SNR_{TRAIN} \in \{10, 15, 20, 25, 30, \text{inf}\}$ dB. The antenna geometry is an UCA with $M = 45$ and $K = 5$*
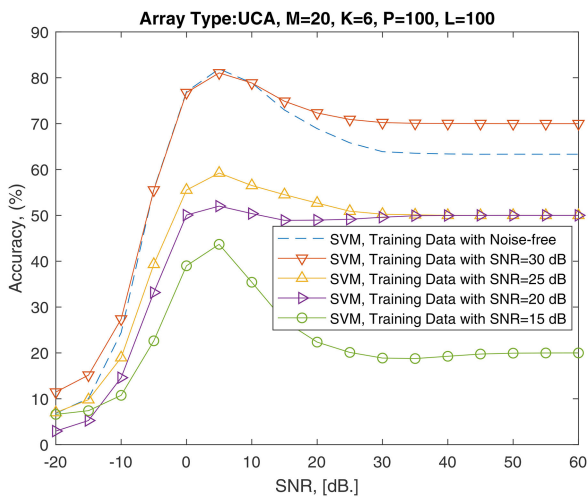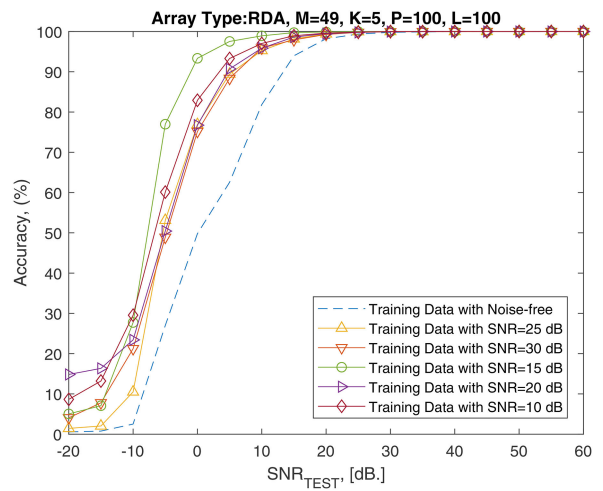

Fig. 7 *Performance of test dataset using SVM with respect to SNR_{TEST} when the training data is prepared with $SNR_{TRAIN} \in \{10, 15, 20, 25, 30, \text{inf}\}$ dB. The antenna geometry is an UCA with $M = 20$ and $K = 6$*

Fig. 8 *Performance of test data using CNN with respect to SNR when the training data is prepared with SNR_{TRAIN} levels of 5, 10, 15, 20, 25 and 30 dB as well as without any noise. The antenna geometry is an RDA with $M = 49$ shown in Fig. 4b and $K = 5$*

for RDA are listed in Table 3. The network achieves high accuracy when M is large and $SNR_{TEST} \geq 10 \, dB$.

4.2.2 Experiment #2: 2D scenario: Finally, we consider cases when the target and the antenna array are not coplanar. We train the CNN structure in Fig. 2 with the data generated for $T = 100$ and $L = 100$. The angles lie on a uniformly spaced elevation and azimuth grids in the planes [90°, 100°] and [0°, 360°), respectively. We set the number of grid points in the elevation and azimuth to $P_\theta = 11$ and $P_\phi = 100$, respectively.

In this experiment, we use an RDA (Fig. 4c) with $M = 16$ and $K = 6$. Table 4 lists the indices of a few RDA antennas that yield the best CRB for different target DoAs: $\phi \in \{30°, 60°, 90°, 120°, 210°\}$ and $\theta \in \{90°, 92°\}$ as K varies. When K increases, a subarray with a larger aperture is to be selected for better DoA estimation performance. When there is even a slight change in the elevation angle, the best subarray changes completely because of the relatively small subarray aperture in the elevation dimension. We prepared the training and validation datasets for different SNR_{TRAIN} values. The accuracies of the two stages are listed in Table 3. We note that the training accuracy of the network in the 2D case is worst than the 1D scenario of RDA because simple 2D arrays are unable to distinguish all elevation angles. As a result, training data samples that are very similar to each other are labelled to different classes with different elevation angles.

We generated a test dataset with $L_{TEST} = 100$ and $T_{TEST} = 100$ to evaluate the CNN for a 2D target scenario. The target directions were drawn from $\phi_{TEST} \sim u\{[0°, 360°]\}$ and $\theta_{TEST} \sim u\{[80°, 100°]\}$ for $P_{TEST} = 100$. The accuracies of the test mode for $J_{TEST} = 100$ trials is shown in Fig. 9 for different SNR_{TEST} levels. Fig. 9 shows that the training datasets with $SNR_{TRAIN} \geq 15 \, dB$ providing sufficiently good performance with an accuracy of ~85% for $SNR_{TEST} \geq 10 \, dB$. However, as seen earlier, poor classification performance results when SNR_{TRAIN} is low (e.g. $\leq 10 \, dB$).

4.2.3 Experiment #3: DoA estimation performance: In this experiment, the DoA estimation performance of the proposed method is presented. Our CNN approach is compared with SVM and random antenna selection (RAS) algorithms. The selected antenna subarrays from CNN and SVM are inserted to the beamforming technique [58] for DoA estimation. As a traditional technique, we consider the RAS algorithm where, instead of all subarray candidates, a number of subarray geometries are realised randomly (i.e. 1000 realisations) and their beamforming spectra are obtained by a search algorithm [59]. We also added the full array performance where $M = K$ for comparison. In Fig. 10, the results are given for an UCA with $M = 20$ and $K = 6$ antennas to be selected. Here, 'best subarray' denotes the beamforming performance of the subarray that gives the lowest CRB. It can be seen that CNN provides better performance as compared with

Table 4 Selected antenna indices for a random array with $M = 16$ and $K \in \{3, 4, 5, 6, 7\}$, $\phi \in \{30°, 60°, 90°, 120°, 210°\}$ and $\theta \in \{90°, 92°\}$

	$\theta = 90°, \phi = 30°$	$\theta = 90°, \phi = 60°$	$\theta = 90°, \phi = 90°$	$\theta = 90°, \phi = 120°$	$\theta = 90°, \phi = 210°$
$k = 3$	{4, 8, 13}	{3, 4, 13}	{2, 14, 15}	{1, 15, 16}	{4, 8, 13}
$k = 4$	{4, 8, 9, 13}	{3, 4, 13, 14}	{2, 3, 14, 15}	{1, 2, 15, 16}	{4, 8, 9, 13}
$k = 5$	{3, 4, 8, 9, 13}	{3, 4, 8, 13, 14}	{2, 3, 4, 14, 15}	{1, 2, 14, 15, 16}	{2, 3, 4, 14, 15}
$k = 6$	{3, 4, 8, 9, 13, 14}	{3, 4, 8, 9, 13, 14}	{2, 3, 4, 13, 14, 15}	{1, 2, 5, 14, 15, 16}	{3, 4, 8, 9, 13, 14}
$k = 7$	{3, 4, 5, 8, 9, 13, 14}	{2, 3, 4, 8, 9, 13, 14}	{1, 2, 3, 4, 13, 14, 15}	{1, 2, 5, 12, 14, 15, 16}	{1, 2, 3, 4, 13, 14, 15}

	$\theta = 92°, \phi = 30°$	$\theta = 92°, \phi = 60°$	$\theta = 92°, \phi = 90°$	$\theta = 92°, \phi = 120°$	$\theta = 92°, \phi = 210°$
$k = 3$	{1, 2, 15}	{1, 8, 16}	{1, 8, 9}	{4, 12, 13}	{1, 2, 15}
$k = 4$	{1, 2, 14, 16}	{1, 9, 12, 16}	{1, 8, 12, 13}	{4, 8, 9, 13}	{1, 2, 14, 16}
$k = 5$	{1, 2, 14, 15, 16}	{1, 5, 8, 12, 16}	{1, 8, 9, 12, 13}	{4, 8, 9, 12, 13}	{1, 2, 14, 15, 16}
$k = 6$	{1, 2, 5, 14, 15, 16}	{1, 5, 9, 12, 15, 16}	{1, 8, 9, 12, 13, 16}	{3, 4, 5, 8, 9, 13}	{1, 8, 9, 12, 13, 16}
$k = 7$	{1, 2, 3, 5, 14, 15, 16}	{1, 5, 8, 9, 12, 15, 16}	{1, 4, 8, 9, 12, 13, 16}	{3, 4, 5, 8, 9, 12, 13}	{1, 2, 3, 5, 14, 15, 16}

Number of snapshots $L = 100$ and $SNR_{TRAIN} = 20$ dB. The antenna indices are given in Fig. 4c.

SVM (32% more accurate) and RAS (72%) and it approaches the performance of the 'best subarray' as expected from the accuracy results given in Fig. 5. SVM performs poorer due to its lower antenna selection accuracy. We present 2D DoA estimation results in Fig. 11 for RDA with $M = 16$ and $K = 6$ with the same settings as for Fig. 9. Similar observations are obtained for the 2D case as compared with the 1D scenario.

We further compare the DoA estimation performance of the selected subarrays with full array ($M = K$) performance in both Figs. 10 and 11. While there is a gap between subarray and the full array performances, antenna selection provides less computation and cost.

In Fig. 12, the DoA estimation performance over time is presented for different antenna selection algorithms. In this test, SNR is first increased from 0 to 20 dB then dropped by 10 dB for every 1000 data snapshot blocks. The target location varies for each 500 data snapshot blocks. In each block, the first 100 snapshots are used for antenna selection (all antennas are used). After antenna selection, the selected antennas are used for DoA estimation (only K antennas are in use) for the next data blocks. While the algorithms have the robust performance to the change in the target DoA, the CNN has the lowest root-mean-square error as compared with the others.

4.3 Computational complexity

The computation times for the algorithms are given in Table 5 in seconds. To fairly compare the algorithms, the results are calculated to include both antenna selection and DoA estimation phases. The computation time only for the beamforming is 0.0384 s. In terms of classification, the computation times for CNN and SVM are 0.0037 and 0.0609 s, respectively, for $M = 20$ and $K = 6$. As a result, CNN provides much faster results and accuracy as compared with both SVM and the conventional DoA estimation technique based on beamforming. The complexity of RAS is due to the computation of the DoA spectra for each subarray realisation (1000 realisations were used for RAS in Table 5). We also compared the computation time for DoA estimation with full array and the CNN with K antennas. We observed that DoA estimation with full array took 0.14 s, whereas CNN takes 0.0535 s of which 0.0035 s used for classification. These results show that the proposed CNN approach provides less computational complexity together with the loss in the DoA estimation performance due to the use of less number of antennas as compared with the full array.

5 Discussion and summary

We introduced a method based on DL to select antennas in a cognitive radar scenario. We constructed a deep NN with convolutional layers as a multi-class classification framework. The training data was generated such that each class indicated an
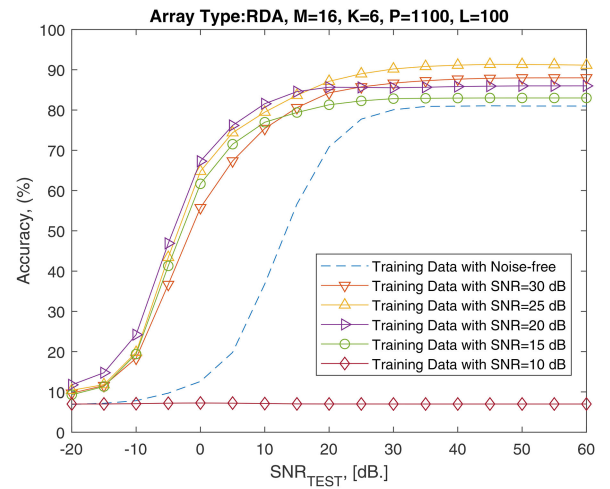

Fig. 9 *Performance of test data using CNN with respect to SNR when the training data is prepared with SNR_{TRAIN} levels of 5, 10, 15, 20, 25 and 30 dB as well as without any noise. The antenna geometry is an RDA with $M = 16$ shown in Fig. 4c and $K = 6$*
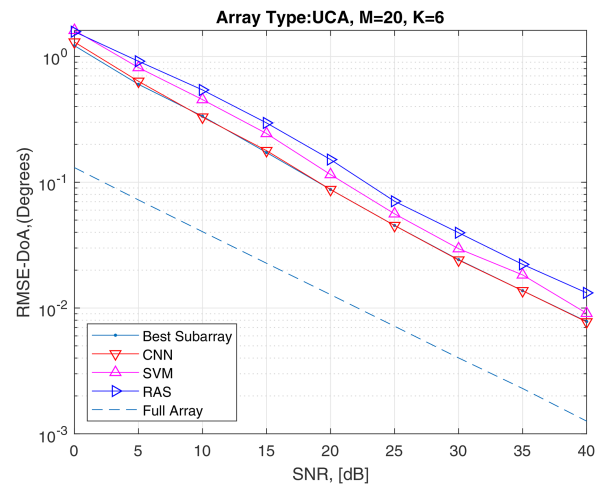

Fig. 10 *DoA estimation performance with respect to SNR. $SNR_{TRAIN} = 20$ dB. The antenna geometry is an UCA with $M = 20$ and $K = 6$*

antenna subarray that provides the lowest minimal error bound for estimating target DoA in a given scenario. Our learning network then cognitively determines a new array whenever the radar receiver acquires echoes from the target scene. We evaluated the
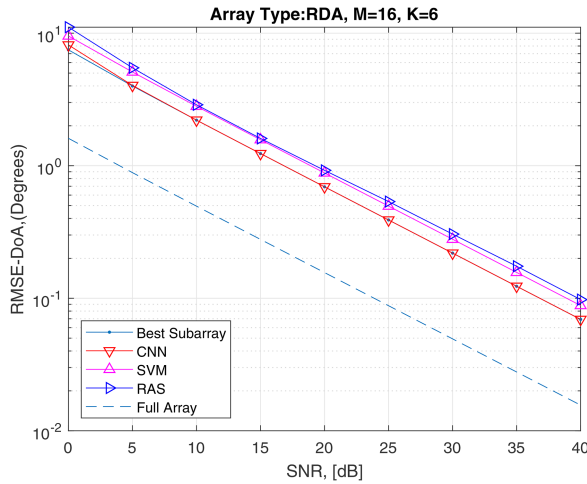
Fig. 11 *DoA estimation performance with respect to SNR.* $SNR_{TRAIN} = 20\,dB$. *The antenna geometry is an RDA with $M = 16$ and $K = 6$*
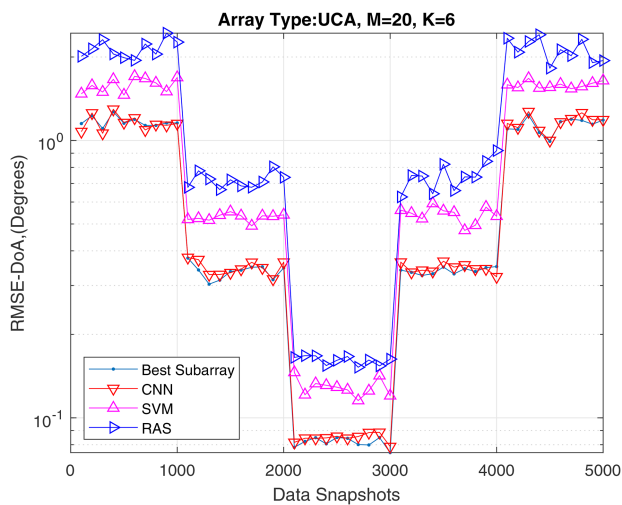

Fig. 12 *DoA estimation performance over time with $SNR_{TRAIN} = 20\,dB$. The antenna geometry is an UCA with $M = 20$ and $K = 6$. SNR_{TEST} changes over time as $[0, 10, 20, 10, 0]^T\,dB$*

Table 5 Computation time in seconds

	CNN	SVM	RAS
computation time	0.0414s	0.0984s	28.7862s

Results include antenna selection and DoA estimation complexity.

performance of the proposed approach for both 1D (azimuth) and 2D (azimuth and elevation) target scenarios using ULA, UCA and RDA structures. The results show the enhanced performance of the proposed network over conventional randomly thinned arrays as well as the traditional SVM-based selection. Our method does not depend on the array geometry and selects optimal antenna subarrays for DoA estimation. As with other classifiers such as SVM, the classification accuracy of our CNN degrades in low SNR conditions because the network cannot distinguish the array data of different classes and, consequently, predicts false results. We were able to partially mitigate this issue by training the network with noisy data samples. We design the training data with both single and multiple SNR levels of noisy data to investigate further the performance. The results show that there is a slight difference in the classification accuracy after including data at multiple SNRs during the training. The proposed CNN structure provides 32% better classification than an SVM and the resulting subarrays yield 72% more accurate DoA estimate than a random array selection. The combined computation time required by CNN for the antenna

selection and DoA estimation was half of that taken by SVM and three orders of magnitude smaller than a random selection.

Although the CNN predicts an optimal subarray for 1D scenarios very well, its performance degrades for 2D cases. This is expected because the simple 2D arrays we considered are unable to distinguish all elevation angles, and thereby lead to some misclassification. We reserve further investigations of this issue for the future.

6 Acknowledgments

7 References

[1] Haykin, S.: 'Cognitive radar: a way of the future', *IEEE Signal Process. Mag.*, 2006, **23**, (1), pp. 30–40
[2] Guerci, J.R.: 'Cognitive radar: a knowledge-aided fully adaptive approach'. IEEE Radar Conf., Washington, DC, USA, 2010, pp. 1365–1370
[3] Smith, G.E., Cammenga, Z., Mitchell, A., *et al.*: 'Experiments with cognitive radar', *IEEE Aerosp. Electron. Syst. Mag.*, 2016, **31**, (12), pp. 34–46
[4] Chen, P., Wu, L.: 'Waveform design for multiple extended targets in temporally correlated cognitive radar system', *IET Radar Sonar Navig.*, 2016, **10**, (2), pp. 398–410
[5] Kilani, M.B., Nijsure, Y., Gagnon, G., *et al.*: 'Cognitive waveform and receiver selection mechanism for multistatic radar', *IET Radar Sonar Navig.*, 2016, **10**, (2), pp. 417–425
[6] Mishra, K.V., Eldar, Y.C.: 'Performance of time delay estimation in a cognitive radar'. IEEE Int. Conf. Acoustics, Speech and Signal Processing, New Orleans, LA, USA, 2017, pp. 3141–3145
[7] Mishra, K.V., Eldar, Y.C., Shoshan, E., *et al.*: 'A cognitive sub-Nyquist MIMO radar prototype', arXiv preprint arXiv:1807.09126, 2018
[8] Bell, K.L., Baker, C.J., Smith, G.E., *et al.*: 'Cognitive radar framework for target detection and tracking', *IEEE. J. Sel. Top. Signal Process.*, 2015, **9**, (8), pp. 1427–1439
[9] Goodman, N.A., Venkata, P.R., Neifeld, M.A.: 'Adaptive waveform design and sequential hypothesis testing for target recognition with active sensors', *IEEE. J. Sel. Top. Signal Process.*, 2007, **1**, (1), pp. 105–113
[10] Stinco, P., Greco, M.S., Gini, F.: 'Spectrum sensing and sharing for cognitive radars', *IET Radar Sonar Navig.*, 2016, **10**, (3), pp. 595–602
[11] Cohen, D., Mishra, K.V., Eldar, Y.C.: 'Spectrum sharing radar: coexistence via Xampling', *IEEE Trans. Aerosp. Electron. Syst.*, 2018, **29**, pp. 1279–1296
[12] Mishra, K.V., Eldar, Y.C.: 'Sub-Nyquist radar: principles and prototypes', arXiv preprint arXiv:1803.01819, 2018
[13] Cohen, D., Eldar, Y.C.: 'Sub-Nyquist radar systems: temporal, spectral and spatial compression', *IEEE Signal Process. Mag.*, 2018, **35**, (6), pp. 35–58
[14] Na, S., Mishra, K.V., Liu, Y., *et al.*: 'TenDSuR: tensor-based sub-Nyquist radar', *IEEE Signal Process. Lett.*, 2019, **26**, (2), pp. 237–241
[15] Baylis, C., Fellows, M., Cohen, L., *et al.*: 'Solving the spectrum crisis: intelligent, reconfigurable microwave transmitter amplifiers for cognitive radar', *IEEE Microw. Mag.*, 2014, **15**, (5), pp. 94–107
[16] Lo, Y.: 'A mathematical theory of antenna arrays with randomly spaced elements', *IEEE Trans. Antennas Propag.*, 1964, **12**, (3), pp. 257–268
[17] Duman, T.M., Ghrayeb, A.: 'Antenna selection for MIMO systems', in (Eds.): '*Coding for MIMO communication systems*' (John Wiley & Sons, Oxford, UK, 2007), pp. 287–315
[18] Gershman, A., Böhme, J.F.: 'A note on most favorable array geometries for DOA estimation and array interpolation', *IEEE Signal Process. Lett.*, 1997, **4**, (8), pp. 232–235
[19] Pal, P., Vaidyanathan, P.P.: 'Coprime sampling and the MUSIC algorithm'. Digital Signal Processing and Signal Processing Education Meeting, Sedona, AZ, USA, 2011, pp. 289–294
[20] Pal, P., Vaidyanathan, P.P.: 'Nested arrays: a novel approach to array processing with enhanced degrees of freedom', *IEEE Trans. Signal Process.*, 2010, **58**, (8), pp. 4167–4181
[21] Mishra, K.V., Kahane, I., Kaufmann, A., *et al.*: 'High spatial resolution radar using thinned arrays'. IEEE Radar Conf., Seattle, WA, USA, 2017, pp. 1119–1124
[22] Rossi, M., Haimovich, A.M., Eldar, Y.C.: 'Spatial compressive sensing for MIMO radar', *IEEE Trans. Signal Process.*, 2014, **62**, (2), pp. 419–430
[23] Cohen, D., Cohen, D., Eldar, Y.C., *et al.*: 'SUMMer: sub-Nyquist MIMO radar', *IEEE Trans. Signal Process.*, 2018, **66**, pp. 4315–4330
[24] Tabrikian, J., Isaacs, O., Bilik, I.: 'Cognitive antenna selection for DOA estimation in automotive radar'. IEEE Radar Conf., Philadelphia, PA, USA, 2016, pp. 1–5
[25] Isaacs, O., Tabrikian, J., Bilik, I.: 'Cognitive antenna selection for optimal source localization'. IEEE Int. Workshop on Computational Advances in Multi-Sensor Adaptive Processing, Cancun, Mexico, 2015, pp. 341–344
[26] Mateos-Núñez, D., González-Huici, M.A., Simoni, R., *et al.*: 'Adaptive channel selection for DOA estimation in MIMO radar'. Int. Workshop on Computational Advances in Multi-Sensor Adaptive Processing, Curacao, Netherlands Antilles, 2017, in press

[27] Shulkind, G., Jegelka, S., Wornell, G.W.: 'Multiple wavelength sensing array design'. IEEE Int. Conf. Acoustics, Speech and Signal Processing, New Orleans, LA, USA, 2017, pp. 3424–3428

[28] Nosrati, H., Aboutanios, E., Smith, D.B.: 'Receiver–transmitter pair selection in MIMO phased array radar'. IEEE Int. Conf. Acoustics, Speech and Signal Processing, New Orleans, LA, USA, 2017, pp. 3206–3210

[29] Molisch, A.F., Ratnam, V.V., Han, S., *et al.*: 'Hybrid beamforming for massive MIMO: a survey', *IEEE Commun. Mag.*, 2017, **55**, (9), pp. 134–141

[30] Zhai, X., Cai, Y., Shi, Q., *et al.*: 'Joint transceiver design with antenna selection for large-scale MU-MIMO mmWave systems', *IEEE J. Sel. Areas Commun.*, 2017, **35**, (9), pp. 2085–2096

[31] Wang, Z., Vandendorpe, L.: 'Antenna selection for energy efficient MISO systems', *IEEE Commun. Lett.*, 2017, **21**, (12), pp. 2758–2761

[32] Demir, O.T., Tuncer, T.E.: 'Antenna selection and hybrid beamforming for simultaneous wireless information and power transfer in multi-group multicasting systems', *IEEE Trans. Wirel. Commun.*, 2016, **15**, (10), pp. 6948–6962

[33] Joshi, S., Boyd, S.: 'Sensor selection via convex optimization', *IEEE Trans. Signal Process.*, 2009, **57**, (2), pp. 451–462

[34] Roy, V., Chepuri, S.P., Leus, G.: 'Sparsity-enforcing sensor selection for DOA estimation'. IEEE Int. Workshop on Computational Advances in Multi-Sensor Adaptive Processing, St. Martin, France, 2013, pp. 340–343

[35] Godrich, H., Petropulu, A.P., Poor, H.V.: 'Sensor selection in distributed multiple-radar architectures for localization: a knapsack problem formulation', *IEEE Trans. Signal Process.*, 2012, **60**, (1), pp. 247–260

[36] Metcalf, J., Blunt, S.D., Himed, B.: 'A machine learning approach to cognitive radar detection'. IEEE Radar Conf., Arlington, VA, USA, 2015, pp. 1405–1411

[37] Gao, Y., Hu, D., Chen, Y., *et al.*: 'Gridless 1-b DOA estimation exploiting SVM approach', *IEEE Commun. Lett.*, 2017, **21**, (10), pp. 2210–2213

[38] Gonnouni, A.E., Martinez-Ramon, M., Rojo-Alvarez, J.L., *et al.*: 'A support vector machine MUSIC algorithm', *IEEE Trans. Antennas Propag.*, 2012, **60**, (10), pp. 4901–4910

[39] Joung, J.: 'Machine learning-based antenna selection in wireless communications', *IEEE Commun. Lett.*, 2016, **20**, (11), pp. 2241–2244

[40] Lecun, Y., Bengio, Y., Hinton, G.: 'Deep learning', *Nature*, 2015, **521**, (7553), pp. 436–444

[41] Bengio, Y., Courville, A., Vincent, P.: 'Representation learning: a review and new perspectives', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2013, **35**, (8), pp. 1798–1828

[42] Wang, C., Wang, J., Zhang, X.: 'Automatic radar waveform recognition based on time–frequency analysis and convolutional neural network'. IEEE Int. Conf. Acoustics, Speech and Signal Processing, New Orleans, LA, USA, 2017, pp. 2437–2441

[43] Mason, E., Yonel, B., Yazici, B.: 'Deep learning for radar'. IEEE Radar Conf., Seattle, WA, USA, 2017, pp. 1703–1708

[44] Schwegmann, C.P., Kleynhans, W., Salmon, B.P., *et al.*: 'Very deep learning for ship discrimination in synthetic aperture radar imagery'. IEEE Int. Geoscience and Remote Sensing Symp., Beijing, China, 2016, pp. 104–107

[45] Jokanović, B., Amin, M.: 'Fall detection using deep learning in range–Doppler radars', *IEEE Trans. Aerosp. Electron. Syst.*, 2018, **54**, (1), pp. 180–189

[46] Mishra, K.V., Gharanjik, A., Shankar, M.R.B., *et al.*: 'Deep learning framework for precipitation retrievals from communication satellites'. European Conf. Radar in Meteorology & Hydrology, Ede, Wageningen, Netherlands, 2018

[47] Renaux, A., Forster, P., Larzabal, P., *et al.*: 'A fresh look at the Bayesian bounds of the Weiss–Weinstein family', *IEEE Trans. Signal Process.*, 2008, **56**, (11), pp. 5334–5352

[48] Jiang, T., Sidiropoulos, N.D., ten Berge, J. .: 'Almost-sure identifiability of multidimensional harmonic retrieval', *IEEE Trans. Signal Process.*, 2001, **49**, (9), pp. 1849–1859

[49] Liu, X., Sidiropoulos, N.D.: 'On constant modulus multidimensional harmonic retrieval'. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, Orlando, FL, USA, 2002, vol. **3**, pp. III–2977

[50] Nion, D., Sidiropoulos, N.D.: 'Tensor algebra and multidimensional harmonic retrieval in signal processing for MIMO radar', *IEEE Trans. Signal Process.*, 2010, **58**, (11), pp. 5693–5705

[51] Ben-Haim, Z., Eldar, Y.C.: 'A lower bound on the Bayesian MSE based on the optimal bias function', *IEEE Trans. Inf. Theory*, 2009, **55**, (11), pp. 5179–5196

[52] Stoica, P., Nehorai, A.: 'MUSIC, maximum likelihood, and Cramér–Rao bound: further results and comparisons', *IEEE Trans. Acoust. Speech Signal Process.*, 1990, **38**, (12), pp. 2140–2150

[53] Friedlander, B., Weiss, A.: 'Direction finding in the presence of mutual coupling', *IEEE Trans. Antennas Propag.*, 1991, **39**, (3), pp. 273–284

[54] Ye, Z., Liu, C.: '2-D DOA estimation in the presence of mutual coupling', *IEEE Trans. Antennas Propag.*, 2008, **56**, (10), pp. 3150–3158

[55] Bilal, A., Jourabloo, A., Ye, M., *et al.*: 'Do convolutional neural networks learn class hierarchy?', *IEEE Trans. Vis. Comput. Graph.*, 2018, **24**, pp. 152–162

[56] Srivastava, N., Hinton, G., Krizhevsky, A., *et al.*: 'Dropout: a simple way to prevent neural networks from overfitting', *J. Mach. Learn. Res.*, 2014, **15**, (1), pp. 1929–1958

[57] Bishop, C.M.: '*Pattern recognition and machine learning*' (Springer, New York, 2006)

[58] Dudgeon, D.E.: 'Fundamentals of digital array processing', *Proc. IEEE*, 1977, **65**, pp. 898–904

[59] Athley, F.: 'Optimization of element positions for direction finding with sparse arrays'. Proc. 11th IEEE Signal Processing Workshop on Statistical Signal Processing (Cat. No. 01TH8563), Singapore, August 2001, pp. 516–519