Alexandria Engineering Journal (2022) 61, 11589–11600

HOSTED BY

Alexandria University

Alexandria Engineering Journal

www.elsevier.com/locate/aej
www.sciencedirect.com

Enhanced feature selection method based on regularization and kernel trick for 5G applications and beyond

Amira Zaki [a], Ahmed Métwalli [a], Moustafa H. Aly [a,b,*], Waleed K. Badawi [a]

[a] Arab Academy for Science, Technology and Maritime Transport, Alexandria 1029, Egypt
[b] OSA Member

Received 18 February 2022; revised 18 May 2022; accepted 18 May 2022

KEYWORDS

Wireless communication;
5G;
Machine learning;
QuaDRiGa platform;
Feature selection;
Regularization

Abstract Prediction of wireless channel scenarios is fundamental for modern wireless communication systems with diverse propagation conditions. Moreover, the type of data extracted from a wireless communication channel impulse response (CIR) is complex. In recent research, machine learning (ML) techniques have proven their success in classification problems of wireless communication scenarios and provide reasonable results. In this paper, a new enhanced feature selection method is proposed to improve the training model and classification performance of the conventional model. This improvement is achieved based on the concept of regularization in which the selection of the best features is considered before training the model under any propagation environment. The adoption of regularization leads to a high Total Explained Variance (TEV) during the process of kernel Principal Component Analysis (k-PCA). As a consequence, two principal features are used instead of three. The proposed model has high generalization ability since it reduces the features dimensionality (computational complexity) and, generally, enhances the ML classification performance. Experimental simulation is executed to compare the proposed model and the conventional one in terms of accuracy, precision and recall. The accuracy is increased from 97% to 99%, from 96% to 99%, from 89% to 97% and from 90% to 98% for k-nearest neighbor (k-NN), support vector machine (SVM), k-Means and Gaussian mixture model (GMM), respectively.

1. Introduction

Today, one of the main key components for a modern data-driven wireless communication system is Artificial Intelligence (AI) including Machine Learning (ML) [1]. The adoption of these techniques is essential at physical layers, middle layers and end user layers [2]. However, this rising trend is pushed by the massive amount of data transceiver devices such as

* Corresponding author at: Department of Electronics and Communications, College of Engineering and Technology, Arab Academy for Science, Technology and Maritime Transport, Alexandria, Egypt.
E-mail address: mosaly@aast.edu (M.H. Aly).
Peer review under responsibility of Faculty of Engineering, Alexandria University.

https://doi.org/10.1016/j.aej.2022.05.024
1110-0168

smartphones, Internet of Things (IoT) sensors, laptops and tablets. These applications achieved a significant social industry, scientific and medical technological progress. Furthermore, by the growth of the IoT, multiple connected sensors implanted in various settings place a strong focus on the identification of scenarios. In other words, IoT techniques are used alongside AI in order to provide suitable wireless communication networks design. In addition to wireless systems deployment, AI and IoT improve data transmission quality. As a result, properly determining wireless channel situations to meet the devoted need of the wireless environment becomes a concern.

A channel scenario is the unique transmission and reception environment for a wireless network such as the suburban, urban, rural macro-cells, satellites, indoor hotspots, etc. [3]. Every reception setting or scenario is mainly divided into two types: Line-of-Sight (LoS) and Non-Line-of-Sight (NLoS) multipath situations. Even for two scenarios in much the same propagation environment, the channel conditions parameters of one scenario significantly differ from the others (e.g., the LoS and NLoS scenarios of the rural macro-cell (RMa) network or urban macro-cell (UMa) network). The recent growth of teledensity in urban areas, pushed by mobile technology, means that the digital gap between rural and urban areas has been widened.

As a natural outcome, various empirical statistical radio propagation models have been developed for specific transmit circumstances, such as the traditional Hata and Okumura models for urban scenarios [4]. These frameworks are built on the hypothesis that the transmission situations are recognized. As a corollary, determining the propagation environment is critical before implementing the channel model, especially in 5G [4]. Unfortunately, real-world signal propagation environments frequently experience a variety of circumstances or different scenarios. This can be witnessed by the users who transport with high speed as for example, the high-speed railway. As the user passes through multiple different channels and scenarios, such as stations, mountains, deserts, and other obstacles, he poses significant difficulties to present communication systems [5]. Currently, most of scenario determination is done manually by observing the surrounding area. This may definitely result in some false classifications or blunders. As a reason, reducing the complexity and classifying accurately for the scenario identification process are critical for improving the reliability of the communication system. Also, one of the current demands is the low latency. ML techniques have been widely used in a variety of scientific disciplines, including speech recognition and wireless technology. This is because of the implementation of deep neural networks and their remarkable success in the object recognition domain [6]. Besides, at the application and infrastructure levels of communication hierarchy, the AI-based solutions address the following issues such as latency, power control, channel capacity, privacy and security [7].

Once, a ML model is trained on the available data, it can successfully, make decisions on unknown data and execute tasks using arithmetic calculations. This would enable ML modeling for mobility, availability, accessibility, inter-process communication management based on 6G data, optimize and automate network performance management. This is important to keep the current key performance indicators within preset limits. Moreover, the administration of 6G mobile networks with smart adaptive cells is also possible using ML.

Beam management, power-saving, fault management, maintenance, operation, power control, network setup, QoS prediction, throughput, and coverage performance will all benefit from this [2]. Deep learning methods also proved its high accuracy in data-driven wireless communication problems [8]. Deep learning approaches are more complex. For example, who focused on the angular information that is collected from CIR to differentiate the NLoS and LoS scenarios in urban areas [9]. Also, where the author exhibited excellent accurate classification performance using a deep convolutional neural network for fingerprint feature extraction and classification of wireless channels built on software-defined radio [10]. ElasticNet regularization also improves the feature selection process even more [11,12]. The most important objective in these complicated applications is to evaluate the input data with fair performance metrics such as accuracy [13], precision and recall in addition to minimal computing cost and time. The spectrum sensing researches that include ML approaches are proving the reliability of ML [14]. In the conventional model [15], authors proved that supervised learning algorithms such as, SVM and k-NN, are two suitable classification strategies to deal with scenario identification alongside two unsupervised learning algorithms such as k-Means and GMM. However, the conventional model uses constant seven features and some of these features may not be suitable for model training in a specific environment [15]. So, the feature selection process of the conventional model [15] needs to be generalized for any propagation environment since choosing input ML data for classification is crucial and unwanted features may result in bad classification accuracy.

The contributions of this paper are:

1. Enhanced feature selection process contribution that suits any propagation environment which determines the input variables before training the ML model. This selection is a cross-validation based regularization technique and does both, regularization and dimension reduction k-PCA.
2. The number of principal orthogonal components processed by k-PCA is determined using TEV. In this work, high TEV is achieved using only two principal components instead of three.
3. A new contributed dataset containing the data describing each scenario. The data are extracted from CIR including small-scale-fading (SSF) and large-scale-fading (LSF) parameters. This environment contains rich wireless scenarios which include first bounce scatterers (FBS) and last bounce scatterers (LBS), which are simulated under the standard of TR mmw 3GPP 38.901 [3].

The rest of this paper is organized as follows. Section 2 provides wireless environment and network configurations. In addition, it includes the parameters extraction that takes place in order to create a structured dataset of each feature. Then, Section 3 shows the proposed data preprocessing workflow, where the feature selection process includes regularization. Consequently, in Section 4, the ML algorithms are applied and evaluated. Hence, a comparison between the proposed model and the conventional model, in terms of ML evaluation metrics, is introduced. Finally, the conclusion is adopted in Section 5.

2. Propagation environment

Generally, a propagation environment is referred as a propagation model that consists of three components: a path-loss-dependent component, shadow fading, and fast fading. Creating a rich, detailed and realistic propagation environment is an essential step toward obtaining a dataset that contains practical CIR snapshots and creating a structured dataset. Propagation environment models are introduced and studied in this section. These environments are generated by the Quasi Deterministic Radio Channel Generator (QuaDRiGa) platform. This platform simulates wireless communication environments and scenarios. Also, The geometry-based stochastic channel modelling technique is used in the QuaDRriGa channel model. The antenna effect can be simulated using a geometry-based technique, while propagation properties are specified using statistical models [3]. This section also discusses the data collection process using the Spatial Consistency model of the QuaDRiGa platform.

The QuaDRiGa platform Spatial Consistency model places FBS and LBS scatterers randomly. They scatter the signal propagated between BS and MT along a linear path of 50 m. These scatterers force the LoS power to be set at 33% ± 10% of total power with a 10% standard deviation which differentiates LoS and NLoS scenarios, as LoS scenarios have a higher power. A snapshot is taken each 1 m, then each 10 snapshots are averaged and denoised into a single snapshot, so that a structured data table containing 500 CIR data points for each scenario is obtained. These scenarios are: 3GPP 38.901 RMa LoS, 3GPP 38.901 RMa NLoS, 3GPP 38.901 UMa LoS and 3GPP 38.901 UMa NLoS [3]. These are the target output of 3GPP 38.901 scenarios as shown in Table 1. UMa is characterized in urban areas which refers to city or town. On the other hand, RMa is specified in rural areas such as country with less population density and less scatterers.

The multiple-input and multiple-output (MIMO) system is adopted [16]. The BS antenna is formed of 31 elements to provide the angular information. Hence, by applying the space-alternating generalized expectation-maximization (SAGE) algorithm, the angular spread and the mean of cluster angles can be extracted [17]. Then, the parameters of a CIR snapshot which describe the LSF and the SSF are the delay spread (DS), path loss (PL), k-factor (KF), elevation spread angle of arrival (esA), elevation spread angle of departure (esD), azimuth spread angle of arrival (asA) and azimuth spread angle of departure (asD) which are combined into a structured data table and are ready for data preprocessing.

Table 1 Emulator communication environment configuration.

Parameter	Value
3GPP 38.901 Scenarios	RMa LoS - UMa LoS RMa NLoS - UMa NLoS
Bandwidth [MHz]	100
Track type	Linear
BS height [m]	25
MT height [m]	1.6
Carrier Frequency [GHz]	2.6
Sample density [cm]	10
No. snapshots for each scenario	500–500–500–500
No. cluster for each scenario	25–10–34–58
No. of BS antenna	31

PL is the degradation of power across distances. So, PL is the relation between the distance d and the actual PL, indicated as P, in a given situation has been shown via the empirical work [18].

$$P_{dB} = PL(d_0) + 10\gamma log\log_{10}\left(\frac{d}{d_0}\right) + X_{\sigma[dB]} \tag{1}$$

The PL exponent is denoted as γ and d_0 is the reference distance while $X_{\sigma[dB]}$ expresses the standard normal distribution.

KF is a typical SSF parameter which represents the ratio between the power of a LoS component over all other multipath components. For each CIR snapshot, KF (K_{dB}) can be expressed as [19].

$$K_{dB} = 10log\left\{\frac{\left(|h(\tau_{m_0})|_{max}\right)^2}{\sum_{\tau_m \neq \tau_{m_0}}\left(|h(t)|\right)^2}\right\} \tag{2}$$

where τ_m represents the current duration of the m_th multipath component delay and the highest amplitude appears in τ_{m_0} index. Also, m = 1, 2, 3 ... M. $h(t)$ represents the CIR with time domain. The value of K_{dB} is generally larger in LoS scenarios than NLoS scenarios.

RMS DS is another typical SSF parameter which represents the channel dispersion of each snapshot from the perspective of time delay. RMS DS (σ_τ) can be represented as [20].

$$\sigma_\tau = \sqrt{\frac{\Sigma_{m=1}^{M}\left(\tau_m - \bar{\tau}\right)|h(\tau_m)|^2}{\Sigma_{m=1}^{M}|h(\tau_m)|^2}} \tag{3}$$

where M is the total number of multipath components and $\bar{\tau}$ is the mean excess delay that can be approximated as

$$\bar{\tau} = \frac{\Sigma_{m=1}^{M}\tau_m|h(\tau_m)|^2}{\Sigma_{m=1}^{M}|h(\tau_m)|^2} \tag{4}$$

The communication channel capacity depends on RMS DS, where a channel with several rich scatterers will result in a greater RMS DS. In other words, the NLoS scenarios have a greater RMS DS than the LoS scenarios.

AS is similar to RMS DS, but, it represents the channel dispersion of each snapshot from the perspective of the angular domain [21]. AS is also denoted as σ_θ and it is calculated as

$$\sigma_\theta = \sqrt{\frac{\Sigma_{m=1}^{M}\theta_{m,\mu}^2.|h(\tau_m)|^2}{\Sigma_{m=1}^{M}|h(\tau_m)|^2}} \tag{5}$$

$$\theta_{m,\mu} = mod\left(\left(\theta_m - \bar{\theta} + \pi, 2\pi\right)\right) - \pi \tag{6}$$

$$\bar{\theta} = \frac{\Sigma_{m=1}^{M}\theta_{m,\mu}|h(\tau_m)|^2}{\Sigma_{m=1}^{M}|h(\tau_m)|^2} \tag{7}$$

The expression of angle θ is valid for azimuth angle of arrival (AOA), azimuth angle of departure (AOD), elevation angle of arrival (EOA) and elevation angle of departure (EOD). The NLoS scenarios contain more clusters than the LoS scenarios. As a consequence, the value of AS is larger in NLoS scenarios.

These characteristic channel variables are selected as the features based on the aforementioned study. Hence, Table 2 shows the average and standard deviation for each scenario, where each data point can be represented as $\Omega_i = \{K_i, P_i, \sigma_{\tau,i}, \sigma_{AOA,i}, \sigma_{AOD,i}, \sigma_{EOA,i}, \sigma_{EOD,i}, t_i\}$, where i represents the counter of rows and t_i is the target variable. In this case, it contains categorical values which are the outcomes of each snapshot such as {UMa LoS, UMa LNoS, RMa LoS, RMa NLoS} for the wireless environment mentioned. This gives the intuition about the classification problem of which ML algorithms will solve.

3. Proposed preprocessing workflow

In this section, the input features used for ML classifiers are produced and preprocessed using both regularization and dimension reduction. Therefore, the k-PCA with Radial Basis Function (RBF) will be adopted to reduce data dimensions [22], after using z-score normalization as shown in Fig. 1. Fig. 1 shows a flowchart for data workflow, where the 7 input features contained in each data point Ω_i are being normalized and regularized in order to get the selected features. If the selected features dimensions are equal or less than 3, the data dimension reduction techniques are not needed, and the classification ML process should be done directly and data can be visualized in 3-D. Else, the k-PCA is used to reduce the dimensions to 3 or less.

3.1. Normalization

The ML models may suffer poor performance and high complexity due to the data containing outliers and high variance [23]. Hence, data normalization should be applied on each data point Ω_i^j by using z-score which measures the distance to its mean value divided by a standard deviation and can be expressed as [23].

$$\Omega_i^j = \frac{\Omega_i^j - \bar{\Omega^j}}{\sigma_{\Omega^j}} \tag{8}$$

where $\bar{\Omega^j}$ and σ_{Ω^j} are the average and standard deviation of each feature j.

3.2. Enhanced feature selection

Selecting the input variables significantly affects the model prediction accuracy, precision and recall. Hence, it is the process of decreasing the number of input features before deploying the model in order to reduce the computational cost and improve the performance. Regularization is a type of regression analysis which plays a vital role in statistical modeling and in turn for performing ML tasks.

Table 3 Data preprocessing complexity comparison in terms of number of features.

Model	Regularization	Dimension Reduction	TEV
Proposed	7 to 4	4 to 2	72%
Conventional [15]	No	7 to 3	71%

Due to regularization, preprocessing workflow may not require the usage of a dimension reduction technique if the number of selected features or dimensions $\mathfrak{R}$ is less than or equal to 3. However, reducing $\mathfrak{R}$ to 3 is essential to visualize data and to reduce computational complexity. So, the data visualization may occur directly alongside applying ML learning algorithms. The conventional work depends on direct dimension reduction from 7 to 3 regarding the importance of each feature [15]. This requires more computational complexity during the preprocessing phase and reduces the accuracy of ML algorithms.

In order to increase the performance of machine learning algorithms such as k-NN, SVM, k-Means and GMM, the process of data preprocessing must contain a regression analysis regularization technique based on cross-validation. This is because unwanted data are translated into noisy data that affect the model later on. So, Least Absolute Shrinkage and Selection Operator (LASSO), Ridge and ElasticNet are all typical cross-validation based regularization techniques. Both LASSO and Ridge regression are loss functions, where LASSO determines which features should be dropped or avoided in order to train the model by adding a penalty coefficient to the output [24]. LASSO regression is L1 regularization type while Ridge is L2 regularization type, where L1 stands for Least Absolute Error and L2 stands for Least Square Errors. The LASSO regression $\widehat{B}^{lasso}$ and ridge regression $\widehat{B}^{ridge}$ can be expressed respectively, as

$$\widehat{B}^{lasso} = argmin\left\{\frac{1}{2}\sum_{i=1}^{N}\left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j\right) + \lambda\sum_{j=1}^{p}|\beta_j|\right\} \quad \text{subject to } \sum_{j=1}^{p}|\beta_j| \le t \tag{9}$$

$$\widehat{B}^{ridge} = argmin\left\{\frac{1}{2}\sum_{i=1}^{N}\left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j\right) + \lambda\sum_{j=1}^{p}\beta_j^2\right\} \quad \text{subject to } \sum_{j=1}^{p}|\beta_j| \le t \tag{10}$$

Table 2 Mean and standard deviation for each feature in each scenario.

Feature	RMa LoS	RMa NLoS	Uma LoS	Uma NLoS
DS (dB)	−66 ± 1.3	−72.1 ± 0.5	68.2 ± 4.2	−62 ± 1.9
KF (dB)	−2.6 ± 0.2	−3 ± 0.05	−2.7 ± 0.5	−3 ± 0.05
PL (dB)	0.77 ± 1.8	0.84 ± 3.5	0.75 ± 1.9	0.92 ± 3.4
asA (deg)	54.5 ± 5.6	21 ± 3.2	79.7 ± 35	87.4 ± 6
asD (deg)	13.8 ± 3.3	7 ± 2.3	13.6 ± 4	11.6 ± 0.8
esA (deg)	3 ± 1.8	2.8 ± 0.2	8.9 ± 2.4	26 ± 3.91
esD (deg)	3.5 ± 1.5	1.3 ± 0.05	2.7 ± 1.7	1 ± 0.5

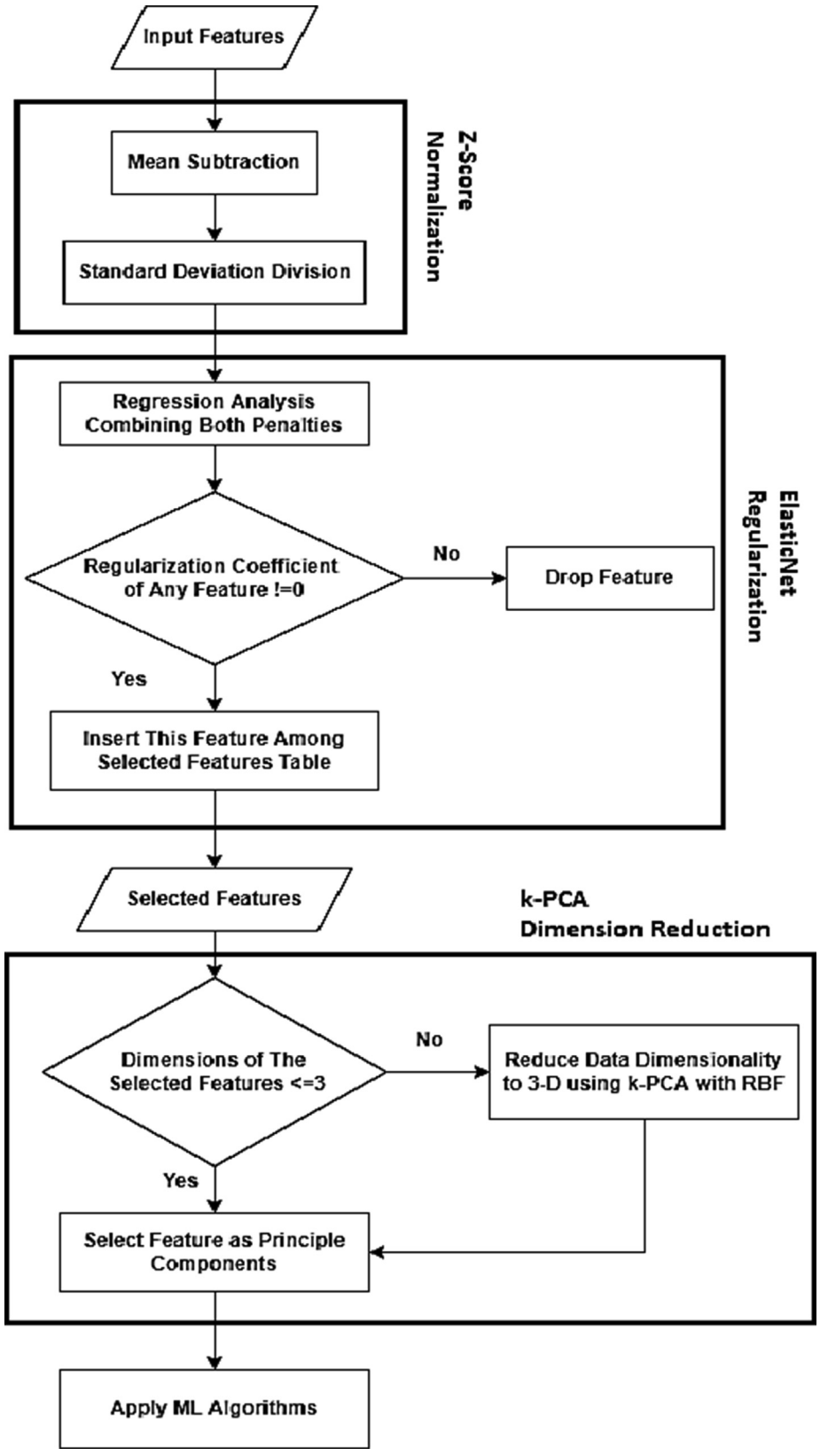


Fig. 1 The proposed approach flow chart.

where λ is the amount of shrinkage. It represents the regularization penalty and must be equal or greater than 0. β_0 is the constant coefficient, $\beta = (\beta_0, \beta_1, \beta_2, \cdots, \beta_N)$ is the coefficient vector and t is a prespecified free parameter that determines the degree of regularization.

The Ridge regression reduces the complexity of the model but cannot reduce the number of input features since it never sets a coefficient to zero but only minimizes it as shown in Fig. 2 (a). Hence, this model is not suitable for feature reduction. On the other hand, in some cases, if there are multiple highly collinear variables, then the LASSO regression randomly chooses one of them. This is the reason why LASSO is not good for data interpretation, but, for this case with this specific environment, it is suitable because it produces results similar to ElasticNet as shown in Fig. 2 (b) and (c). Hence, ElasticNet solves these limitations by combining the regularization of both LASSO and Ridge. ElasticNet regression can be expressed as

$$\widehat{B}^{elastic} = argmin\left\{ \sum_i \frac{1}{2} \left| y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \right|_2^2 + \lambda\left(\alpha_1 \sum_{j=1}^{p} \beta_j + \alpha_2 \sum_{j=1}^{p} \beta_j^{\,2} \right) \right\}, \quad (11).$$

where both α_1, α_2 control the ratio of penalty and $\alpha_1 + \alpha_2 = 1$.

The features to be eliminated are those which have coefficients equal to 0. LASSO and ElasticNet eliminated 3 variables: {esD, asD, DS} and kept the 4 variables: {esA, KF, PL, asA} for this propagation environment dataset. However, it is generalized and so, it can be used for any further propagation environment dataset as it will choose other important features that suite the regression analysis.

Both LASSO and ElasticNet can be used in the enhanced feature selection process before employing and other data preprocessing such as dimension reduction, unlike Ridge regression which has not any zero coefficient. So, it does not drop any feature which leads to more complexity and poor performance. Hence, for more stability and reliability, ElasticNet is a good candidate to be chosen.

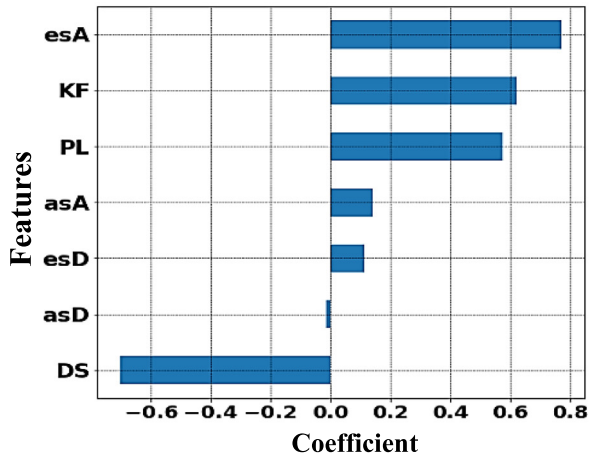
3.3. Dimension reduction

A typical preprocessing step is data projection or reduction since it decreases the processing time and complexity. Most of the models deploy PCA as a dimension reducer. This is because it turns a subset of dependent variables into a set of measurements of linearly independent features via an orthogonal transformation. However, the direct usage of PCA on a complex distributed dataset is not efficient. In this paper, the most recent PCA technique is k-PCA and also called kernel trick [25]. Any algorithm that can be performed successfully on the basis of the inner product can benefit from the kernel technique. The kernel type used in this work is RBF and can be calculated as

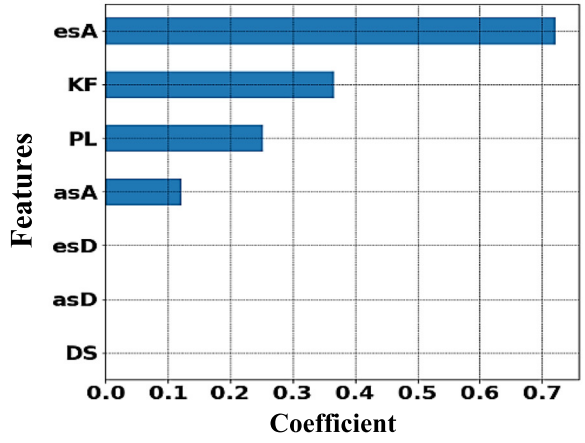
$$K(\Omega_a, \Omega_b) = exp\left(-\Gamma ||\Omega_a - \Omega_b||^2\right), \quad (12)$$

where Ω_a, Ω_b are two different points and Γ is a threshold hyper-parameter which is set to 0.5 [25] and [26].

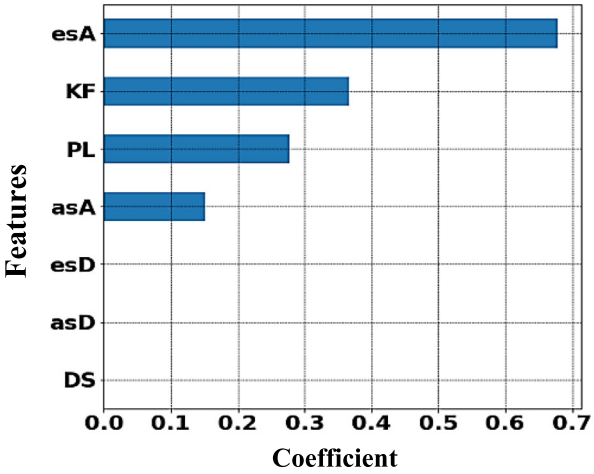
To evaluate the selected principal components provided by k-PCA, the total explained variance is used as an evaluation metric as it is the sum of all selected components. As previously mentioned, regularization reduces dimensions from 7 to 4, then there are 4 principal components in k-PCA as shown in Fig. 3. Hence, selecting principal features that have their sum of the explained variance is more than 50% or 60% of total explained variance is essential before training the ML model. In the proposed model, choosing the first 3 components



(a) Feature selection using Ridge regression.



(b) Feature selection using LASSO regression.



(c) Feature selection using ElasticNet regression.

Fig. 2 Regularization using Ridge, LASSO and ElasticNet regression.

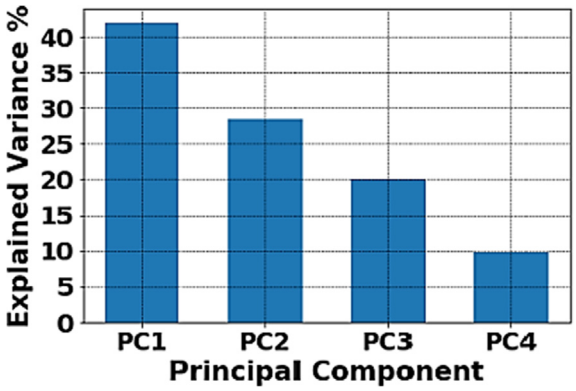


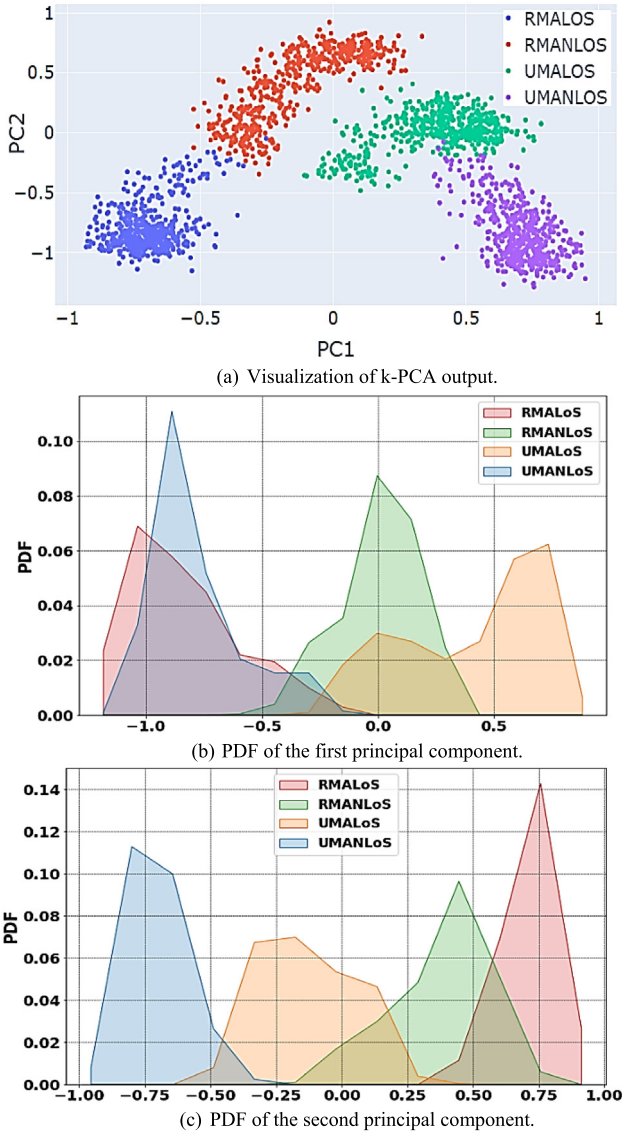
Fig. 3 Total explained variance for the proposed model.

will result in 90% of total explained variance. Moreover, using the first 2 principal components 1 and 2, will result in 70% of total explained variance. This is considerably enough for a reliable training dataset provided to ML models. As a consequence, the ML input predictors are reduced and the computational complexity and computational time are also reduced.

So, the dimensions ℜ are reduced from 7 to 4 during regularization and then from 4 to 2 during dimension reduction. This will enhance the classification performance during the ML process. The visualization of the data principal features is shown in Fig. 4 (a), where each scenario can be differentiated, visually, using only 2 principal components. Fig. 4 (b) shows the Probability Density Function (PDF) of each scenario in the first principal component and it is clear that it is hard to distinguish between each of them with this component. The first principal component contains an overlap of multiple classes. This overlap results in misclassification due to the confused data points such as the cases of RMA NLoS and UMA NLoS. So, Fig. 4 (c) presents the PDF of the second principal component, showing an obvious difference between the four classes. It shows another dimension of the information as for an example, the UMa NLoS can be differentiated easily from RMa LoS. Hence, using these 2 principal components is enough for classification tasks.

3.4. Statistical hypothesis testing

Statistical hypothesis testing is used to determine whether the result of a data set is statistically significant. Statistical significance is the likelihood that the difference in conversion rates between a given variation and the baseline is not due to random chance. This is indicated by the P-value which is the level of marginal significance within a statistical hypothesis test, representing the probability of the occurrence of a given event. The T-test is a typical type of inferential statistic used to determine if there is a significant difference between the means of the two groups, which may be related in certain features [27]. The two groups in this case are the population and the sample for both PC1 and PC2. The null hypothesis (H_0) is that the mean of the sample set is the same as the population set. On the other hand, the alternative hypothesis (H_A) is that both means are different. So the problem formulation can be expressed as



(a) Visualization of k-PCA output.

(b) PDF of the first principal component.

(c) PDF of the second principal component.

Fig. 4 k-PCA output principal components.

$$H_0: \; Mean\left(PC[i]_{Sample}\right) = Mean\left(PC[i]_{Population}\right), \tag{13}$$

$$H_A: \; Mean\left(PC[i]_{Sample}\right) \neq Mean\left(PC[i]_{Population}\right), \tag{14}$$

where $i = 1, 2$ which represents both PC1 and PC2. The T-test shows a resultant P-value of 0.96 and 0.97 for PC1 and PC2 respectively. A p-value higher than 0.05 is not statistically significant and indicates strong evidence for the null hypothesis. So the result is that the null hypothesis is failed to be rejected and there is no statistical significance. In other words, the sample distribution is the same as the population. Thus, the test shows the robustness of the dataset that will be used as predictors on the next section.

4. ML and evaluation

The type of problem that ML will solve is a classification based problem, where the output variable is called Target, and it has 4 possible categorical values. These categories are: RMa LoS,

RMa NLoS, UMa LoS, and UMa NLoS. When these labels are used during the ML training, then, this process is called supervised learning. On the other hand, unsupervised learning approaches do not need labels as the main objective is to differentiate the groups by clusters. In addition, the input variables for both supervised and unsupervised models are the k-PCA output principal components 1 and 2 for the proposed model. On the other hand, the conventional work has 3 fixed principal components and this will be taken into consideration during comparison.

In this section, the focus is upon testing the performance of the proposed model and comparing it with the recent conventional model [15]. This comparison is performed using the same algorithms on the same dataset. The evaluation results are obtained easily using confusion matrices [28]. These confusion matrices can also provide the 3 key metrics of evaluation: accuracy, precision and recall.

4.1. Supervised learning techniques

Supervised learning, also known as supervised machine learning, is a subcategory of machine learning and artificial intelligence. It is defined by its use of labeled datasets to train algorithms that to classify data or predict outcomes accurately.

In supervised learning, data must be split mainly into training and testing data. As previously mentioned, the classification process of the proposed model has 4 equally probable scenarios, where each one contains 500 data points. Moreover, in supervised approaches, data are mainly split into training set and testing set.

Recent research is directed toward employing an efficient supervised algorithm with high performance. Hence, k-NN algorithm in a simple way is a basic classifier, where it does not require an awareness or background knowledge of statistics and distributions. It classifies a new inserted data point according to the majority of its nearest surrounding points [29]. However, number of neighbors, k, must be a positive odd integer. The parameter k determines the number of surrounding data points to be taken in calculation. This method is based on the Euclidean distances of the nearest neighbors with the inserted data point.

Also, SVM proved its robustness in both classification and regression since it is robust to the over-fitting problems that occur during the training. In addition to its flexibility with different kernels [30] as it handles training data sets containing outliers, non-normal distributed data, nonlinear correlations, noisy and complex data [31]. The main idea of SVM is to create a line or hyper-plane to separate each class of the data and can support multiple dimensions. The goal is to set different support vectors and find the optimal hyper-planes in such a way to minimize the error and increase maximal margins between each class [32].

4.2. Unsupervised learning techniques

Unsupervised learning is a branch of machine learning that is used to find underlying patterns in data and is often used in exploratory data analysis. Unsupervised learning does not use labeled data like supervised learning. Instead, it focuses on the data's features.

The main goal of unsupervised learning is to cluster the classes or groups without having labels. So it trains the model only with input features in order to discover the patterns of data and map each group of the data into a cluster.

A simple form of clustering can be represented in k-Means clustering technique as it is widely used in data mining. However, it is an old method, but, is still being used since it provides a suitable performance, convergence and simplicity of implementation [33].

Given a numerical input dataset $U = \{u_1, u_2, ..., u_n\}$, the k-Means attempts to cluster N data points into M sets $S = \{s_1, s_2, ..., s_m\}$, iteratively, to minimize the distance between the mean and the data points inside each cluster. An objective function is included for the clustering algorithms (K-Means and GMM), where the function is to minimize the distance between the centroid of the class and the surrounding data points. It can be considered as an optimization problem and expressed as [33].

$$argmin \sum_{m=1}^{M} \sum_{u_j \epsilon S_j} \|u_j - \mu_j\|, \tag{15}$$

where μ_j is the mean of a subcluster S_m.

k data points are selected iteratively in a random form so that k-Means converges at each iteration. The Elbow method [33] is used to determine the number of clusters as shown in Fig. 5, where the number of clusters is found at the inflection point equal to 4.

Moreover, to make use of the flexibility in distribution fitting, GMM uses probabilities to estimate the likelihood of a data point in a set [34]. It uses Z multivariate Gaussian distributions, where each z_{th} component can be calculated by the mean of each component μ_z, covariance matrix Λ_z and component weight coefficient ω_z. Hence, the PDF of GMM can be expressed as

$$p(u) = \sum_{z=1}^{Z} \omega_z \mathbb{N}(u|\mu_z, \Lambda_z), \tag{16}$$

$$\mathbb{N}(u|\mu_z, \Pi_z) = \frac{1}{\sqrt{(2\Pi)^Z |\Lambda_z|}} exp\left(-\frac{1}{2}(u-\mu_z)^T \Lambda_z (u-\mu_z)\right), \tag{17}$$

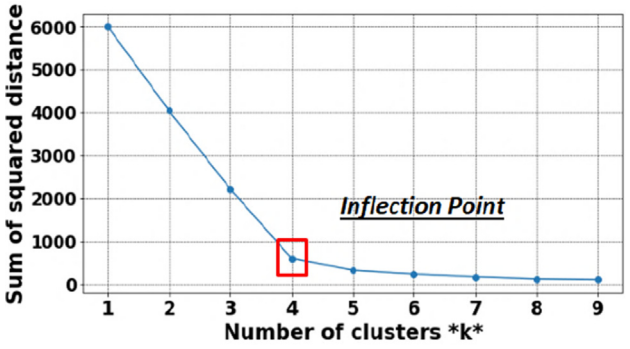
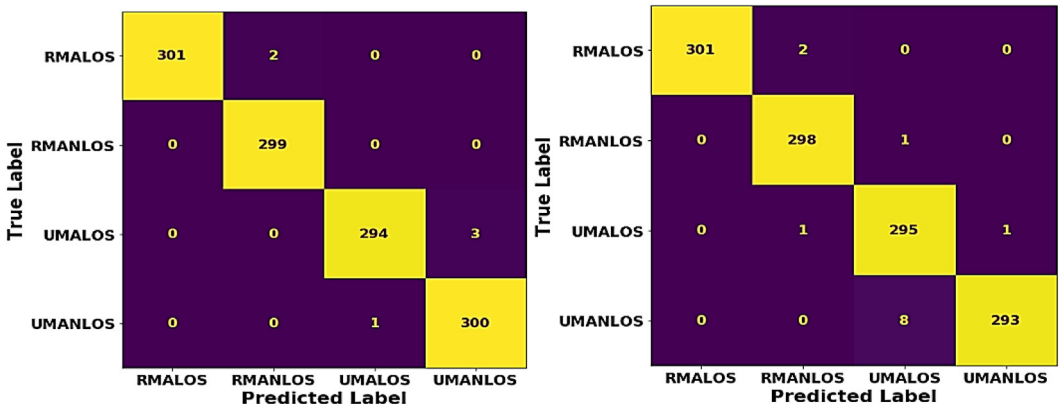


Fig. 5 Elbow method for k-Means clustering.



(a) k-NN confusion matrix. (b) SVM confusion matrix.

Fig. 6 The supervised learning confusion matrices.

$$\sum_{z=1}^{Z} \omega_z = 1. \tag{18}$$

Hence, the GMM can fit most of the probability distributions with a known number of clusters. The expectation-maximization (EM) technique is used to estimate the hyper-parameters of GMM [15] and [35].

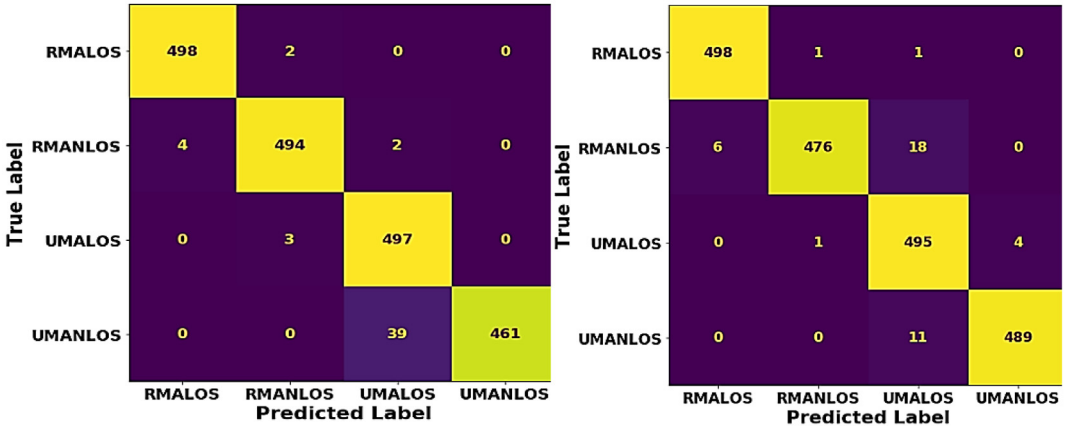
As the unsupervised learning has no label. Then, evaluating the model may require multiple trials and errors in order to determine the known classes.

4.3. Evaluation and comparison

In ML, the model accuracy is possibly calculated as the percentage of correctly predicted tests by dividing the number of correct predictions by the number of total predictions. But, although using accuracy as a defining parameter and the only evaluation metric for the proposed model is not enough, it is usually recommended to include precision and recall as well. Recall is a metric that measures how many correct positive predictions are produced out of all possible positive predictions. The precision is the number of true positives divided by the total number of positive predictions (i.e., the number of true positives plus the number of false positives). These 3 metrics can be obtained easily by a confusion matrix. Figs. 6 and 7 display the confusion matrix for these algorithms, where the results of accuracy, precision and recall are shown in Tables 4, 5 and 6, respectively.

In addition, for the comparison purpose, the conventional model [15] has no regularization technique adopted and has 3 principal components tested on the same dataset. Table 4 shows a comparison in performance between the proposed model and the conventional one for all of the above-mentioned algorithms when performed on the dataset in terms of classification accuracy.

In Table 3, where the ElasticNet regularization took into consideration 4 important features, so, the features are dropped from 7 to 4. In addition, a dimension reduction complexity is achieved since the number of features is reduced earlier so that the sum of principal features explained variance TEV is 72% with only 2 principal components. As a consequence, a dimension reduction took place by reducing the dimension from 4 to 2 and these 2 principal components are the ML input variables. However, with only 2 principal com-



(a) k-Means confusion matrix. (b) GMM confusion matrix.

Fig. 7 The unsupervised learning confusion matrices.

Table 4 Overall accuracy comparison between the proposed and the conventional one.

Models	k-NN	SVM	k-Means	GMM
Proposed	99%	99%	97%	98%
Conventional	97%	96%	89%	90%

Table 5 Precision comparison between the proposed model and the conventional one.

Classes **Proposed / Conventional**	k-NN	SVM	k-Means	GMM
RMA LoS	100 / 96	100 / 94	99 / 77	99 / 79
RMA NLoS	99 / 99	99 / 99	99 / 85	100 / 84
UMA LoS	100 / 93	97 / 89	92 / 98	94 / 98
UMA NLoS	99 / 99	100 / 100	99 / 99	99 / 98
Overall %	99.5 / 96.7	99 / 95.5	97.3 / 89.5	98 / 90

Table 6 Recall comparison between the proposed model and the conventional one.

Classes **Proposed / Conventional**	k-NN	SVM	k-Means	GMM
RMA LoS	99 / 94	99 / 91	99 / 82	100 / 82
RMA NLoS	100 / 100	100 / 99	99 / 99	95 / 100
UMA LoS	99 / 96	99 / 95	99 / 75	99 / 74
UMA NLoS	100 / 98	97 / 97	92 / 96	98 / 98
Overall %	99.5 / 97	98.8 / 95.5	97.3 / 88	98 / 88.5

ponents, ML algorithms achieved higher accuracy, especially in unsupervised learning algorithms.

Table 4 shows that all of the above-mentioned algorithms have better classification performance in the proposed model, especially the unsupervised algorithms. This is because they could cluster more accurately when the model is regularized. Moreover, the k-NN gives the highest accuracy and compared with the conventional model. It increases from 97% to 99% and the SVM accuracy is increased from 96% to 99%. In unsupervised learning, there is a dramatic increment of accuracy where, the k-Means accuracy increases from 89% to 97%. The GMM showed a great result with an accuracy increment from 90% to 98%.

Tables 5 and 6 display the comparison between both models in terms of precision and recall, respectively, for each class. The total precision is increased in all of the algorithms alongside the total recall. The proposed model is reliable for classification tasks. Moreover, clustering tasks of k-Means and GMM are showing high precision and recall of more than 97% and 98%, respectively.

5. Conclusion

In this work, an enhanced feature selection technique in 5G and beyond with generalization ability in wireless communication scenario prediction is introduced. The proposed model is used to improve the classification performance of ML predictive models employed to MIMO wireless communication systems. The proposed model reduces the computational complexity for each ML classifier due to less input variables. This process includes regularization of ElasticNet in order to select the most important features among a dataset of SSF and LSF parameters such as DS, KF, PL, esA, esD, asA and asD. In addition to, using k-PCA reduces the data dimensionality even more. Moreover, by comparing the proposed method with the conventional one [15], the model could take only two principal features instead of three and achieves higher TEV of 72%. So, its computational complexity is also reduced. The four ML algorithms, k-NN, SVM, k-Means and GMM, are tested with same parameters for both the proposed and the conventional models. The ML input variables are the two principal components of the proposed model and the three principal components of the conventional model. The accuracy of the proposed model for each classifier is increased by 2%, 3%, 8% and 8% for k-NN, SVM, k-Means and GMM, respectively.

As a future work, the adoption of Long Short Term Memory (LSTM) can be integrated into the wireless communication scenario classification problems. The exploration of the features of Doppler spectrum can also be considered.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] K. Chen, Q. Kong, Y. Dai, Y. Xu, F. Yin, L. Xu, S. Cui, Recent advances in data-driven wireless communication using Gaussian processes: A comprehensive survey, China Commun. 19 (1) (2022) 218–237, https://doi.org/10.23919/JCC.2022.01.016.

[2] J. Kaur, M.A. Khan, M. Iftikhar, M. Imran, Q. Emad Ul Haq, Machine learning techniques for 5G and beyond, IEEE Access 9 (2021) 23472–23488, https://doi.org/10.1109/ACCESS.2021.3051557.

[3] F. Burkhardt, S. Jaeckel, E. Eberlein, R. Prieto-Cerdeira, QuaDRiGa: A MIMO channel model for land mobile satellite, The 8th European Conference on Antennas and Propagation (EuCAP 2014), The Hague, Netherlands, 2014, pp. 1274–1278, doi: 10.1109/EuCAP.2014.6902008.

[4] N.H.M. Adnan, I.M. Rafiqul, A.H.M.Z. Alam, Massive MIMO for fifth generation (5G): Opportunities and challenges, 2016 International Conference on Computer and Communication Engineering (ICCCE), Kuala Lumpur, Malaysia, 2016, pp. 47–52, doi: 10.1109/ICCCE.2016.23.

[5] L.i. Yan, X. Fang, L.i. Hao, Y. Fang, A fast beam alignment scheme for dual-band HSR wireless networks, IEEE Trans. Veh. Technol. 69 (4) (2020) 3968–3979, https://doi.org/10.1109/TVT.2020.2971856.

[6] T.R.N, R. Gupta, A Survey on machine learning approaches and its techniques, 2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), Bhopal, India, 2020, pp. 1-6, doi: 10.1109/SCEECS48394.2020.190.

[7] S. Wenhui, W. Kejin, Z. Aichun, The development of artificial intelligence technology and its application in communication security, 2020 International Conference on Computer Engineering and Application (ICCEA), Guangzhou, China, 2020, pp. 752-756, doi: 10.1109/ICCEA50009.2020.00164.

[8] S. Mahajan, R. Harikrishnan, K. Kotecha, Prediction of network traffic in wireless mesh networks using hybrid deep learning model, IEEE Access 10 (2022) 7003–7015, https://doi.org/10.1109/ACCESS.2022.3140646.

[9] C. Huang, A.F. Molisch, R. Wang, P. Tang, R. He, Z. Zhong, Angular information-based NLOS/LOS identification for vehicle to vehicle MIMO system, 2019 IEEE International Conference on Communications Workshops (ICC Workshops), Shanghai, China, 2019, pp. 1-6, doi: 10.1109/ICCW.2019.8756726.

[10] Y. Yu, F. Liu, S. Mao, Fingerprint extraction and classification of wireless channels based on deep convolutional neural networks, Neural Process Lett. 48 (3) (2018) 1767–1775, https://doi.org/10.1007/s11063-018-9800-1.

[11] S. Zhang, W. Xing, Object tracking with adaptive elastic net regression, 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 2017, pp. 2597-2601, doi: 10.1109/ICIP.2017.8296752.

[12] M. Tanveer, H.-K. Tan, H.F. Ng, M.K. Leung, J.H. Chuah, Regularization of deep neural network with batch contrastive loss, IEEE Access 9 (2021) 124409–124418, https://doi.org/10.1109/ACCESS.2021.3110286.

[13] P.M. Hasan, N.A. Sulaiman, F. Soleymani, A. Akgül, The existence and uniqueness of solution for linear system of mixed Volterra-Fredholm integral equations in Banach space, AIMS Math. 5 (1) (2020) 226–235, https://doi.org/10.3934/math.2020014.

[14] M.A. Abusubaih, S. Khamayseh, Performance of machine learning-based techniques for spectrum sensing in mobile cognitive radio networks, IEEE Access 10 (2022) 1410–1418, https://doi.org/10.1109/ACCESS.2021.3138888.

[15] J. Zhang, L. Liu, Y. Fan, L. Zhuang, T. Zhou, Z. Piao, Wireless channel propagation scenarios identification: A perspective of machine learning, IEEE Access 8 (2020) 47797–47806, https://doi.org/10.1109/ACCESS.2020.2979220.

[16] Amira I. Zaki, Mahmoud Nassar, Moustafa H. Aly, Waleed K. Badawi, A generalized spatial modulation system using massive MIMO space time coding antenna grouping, Entropy 22 (12) (2020), 1350(1-10).

[17] T. Someya, T. Ohtsuki, SAGE algorithm for channel estimation and data detection with tracking the channel variation in MIMO system, IEEE Global Telecommunications Conference, 2004. GLOBECOM'04, vol. 6, Dallas, TX, USA, 2004, pp. 3651-3655. doi: 10.1109/GLOCOM.2004.1379050.

[18] A.M. Al-Samman, M.N. Hindia, T.A. Rahman, Path loss model in outdoor environment at 32 GHz for 5G system, in: 2016 IEEE 3rd International Symposium on Telecommunication Technologies (ISTT), 2016, pp. 9–13, https://doi.org/10.1109/ISTT.2016.7918076.

[19] A. Doukas, G. Kalivas, Rician K factor estimation for wireless communication systems, 2006 International Conference on Wireless and Mobile Communications (ICWMC'06), Bucharest, Romania, 2006, pp. 69-69, doi: 10.1109/ICWMC.2006.81.

[20] H. Arslan, T. Yucek, Delay spread estimation for wireless communication systems, The 8th IEEE Symposium on Computers and Communications. ISCC 2003, vol. 1, 2003, pp. 282-287. doi: 10.1109/ISCC.2003.1214135.

[21] A. Alshammari, S. Albdran, M.A.R. Ahad, M. Matin, Impact of angular spread on massive MIMO channel estimation, 2016 19th International Conference on Computer and Information Technology (ICCIT), Dhaka, Bangladesh, 2016, pp. 84-87, doi: 10.1109/ICCITECHN.2016.7860173.

[22] Q. Li, J. Zhao, X. Zhu, A kernel PCA radial basis function neural networks and application, in: 2006 9th International Conference on Control, Automation, Robotics and Vision, Singapore, 2006, pp. 1–4, https://doi.org/10.1109/ICARCV.2006.345230.

[23] V.N.G. Raju, K.P. Lakshmi, V.M. Jain, A. Kalidindi, V. Padma, Study the influence of normalization/transformation process on the accuracy of supervised classification, 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2020, pp. 729-735, doi: 10.1109/ICSSIT48917.2020.9214160.

[24] R. Muthukrishnan, R. Rohini, LASSO: A feature selection technique in predictive modeling for machine learning, 2016 IEEE International Conference on Advances in Computer Applications (ICACA), Coimbatore, India, 2016, pp. 18-20, doi: 10.1109/ICACA.2016.7887916.

[25] N. Kwak, Nonlinear projection trick in kernel methods: An alternative to the kernel trick, IEEE Trans. Neural Networks Learn. Syst. 24 (12) (Dec. 2013) 2113–2119, https://doi.org/10.1109/TNNLS.2013.2272292.

[26] F. Soleymani, A. Akgül, Improved numerical solution of multi-asset option pricing problem: A localized RBF-FD approach, Chaos, Solitons Fract., Elsevier 119 (2019) 298–309, https://doi.org/10.1016/j.chaos.2019.01.003.

[27] M.T. Vu, T.J. Oechtering, M. Skoglund, Hypothesis testing and identification systems, IEEE Trans. Inf. Theory 67 (6) (2021) 3765–3780, https://doi.org/10.1109/TIT.2021.3076497.

[28] B.P. Salmon, W. Kleynhans, C.P. Schwegmann, J.C. Olivier, Proper comparison among methods using a confusion matrix, in: 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2015, pp. 3057–3060, https://doi.org/10.1109/IGARSS.2015.7326461.

[29] K. Taunk, S. De, S. Verma, A. Swetapadma, A brief review of nearest neighbor algorithm for learning and classification, 2019 International Conference on Intelligent Computing and Control Systems (ICCS), Madurai, India, 2019, pp. 1255-1260, doi: 10.1109/ICCS45141.2019.9065747.

[30] A. Akgül, A novel method for a fractional derivative with non-local and non-singular kernel, Chaos, Solitons Fract. 114 (2018) 478–482, https://doi.org/10.1016/j.chaos.2018.07.032.

[31] S. Ghosh, A. Dasgupta, A. Swetapadma, A study on support vector machine based linear and non-linear pattern classification, 2019 International Conference on Intelligent Sustainable Systems (ICISS), Palladam, India, 2019, pp. 24-28, doi: 10.1109/ISS1.2019.8908018.

[32] W.K. Badawi, Z.M. Osman, M.A. Sharkas, M. Tamazin, A classification technique for condensed matter phases using a combination of PCA and SVM, Progress In Electromagnetics Research Symposium - Spring (PIERS), St. Petersburg, Russia, May 2017, pp. 326-331.

[33] K.P. Sinaga, M.-S. Yang, Unsupervised K-means clustering algorithm, IEEE Access 8 (2020) 80716–80727, https://doi.org/10.1109/ACCESS.2020.2988796.

[34] H. Li, Y. Li, S. Zhou, J. Wang, Wireless channel feature extraction via GMM and CNN in the tomographic channel model, J. Commun. Inform. Networks 2 (1) (March 2017) 41–51, https://doi.org/10.1007/s41650-017-0004-z.

[35] F. Soleymani, A. Akgül, European option valuation under the Bates PIDE in finance: A numerical implementation of the Gaussian scheme, Discrete Continuous Dyn. Syst.-S 13 (3) (2020) 889–909.