Transmit Antenna Selection in MIMO Wiretap Channels: A Machine Learning Approach

Dongxuan He, Chenxi Liu, *Member, IEEE*, Tony Q. S. Quek, *Fellow, IEEE*, and Hua Wang, *Member, IEEE*

***Abstract*—In this letter, we exploit the potential benefits of machine learning in enhancing physical layer security in multi-input multi-output multi-antenna-eavesdropper wiretap channels. To this end, we focus on the scenario where the source adopts transmit antenna selection (TAS) as the transmission strategy. We assume that the channel state information (CSI) of the legitimate receiver is available to the source, while the CSI of the eavesdropper can be either known or not known at the source. By modeling the problem of TAS as a multiclass classification problem, we propose two machine learning-based schemes, namely, the support vector machine-based scheme and the naive-Bayes-based scheme, to select the optimal antenna that maximizes the secrecy performance of the considered system. Compared to the conventional TAS scheme, we show that our proposed schemes can achieve almost the same secrecy performance with relatively small feedback overhead. The work presented here provides insights into the design of new machine learning-based secure transmission schemes.**

***Index Terms*—Machine learning, physical layer security, transmit antenna selection, support vector machine, naive-Bayes.**

I. Introduction

PHYSICAL layer security has been regarded as a promising supplementary component to traditional key-based cryptographic techniques applied to upper layers, due to the fact that it can provide secure data transmissions without requiring secret keys and complex algorithms [1]. This paradigm was initialized by the pioneering work of Wyner [2], in which it was shown how secrecy can exist in single-input single-output wiretap channels. Motivated by the benefits provided by multi-input multi-output (MIMO) techniques, such as high reliability and high data rate, physical layer security in MIMO wiretap channels has recently attracted significant research attention [3]–[7].

On the other hand, machine learning has been recognized as an emerging technique and has been shown to be effective in a wide range of applications from image processing to economics [8], due to its suitability for classification and decision making based on limited information. Since applying MIMO techniques, such as beamforming [3], artificial noise [4], and transmit antenna selection (TAS) [5]–[7], in physical layer security can also be characterized as a classification problem and/or a decision making problem, then a natural question to ask is: "How machine learning can be utilized to improve physical layer security in wireless communications?" To answer this question, in this letter we leverage machine learning to enhance physical layer security in multi-input multi-output multi-antenna-eavesdropper (MIMOME) wiretap channels. We assume that the source adopts TAS as the transmission strategy. We first characterize the problem of TAS as a multi-class classification problem. Then we propose two machine learning-based schemes to select the optimal antenna that maximizes the secrecy performance of the considered system. The proposed schemes are based on support vector machine (SVM) [9] and naive-Bayes (NB) [10], respectively. We show that our proposed schemes can achieve almost the same secrecy performance as that of the conventional TAS scheme, while requiring a smaller feedback overhead.

We note that the machine learning based TAS has recently been presented in [11], where SVM and k-nearest neighbors (KNN) were applied to select the optimal antenna that minimizes the bit error rate of the system without secrecy constraints. Compared to [11], this letter is different as follows: 1) we focus on examining how machine learning can be used to enhance physical layer security of MIMOME wiretap channels; 2) Instead of using KNN, we propose a NB-based scheme, thereby providing a better performance with relatively low implementation complexity.

II. System Model and Conventional Transmit Antenna Selection Scheme

We consider a wiretap channel, where a source transmits to a legitimate receiver in the presence of an eavesdropper. The source, the legitimate receiver, and the eavesdropper are equipped with N_s, N_r and N_e antennas, respectively. We assume that all the channels are subject to identical and independent distributed (i.i.d) Rayleigh fading. We denote $\mathbf{H} \in \mathbb{C}^{N_r \times N_s}$ and $\mathbf{G} \in \mathbb{C}^{N_e \times N_s}$ as the main channel between the source and the legitimate receiver and the channel between the source and the eavesdropper, respectively. We also assume that $\mathbf{H}$ is known at the source, while $\mathbf{G}$ can be either known or not known at the source. Henceforth, we refer the case where $\mathbf{G}$ is known at the source to as the full CSI case, and the case where $\mathbf{G}$ is not known at the source to as the partial CSI case.

Manuscript received November 7, 2017; revised January 3, 2018; accepted February 8, 2018. Date of publication February 14, 2018; date of current version August 21, 2018. This work was supported in part by the China Scholarship Council (2017) under Grant 3109, and in part by the National Natural Science Foundation of China under Grant 61471037 and Grant 61201181. The associate editor coordinating the review of this paper and approving it for publication was P. P. Markopoulos. *(Corresponding author: Chenxi Liu.)*

D. He and H. Wang are with the School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China (e-mail: hdxbit@bit.edu.cn; wanghua@bit.edu.cn).

C. Liu and T. Q. S. Quek are with the Singapore University of Technology and Design, Singapore 487372 (e-mail: chenxi_liu@sutd.edu.sg; tonyquek@sutd.edu.sg).

Digital Object Identifier 10.1109/LWC.2018.2805902

We consider that the source adopts TAS as the transmission strategy. As such, in the following we detail the conventional TAS scheme [5]–[7]. In the conventional TAS scheme, the source selects one of its N_s antennas that maximizes the secrecy performance to transmit the confidential message. Then, the legitimate receiver and the eavesdropper adopt maximal ratio combining [12] as the receiving strategy. In our system, we consider both the full CSI case and the partial CSI case. In order to evaluate the secrecy performance of our system, we adopt the achievable secrecy rate [3] and the secrecy outage probability (SOP) [5] as the performance metric for the full CSI case and the partial CSI case, respectively. As per the rules of the conventional TAS scheme, the index of the selected antenna is expressed as

$$n^* = \begin{cases} \text{argmax}_{1\leq n\leq N_s}\ C_{s,n}, & \text{for the full CSI case} \\ \\ \text{argmin}_{1\leq n\leq N_s}\ \|\mathbf{h}_n\|, & \text{for the partial CSI case.} \end{cases} \tag{1}$$

In (1), $C_{s,n} = \max(C_{b,n} - C_{e,n}, 0)$ denotes the achievable secrecy rate of our system when the n-th antenna is selected, where $C_{b,n}$ and $C_{e,n}$ denotes the achievable rate of the main channel and the achievable rate of the eavesdropper's channel when the n-th antenna is selected, respectively, and $\mathbf{h}_n$ denotes the n-th column of $\mathbf{H}$. Correspondingly, the secrecy outage probability achieved by the n-th antenna is defined as the probability that $C_{s,n}$ is no larger than a certain target secrecy rate R_s. Mathematically, it is given by

$$P_{so,n}(R_s) = \Pr\big(C_{s,n} \leq R_s\big). \tag{2}$$

After the selection, the maximum achievable secrecy rate for the full CSI case and the minimum secrecy outage probability for the partial CSI case achieved by the n^*-th antenna are denoted by C_s^* and $P_{so}^*(R_s)$, respectively.

III. Machine Learning Based Transmit Antenna Selection Scheme

In this section, we detail our proposed machine learning based TAS schemes. Since the source is equipped with multiple antennas, we first model the problem in (1) as a multi-class classification problem. Through extracting features from the CSIs, we then apply two different machine learning approaches, namely, SVM and NB, to construct the classification model and predict the class label that the current channel belongs to.[1] We note that the belonged class represents an ideal antenna index to select that may optimize the secrecy performance of the current channel. We also note that the construction of the classification model needs a sufficiently large training data set and can be completed offline.

A. Preparation of Learning

In order to construct the classification model, we first obtain a training data set containing M varied training CSI examples, which is given by $[(\mathbf{H}^1, \mathbf{G}^1), (\mathbf{H}^2, \mathbf{G}^2), \ldots, (\mathbf{H}^M, \mathbf{G}^M)]$ for the full CSI case and $[\mathbf{H}^1, \mathbf{H}^2, \ldots, \mathbf{H}^M]$ for the partial CSI case. $\mathbf{H}^m$ and $\mathbf{G}^m$ are the m-th training main channel matrix and the m-th training eavesdropper's channel matrix, respectively.

Then, we perform three steps before the model construction and the class prediction [11]: 1) generate the feature vector for each training CSI example, 2) design the key performance indicator (KPI), 3) determine the class label of each training CSI example based on KPI.

1) Feature Vector Generation: Feature vectors are the inputs of the learning system. Unlike the CSI matrices containing complex-valued elements, feature vectors contain real-valued elements extracted from the training CSI matrices. Specifically, we first choose the absolute value of each element of $\mathbf{H}^m$ and $\mathbf{G}^m$ as the element of the feature vector. Then we normalize the feature vectors in order to avoid bias in the learning process. The detailed feature vector generation is summarized as follows:

Step 1: Generate a $1 \times N$ real vector $\mathbf{d}^m = [d_1^m, d_2^m, \cdots, d_N^m]$, where N equals to $N_s \times (N_r + N_e)$ for the full case and $N_s \times N_r$ for the partial case. For the full CSI case, $\mathbf{d}^m$ is expressed as

$$\mathbf{d}^m = \Big[|h_{1,1}^m|, \ldots, |h_{1,N_s}^m|, |h_{2,1}^m|, \ldots, |h_{N_r,N_s}^m|, \ldots, \\ |\mathrm{g}_{1,1}^m|, \ldots, |\mathrm{g}_{1,N_s}^m|, |\mathrm{g}_{2,1}^m|, |\mathrm{g}_{2,2}^m|, |\mathrm{g}_{N_e,N_s}^m|\Big], \tag{3}$$

and for the partial CSI case, $\mathbf{d}^m$ is expressed as

$$\mathbf{d}^m = \Big[|h_{1,1}^m|, \ldots, |h_{1,N_s}^m|, |h_{2,1}^m|, |h_{2,2}^m|, \ldots, |h_{N_r,N_s}^m|\Big], \tag{4}$$

where $h_{i,j}^m$ and $g_{i,j}^m$ denote the $(i,j)th$ element of $\mathbf{H}^m$ and $\mathbf{G}^m$, respectively.

Step 2: Repeat **Step 1** for all M training CSI examples and generate M feature vectors, i.e., $\mathbf{d}^1$, $\mathbf{d}^2$, $\cdots$, $\mathbf{d}^M$.

Step 3: Normalize $\mathbf{d}^m$ and generate the normalized feature vector $\mathbf{t}^m \in \mathbb{R}^{1\times\mathrm{N}}$, for $\mathrm{m} \in \{1, 2, \ldots, M\}$. The n-th element of $\mathbf{t}^m$ can be expressed as

$$\mathrm{t}_n^m = \big(\mathrm{d}_n^m - \mathbb{E}\big[\mathbf{d}^m\big]\big) / \big(\max\big(\mathbf{d}^m\big) - \min\big(\mathbf{d}^m\big)\big), \tag{5}$$

where d_n^m is the n-th element of $\mathbf{d}^m$.

2) KPI Design: KPI is the metric to classify the training CSI examples. Aiming at maximizing the secrecy performance, we choose the achievable secrecy rate as the KPI for the full CSI case. As for the partial CSI case, we adopt the achievable rate of the main channel C_b as the KPI.

3) Classification of Training CSI Examples: In order to determine the class label of one training CSI example, we first calculate the KPI for each antenna. Then we choose the class label of this training CSI example as the index of the antenna that achieves the maximum KPI.

B. SVM-Based Scheme

In this subsection, we present the proposed SVM-based scheme. We note that there are two methods to build SVM classifiers, i.e., one-against-one and one-against-the rest (OVR). In this letter, we focus on using OVR to solve the problem in (1). As such, we construct the model of OVR by

[1] In addition to TAS schemes, we note that our proposed SVM-based approach and NB-based approach are also suitable for other classification-based MIMO schemes, such as beamforming vector selection schemes [13].

solving the following alternative logistic regression problem

$$\boldsymbol{w}_l = \underset{\boldsymbol{w}_l}{\operatorname{argmin}}\, C \sum_{m=1}^{\mathrm{M}} \Big[b_l[m] \mathrm{cost}_1\Big(\boldsymbol{w}_l^T \boldsymbol{f}\big(\mathbf{t}^m\big)\Big) + (1 - b_l[m]) \\ \times \mathrm{cost}_0\Big(\boldsymbol{w}_l^T \boldsymbol{f}\big(\mathbf{t}^m\big)\Big)\Big] + ||\boldsymbol{w}_l||_2^2/2, \quad (6)$$

where $\boldsymbol{w}_l \in \mathcal{R}^{M \times 1}, l \in \{1, 2, \ldots, N_s\}$ denotes the learning parameter, C is a nonnegative scalar, which represents a trade-off between bias and overfitting, $b_l[m] = 1$ if the m-th training CSI example is labeled as l, otherwise $b_l[m] = 0$, $\boldsymbol{f}(\mathbf{t}^m) \in \mathbb{R}^{1 \times N}$ denotes the Gaussian radial-based kernel function, the q-th element of which is expressed as $f_q(\mathbf{t}^m) = \exp(-\|\mathbf{t}^m - \mathbf{t}^q\|/(2\sigma^2))$ with σ^2 denoting the variance of $\boldsymbol{f}(\mathbf{t}^m)$, and $\mathrm{cost}_k(z) = \max((-1)^k z + 1, 0)$ is the cost function. In (6), both C and σ are design parameters.

The SVM-based learning model is ready for TAS when all parameters $\boldsymbol{w}_l$ are obtained. For a new CSI example, we generate a normalized feature vector $\mathbf{t} \in \mathbb{R}^{1 \times N}$ using the feature vector generation steps described in Section III-A. the class label of current channel can be determined by using $\mathbf{t}$ to replace $\mathbf{t}^{\mathrm{m}}$ in (6). Then the class label l^* of current channel, i.e., the antenna that should be selected, is the one that achieves the largest $\boldsymbol{w}_l^T \boldsymbol{f}(\mathbf{t})$ among all classes.

C. NB-Based Scheme

In this subsection, we present the proposed NB-based scheme. Different from the SVM-based scheme, which is achieved by maximizing the margins between different classes, our proposed NB-based scheme solves the problem in (1) by utilizing the conditional probability. Specifically, we construct the NB classification model by calculating the probability distribution for each element of the normalized feature vector for all label classes.

We first express the posterior probability that a typical CSI example with normalized feature vector $\mathbf{t}$ belongs to class l as

$$\Pr(c = l|\mathbf{t}) = \frac{\Pr(\mathbf{t}|c = l)\Pr(c = l)}{\Pr(\mathbf{t})} \\ = \frac{\prod_{n=1}^{N} \Pr(\mathrm{t}_n|c = l)\Pr(c = l)}{\Pr(\mathbf{t})}, \quad (7)$$

where $\Pr(\mathbf{t}|c = l)$ is the probability of the occurrence of feature vector $\mathbf{t}$ given the class label l, $\Pr(c = l)$ is the prior probability of the class label l, $\Pr(\mathbf{t})$ is the probability of the occurrence of the feature vector $\mathbf{t}$, and t_n is the n-the element of $\mathbf{t}$. Based on (7), we find that selecting the optimal antenna is equivalent to select the class label that achieves the maximal posterior probability given the feature vector $\mathbf{t}$.

Due to the fact that $\Pr(\mathbf{t})$ is independent of the class label and the prior probability $\Pr(c = l)$ is a constant for all classes, we note that $\Pr(\mathbf{t})$ and $\Pr(c = l)$ have no influence on the selection. As such, the problem of selecting the optimal antenna can be reduced to the problem of selecting the class label that achieves the maximal probability of the occurrence of feature vector $\mathbf{t}$. Mathematically, it can be formulated as

$$l^* = \underset{l \in \{1, 2 \ldots, N_s\}}{\operatorname{argmax}} \prod_{n=1}^{N} \Pr(\mathrm{t}_n|c = l). \quad (8)$$

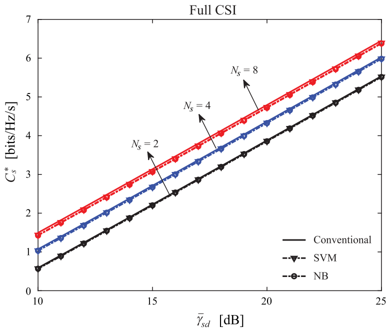

Fig. 1. C_s^* versus $\overline{\gamma}_{sd}$ for different transmission schemes and values of N_s with $\overline{\gamma}_{se} = 10$ dB.

The construction of the NB learning model can be summarized as follows:

1) Find all the feature-label pairs $\{\mathbf{t}^m, \mathrm{c}^m\}$ that satisfies $\mathrm{c}^m = l$ and construct a new set $\mathbb{L} = \{\mathbf{t}^m|\mathrm{c}^m = l, m \in \{1, 2, \ldots, M\}\}$;
2) Use all t_n^m of $\mathbf{t}^{\mathrm{m}} \in \mathbb{L}$ to calculate the probability distribution of $\Pr(\mathrm{t}_n|c = l)$ for $n \in \{1, 2, \ldots N\}$;
3) Repeat 1) and 2) for all $l \in \{1, 2, \ldots, N_s\}$.

We note that the NB-based learning model is ready for TAS after the probability set, given by $\mathbb{P} = \{\Pr(\mathrm{t}_n|c = l)|n \in \{1, 2, \ldots N\}, l \in \{1, 2, \ldots N_s\}\}$, is obtained.

IV. Numerical Results

In this section, we present numerical results to validate our proposed machine learning based TAS schemes. Specifically, we first compare the achievable secrecy rates of the proposed schemes with that of the conventional TAS scheme for the full CSI case. We then compare the SOPs of the proposed scheme with that of the conventional scheme for the partial CSI case. In addition, we quantify the performances and feedback overhead of the proposed schemes. Throughout this section, we generate 10^4 training CSI examples to train the SVM-based learning model and the NB-based learning model.

In Fig. 1, we plot the maximum achievable secrecy rate, C_s^*, versus the average signal-to-noise ratio (SNR) in the main channel, $\overline{\gamma}_{sd}$, for different values of N_s and the full CSI case. In this figure, we consider that the average SNR in the eavesdropper's channel, $\overline{\gamma}_{se}$, is 10 dB. We consider three different transmission schemes, namely, the conventional TAS scheme, the SVM-based scheme, and the NB-based scheme. We first see that C_s^* increases as $\overline{\gamma}_{sd}$ increases and N_s increases. We also see that our proposed schemes (i.e., the SVM-based scheme and the NB-based scheme) achieve almost the same secrecy rates as that of the conventional scheme for all values of N_s, thus validating the effectiveness of our proposed schemes.

In Fig. 2, we plot the minimum SOP, $P_{so}^*(R_s)$, versus $\overline{\gamma}_{sd}$ for different values of N_s and the partial CSI case. In this figure, we consider that $\overline{\gamma}_{se} = 10$ dB and $R_s = 2$ bits/Hz/s. As expected, we observe similar trends as observed in Fig. 1. Specifically, we see that P_{so}^* decreases as $\overline{\gamma}_{sd}$ increases and N_s increases. We also see that $P_{so}^*(R_s)$ achieved by our proposed schemes is almost the same as that of the conventional TAS scheme for all values of N_s.

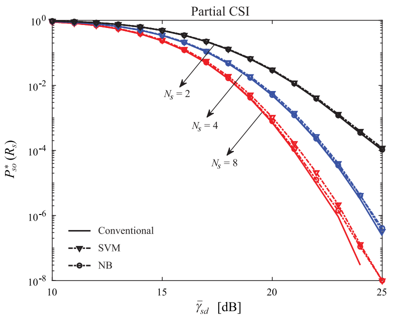

Fig. 2. $P^*_{so}(R_s)$ versus $\overline{\gamma}_{sd}$ for different transmission schemes and values of N_s with $\overline{\gamma}_{se} = 10$ dB and $R_s = 2$ bits/Hz/s.

TABLE I
NMSEs of Different TAS Schemes

Scheme	Full CSI		Partial CSI	
	SVM	NB	SVM	NB
$N_s = 2$	−45.14 dB	−43.61 dB	−47.74 dB	−62.69 dB
$N_s = 4$	−39.65 dB	−40.24 dB	−39.48 dB	−57.69 dB
$N_s = 8$	−37.18 dB	−35.48 dB	−34.58 dB	−48.87 dB

In addition, we see that our SVM-based scheme outperforms our NB-based scheme for the full CSI case, while performing worse than our NB-based scheme for the partial CSI case. This is because SVM is a deterministic classification method based on the feature vector. As such, its performance will be degraded when only partial CSI is available to generate the feature vector. On the other hand, NB is a probabilistic method based on the corresponding probability, which will not be impacted when only partial CSI is available. Although not shown here, we note that our proposed schemes achieve almost the same performance as the conventional TAS scheme even when the number of selected antennas $N_c > 1$, indicating that our proposed scheme can also be applied to scenarios where multiple antennas are selected at the source.

The performances of machine learning based TAS schemes are quantified in Table I. Specifically, we adopt the normalized mean square error (NMSE) to assess the performances of different TAS schemes, given by

$$\mathrm{NMSE} = 10\log_{10}\frac{||\mathbf{s}_{\mathrm{con}} - \mathbf{s}_{\mathrm{ML}}||^2}{||\mathbf{s}_{\mathrm{con}}||^2}, \tag{9}$$

where $\mathbf{s}_{\mathrm{con}}$ denotes the secrecy rates (or secrecy outage probabilities) for all SNRs of the conventional scheme and $\mathbf{s}_{\mathrm{ML}}$ denotes the secrecy rates (or secrecy outage probabilities) of the proposed machine learning based schemes. Note that the value of NMSE represents the performance gap between the proposed schemes and the conventional scheme. The smaller the value of NMSE is, the smaller the performance gap is. We can see that the NMSE of our proposed scheme for all values of N_s are below −30 dB, which again verifies the effectiveness of our proposed schemes.

Finally, we examine the selection complexities and the feedback overheads of different TAS schemes in Table II. In this table, $|\mathcal{L}|$ denotes the number of selected antenna combinations, e.g., $|\mathcal{L}| = \binom{N_s}{N_c} = \frac{N_s!}{(N_s-N_c)!N_c!}$ when N_c antennas are selected at the source, $N = N_s(N_r + N_e)$ for the full CSI case, and $N = N_sN_r$ for the partial CSI case. We can see that, due to the combinatorial search across all possible antenna combinations, the selection complexities of the conventional scheme and our NB-based scheme are higher than that of our SVM-based scheme when $|\mathcal{L}|$ is relatively large. We also see that the feedback overhead of our proposed schemes is half of that of the conventional scheme, showing that our schemes can be applied to the scenarios where the feedback is limited.

TABLE II
Complexities and Overheads of Different TAS Schemes

Scheme	SVM	NB	Conventional
Selection complexity	$\mathcal{O}(N^2)$	$\mathcal{O}(\lvert\mathcal{L}\rvert N+$ $\lvert\mathcal{L}\rvert\log(\lvert\mathcal{L}\rvert))$	$\mathcal{O}(N+$ $\lvert\mathcal{L}\rvert\log(\lvert\mathcal{L}\rvert))$
Feedback overhead	N real values		N complex values

V. Conclusion

In this letter, we characterize the problem of TAS in MIMO wiretap channels as a multi-class classification problem, which enables us to utilize machine learning based schemes (i.e., SVM and NB) to construct the classification model and select the optimal antenna that maximizes the secrecy performance. Both the case where the CSI of the eavesdropper is known at the source and the case where the CSI of the eavesdropper is not known at the source are considered. We show how our proposed machine learning based schemes can achieve almost the same secrecy performance as the conventional TAS scheme, while only requiring the half amount of feedback overhead of the conventional TAS scheme.

References

[1] N. Yang *et al.*, "Safeguarding 5G wireless communication networks using physical layer security," *IEEE Commun. Mag.*, vol. 53, no. 4, pp. 20–27, Apr. 2015.

[2] A. D. Wyner, "The wire-tap channel," *Bell Syst. Tech. J.*, vol. 54, no. 8, pp. 1355–1367, Oct. 1975.

[3] A. Khisti and G. W. Wornell, "Secure transmission with multiple antennas—Part II: The MIMOME wiretap channel," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5515–5532, Nov. 2010.

[4] C. Liu, N. Yang, J. Yuan, and R. Malaney, "Location-based secure transmission for wiretap channels," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 7, pp. 1458–1470, Jul. 2015.

[5] N. Yang, P. L. Yeoh, M. Elkashlan, R. Schober, and I. B. Collings, "Transmit antenna selection for security enhancement in MIMO wiretap channels," *IEEE Trans. Commun.*, vol. 61, no. 1, pp. 144–154, Jan. 2013.

[6] G. Pan *et al.*, "On secrecy performance of MISO SWIPT systems with TAS and imperfect CSI," *IEEE Trans. Commun.*, vol. 64, no. 9, pp. 3831–3843, Sep. 2016.

[7] H. Zhao, Y. Tan, G. Pan, Y. Chen, and N. Yang, "Secrecy outage on transmit antenna selection/maximal ratio combining in MIMO cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 12, pp. 10236–10242, Dec. 2016.

[8] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of Machine Learning*. Cambridge, MA, USA: MIT Press, 2012.

[9] S. R. Gunn, "Support vector machines for classification and regression," Image Speech Intell. Syst. Res. Group, Univ. Southampton, Southampton, U.K., Rep. 1997.

[10] P. Harrington, *Machine Learning in Action.* Greenwich, CT, USA: Manning, 2012.

[11] J. Joung, "Machine learning-based antenna selection in wireless communications," *IEEE Commun. Lett.*, vol. 20, no. 11, pp. 2241–2244, Nov. 2016.

[12] P. A. Dighe, R. K. Mallik, and S. S. Jamuar, "Analysis of transmit-receive diversity in Rayleigh fading," *IEEE Trans. Commun.*, vol. 51, no. 4, pp. 674–703, Apr. 2003.

[13] X. Chen and Y. Zhang, "Mode selection in MU-MIMO downlink networks: A physical-layer security perspective," *IEEE Syst. J.*, vol. 11, no. 2, pp. 1128–1136, Jun. 2017.