Research Journal of Applied Sciences 13 (9): 499-512, 2018
ISSN: 1815-932X

Indoor Wireless LAN Fingerprints Parameterisation and Classification in Academic Environments

Moses E. Ekpenyong
Department of Computer Science, University of Uyo, PMB. 1017 520003 Uyo, Nigeria

Abstract: In this study, we explore Machine Learning (ML) techniques to indoor Wireless Local Area Network (WLAN) Fingerprints (FPs) parameterisation and classification in academic environments. First, relevant indoor location (received signal strength indication and site specific) features were abstracted from the proposed area of study (University of Uyo, Nigeria) in a previous research to serve as fingerprints to the current research. Second, an unsupervised principal component analysis methodology was employed to produce Principal Component Dominant Features (PCDFs) for the first three principal components (components with eigenvalues of at least unity). These components revealed the degree of variances exhibited by the selected FPs. Third, using three ML classifiers (Support Vector Machine: SVM, k-Nearest Neighbour: k-NN, decision tree and Adaptive Neuro-Fuzzy Inference System: ANFIS) a classification of the PCDFs was performed. Results obtained showed that decision tree and linear SVM classifiers were excellent at predicting large datasets an important precursor to accommodating scalability in WLAN environments and areas with localisation challenges such as difficult terrains, heavy interference and spatial or uneven distribution of wireless infrastructure as these classifiers maintained high classification accuracies of above 90%. For small datasets, ANFIS gave good classification accuracy when compared with other classifiers.

Key words: Fingerprint, indoor localisation, pattern classification, principal component analysis, received signal strength, machine learning

INTRODUCTION

Positioning systems are usually classified based on the target environment either as indoor, outdoor or mixed type. Indoor positioning systems provide precise location of structures within close proximity (of environments such as shopping malls, airports, subways and academic campuses). Given the complexities inherent in indoor environments the development of indoor localisation techniques are always associated with numerous challenges such as high Non-Line of Sight (NLOS), influence of obstacles such as walls, equipment, human and vehicular movements, fewer dimensions, doors and more. In outdoor positioning systems, global navigation satellite systems such as the Global Positioning System (GPS) have been consistently used in a wide range of applications including tracking and asset management, transport navigation and tour guides, synchronisation of telecomm networks and geodetic survey. GPS works best in outdoor positioning but fails to perform well in urban environments which are mostly impaired by poor/difficult terrain, presence of barriers (such as walls, buildings, trees, cars and underground structures) as GPS satellite signals are weakened by these barriers, thus, rendering GPS ineffective for indoor localisation (Gu *et al*., 2009).

In Wireless Local Area Network (WLAN) positioning, the characteristic patterns of the perceived service quality is highly dependent on the environmental features or characteristics as well as installed Access Points (APs). Hence, constraining positioning algorithms to a subset of Reference Points (RPs) with related characteristics emerged as a viable approach (Kushki *et al*., 2006). This approach compressed the search space of the user location to a smaller number of RPs, followed by a finer search on the refined set of RPs (Azizyan *et al*., 2009). Performance criteria associated with localisation systems include accuracy, responsiveness, coverage, adaptivity, scalability and complexity and cost (Farid *et al*., 2013). The accuracy (or location error) of a system is a vital user requirement of positioning systems and may be reported as the error distance between the estimated location and the actual mobile location. Responsiveness determines how quickly the location estimate of a moving target is updated. Coverage is necessary when evaluating the effectiveness of a positioning system and is closely related to accuracy. Adaptivity concerns the ability of the localisation system to cope with changes or sudden influence within the localisation environment and is also a direct correlate of accuracy. An adaptive system can also prevent the need for repeated calibration. Scalability suggests how well the system performs when it operates with a larger number of location requests and a larger coverage. Cost and complexity are tradeoffs between the system complexity and the accuracy. They affect the overall cost of the system.

Severe fluctuation of the Received Signal Strength Indication (RSSI) constitutes a challenging problem in WLAN localisation, even for a stationary client (Lu *et al*., 2016). Previous research have evolved various approaches to improving location accuracy. The RSSI location estimation systems can be classified along two architectures: deterministic (fingerprint or map-based) and probabilistic (prediction-based) architectures. In deterministic techniques, the wireless device learns by listening to the channel as it receives beacons periodically set by Access Points (APs) and records their Relative Signal Strength (RSS) values at known positions of the physical space (Bahl and Padmanabhan, 2000). Subsequently, the system records RSS values from received beacons but at random unknown positions. Deterministic algorithms use a similarity metric to differentiate online signal measurement and fingerprint data and the target is found at the closest fingerprint location in signal space (Han *et al*., 2014). Euclidean distance and cosine similarity are recognised metric implemented for signal comparison of deterministic architecture. The major advantage of deterministic methods is the ease of implementation and can be achieved using an algorithm such k-Nearest Neighbour (k-NN) with low computational complexity. Some more advanced deterministic algorithms such as Support Vector Machines (SVMs), (Wu *et al*., 2004) and linear discriminant analysis (Nuno-Barrau and Paz-Borrallo, 2006) show better localization accuracy with higher computational cost. However, due to the unpredictable nature of RSS measurements, most deterministic systems show increased computational cost. Probabilistic techniques employ RSS and radio propagation models to locate the distance of a wireless user from the AP. Probabilistic algorithms are based on statistical inference between the target signal measurement and stored fingerprints (Mirowski *et al*., 2014). Hence, given a training set these algorithms find the target's location with the maximum likelihood (He and Chan, 2016).

Literature review: Advances in localisation-based technologies coupled with the increased benefits of ubiquitous computing and context-dependent information have inspired great interest in location-based services and applications (Farid *et al*., 2013). The development of communication technology certainly demands the use of location-based services. Luo and Fu (2017) proposed an algorithm for indoor localization based on RSS collected from APs. The localisation algorithm contained offline information acquisition phase and online positioning phase. First, they selected working APs using a selection algorithm that scans their functioning status based on signal stability. Second, a Kernel Principal Component Analysis (KPCA) was used to eliminate redundant data and maintain useful characteristics for nonlinear feature extraction. Third, the Affinity Propagation Clustering (APC) algorithm utilising RSS values was used to classify data samples to narrow the positioning range. They found that their algorithm improved the accuracy and computational complexity. Several challenges facing WLAN fingerprint localisation schemes have also been identified by (Fang *et al*., 2008; Kupershtein *et al*., 2013; Xiang *et al*., 2004). By Fang and Lin (2012) an approach to developing a WLAN-based location fingerprinting system was proposed. Their algorithm intelligently transformed Received Signal Strength (RSS) into Principal Components (PCs) such that the information from all APs is efficiently utilised. Results of their experiments conducted in a realistic WLAN environment showed that the mean error was reduced by about 33% and the complexity by 40% when compared to existing methods.

Ekpenyong *et al*. (2017) obtained the perceived performance in a field survey from an academic environment and using the Interval Type-2 Fuzzy Logic (IT2FL), uncertainties inherent in the field data were efficiently modelled for accurate estimation of the Service Quality (SQ). The expected performance was then modelled using two unsupervised tools: PCA and SOM, to abstract the most relevant features and observe similarity patterns between the abstracted features, respectively. An Adaptive Neuro-Fuzzy Inference System (ANFIS) was finally used to optimise the system performance. Results obtained showed that ANFIS sufficiently modelled the SQ as the Root Mean Square Error (RMSE) values of the train and test data set were approximately the same for the three study sites considered. However, combining the three campuses produced the least Mean Absolute Error (MAE). Ekpenyong *et al*. (2018) presents the design and construction of WLAN test bed infrastructure that intelligently supports the tuning and visualisation of the SQ where evolutionary PSO-ANFIS and GA-ANFIS are independently trained to fine-tune the test bed performance. Their results showed that both systems performed well as their Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) for both test and train data were very close.

The essential challenge in WLAN positioning system is the highly uncertain and nonlinear RSS which degrade the positioning accuracy and increase the data collection cost. To address this challenge, Deng *et al.* (2012) proposed the nonlinear discriminative feature extraction of RSS using Kernel Direct Discriminant Analysis (KDDA) to extract location features in a kernel space. Furthermore, they employed location clustering to localise the feature extraction and effectively avoid the suboptimality caused by variability of RSS over physical space. Finally, the relationship between extracted features and physical locations was established by Support Vector Regression (SVR). Experimental results showed that the proposed approach had higher accuracy while reducing the data collection cost significantly. Applying machine learning algorithms to cognitive radio appears to be the state-of-art, given their robustness in feature classification. Hou *et al.* (2011) proposed a combination of dimensionality reduction and SVM classification. Using measured Wi-Fi signals with high Signal to Noise Ratio (SNR), the Degrees of Freedom (DOF) of Wi-Fi signals was extracted by dimensionality reduction techniques. Their results showed dimensionality reduction improved the classification performance, as the classification error rates obtained from the reduced features were better than the error rates obtained from all the features. Song and Wang (2017) proposed semi-supervised learning algorithm called Co-Forest was used to create and iteratively refine a random forest ensemble classifier that performs well for location estimation. They demonstrated their system on a large number of unlabeled RSS samples with some labelled RSS samples. Results of experiments conducted in a real indoor environment showed that their proposed strategy reduced the demand for large quantities of labelled samples and achieved good positioning accuracy.

Positioning methods using existing wireless LAN APs have been well explored (Yim, 2008). Among the methods explored, the fingerprint methods appear to be the most promising and probabilistic method, k-NN (Nearest Neighbour), Bayesian classification and neural networks constitute the most used techniques. Yim (2008) proposed a technique that builds a decision tree during the off-line phase and determines a user's location referring to the tree. A time complexity and experimental accuracy analysis of the proposed technique was also presented. Obtaining training data is particularly challenging due to a high number of possible interfering devices, difficulty in obtaining precise timings and the need to measure the devices in varying conditions. Longi *et al.* (2017) focused on semi-supervised learning using convolutional network to identify frequency devices from signal data. They aimed to minimise the need for reliable training samples while utilizing larger amounts of unsupervised labels to improve the accuracy. They showed that with few seconds of training data for each device their method is sufficient for highly accurate recognition. Narayanan *et al.*, (2016) considered indoor user localisation as a pattern classification problem using Fuzzy Decision Tree (FDT) for locating users. Li *et al.* (2014) used the UJIndoorLoc database as data set and employed PCA for feature selection and building prediction models based on decision tree, gradient boosting, kNN and SVM, respectively. Their experiment results indicate that combining kNN and Gradient Boosting provided high prediction accuracy for Indoor positioning. kNN showed good performance for large dataset and Gradient Boosting had small cross validation error for small data volume and robust to missing data. By Roy *et al.* (2015) a decision tree based classification was proposed to find the best node based on signal strength and the environmental conditions. They simulated the scenario using NS2 platform. An ensemble based learning algorithm with bagging and adaptive boosting in C4.5 was also employed for improving the performance. A performance comparison showed that the boosted decision tree algorithm gave the highest classification accuracy of 95.73%. Node localization is employed in many wireless networks as it can be used to improve routing and enhance security. In Almuzaini and Gulliver, (2012) a new algorithm based on decision tree classification and k-means clustering was proposed for node localisation in wireless networks. They observed that their algorithm performed better than the Linear Least Squares (LLS) and Weighted Linear Least Squares based on Singular Value Decomposition (WLS-SVD) algorithms, even when the geometric anchor distribution about an un-localised node degraded.

MATERIALS AND METHODS

Proposed system framework: The proposed system framework for the parameterisation and classification of indoor location features is presented in Fig. 1. The framework has three phases. The first phase deals with RSSI information measurement which captures signal strength data from the various mobile and fixed infrastructures (APs and mobile users). A selection of fuzzifiable features for efficient signal-print representation

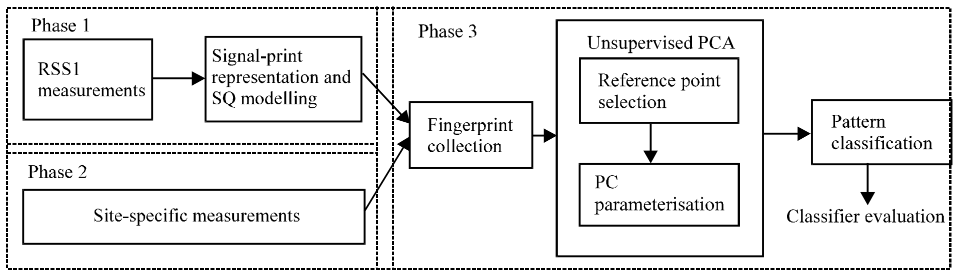

Fig. 1: System framework

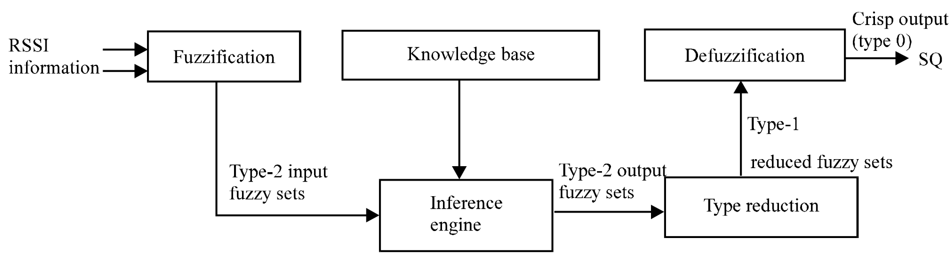

Fig. 2: Structure of interval type-2 fuzzy logic for WLAN SQ determination (Ekpenyong *et al*., 2018)

and service quality modelling is then carried out. In phase 2, measurement of site-specific features of the various indoor structures or buildings was performed. In phase 3, data obtained from these phases were then merged into a single dataset to form a collection of fingerprints where acceptable reference points were selected and parameterised using an unsupervised PCA technique. The first three principal components with dominant features are finally subjected to different classifiers, to examine which classification technique produces the best accuracy for modelling the parameterised location features.

Signal strength data capture and feature extraction: A capture of the Relative Signal Strength Indication (RSSI) information of the service area was performed through a scan of the study environments (where APs are located) using the Acrylic WiFi professional a WiFi analyser software that identifies access points and WiFi channels and is useful for analysing and resolving incidences on 802.11a/b/g/n/ac wireless networks in real time. The functionalities of Acrylic include:

- Efficient visualisation of wireless network performance and connected users
- Access point data transmission speeds identification and channels optimisation
- Access WiFi network detailed information collation and visualisation including hidden wireless networks

Site specific information were also gathered from all the buildings. A list of the RSSI and site information is summarised in Table 1. A full methodological process employed during the survey and sample data of RSSI and site-specific measurements captured is found in Ekpenyong *et al*. (2018). To obtain precise SQ, parameters with Fuzzy Membership Function (FMF) were characterized. The RSSI data formed our major parameters of interest and were passed as inputs to the Fuzzy-type-2 Logic System (F2FLS) to provide precise representation of the SQ. The proposed framework as shown in Fig. 2 has five components namely, fuzzifier, knowledge base, inference engine, type-reducer and defuzzifier. The fuzzification module maps the crisp input (RSSI information) to Interval Type-2 Fuzzy Sets (IT2FSs) using a defined Triangular Membership Function (TMF) method. The following RSSI parameters were considered as inputs: RSSI, Number of Channels (NChannels) and Maximum Baud Rate (MBR) while SQ represents the output variable. The Universe of Discourse (UoD) for the input and output variables and the domain intervals of the variables, as well as the range of each variable used to establish the fuzzy models are as defined in Table 2. TMFs were adopted to evaluate each input and output MFs. Hence, the TMF (for a given input/output, x) μ (x) as shown in Eq. 1 depends on three parameters p-p_2 and indicates the mapping of each input (RSSI, NChannels and MBR) measurements or output (SQ) parameters, required to obtain the membership values:

Table 1: RSSI and site information captured form the service area

RSSI information-captured using Acrylic Professional software		
RSSI parameters	Meaning	Data type
SSID	Service Set Identifier	String
MAC	Media Access Control address	String
RSSI	Received Signal Strength Indicator	Number
SNR	Signal to Noise Ratio	Number
NoChan	Signal communication Channel	Number
ChanWidth	Channel bandwidth	Number
802.11	Infrastructure type	String
MBR	Maximum Baud Rate	Number
WEP	Wired Equivalent Privacy	String
Vendor	Infrastructure Vendor	String
Mgt	Number of traffic Managed	Number
VenType	Vendor Type	String
Latitude	Geographic coordinate of study location (center of a building), north-south on the earth's surface	Number Number
Longitude	Angular distance of study location (center of a building), East-West of the equator	
Time	Time of capture	String
Site information-captured during site survey		
Site parameter	Meaning	
BID	Building Identifier	String
Bloc	Building location indoor/outdoor?	String
BType	Building type	Number
BSize	Building size	Number
BPurp	Purpose for which building is used	Number
BHeight	Building height	Number
DFNOC	Distance of building from NOC	Number
Floor	Number of floors	Number
NOR	Number of Rooms	Number
Pathloss	Signal propagation pathloss	Number

Table 2: Domain intervals of input and output variables

Variables	Lower boundss	Upper bound	Units
	Input variables		
RSSI	-100	-5	dBm
Nchannels	0	20	-
MBR	0	350	msec
	Output variable		
SQ	0	100	%

$$\mu(x)=\begin{cases}0; & \text{if } x<p_1\\ \dfrac{x-p_1}{p-p_1}; & \text{if } p_1\le x\le p\\ \dfrac{p_2-x}{p_2-p}; & \text{if } p\le x\le p_2\\ 0; & \text{if } x>p_2\end{cases} \tag{1}$$

where, p, defines the triangular peak location while p_1 and p_2, define the triangular end points. Figure 3 shows the triangular shape IT2FS with its principal T1FS, bounded by an UMF and a LMF. Applying the prameters in Fig. 3 to Eq. 1, we derived Eq. 2 and 3 the detailed computation formulae for the UMF($\bar{\mu}$ x) and LMF ($\underline{\mu}$ x), respectively, given an input/output variable (x):

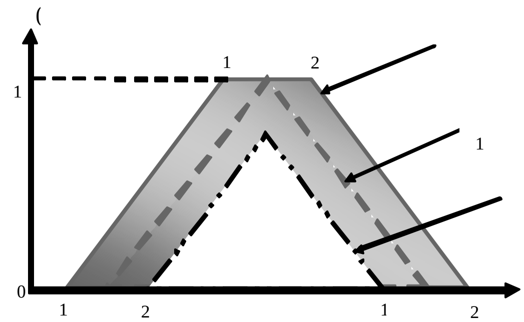

Fig. 3: Triangular shape IT2FS

$$\bar{\mu}(x)=\begin{cases}0; & \text{if } x<l_1\\ \dfrac{x-l_1}{p_1-l_1}; & \text{if } l_1\le x<p_1\\ 1; & \text{if } p_1\le x\le p_2\\ \dfrac{r_2-x}{r_2-p_2}; & \text{if } p_2<x\le r_2\\ 0; & \text{if } x>r_2\end{cases} \tag{2}$$

$$\underline{\mu}(x)=\begin{cases}0; & \text{if } x<l_2\\ \dfrac{x-l_2}{p_2-l_2}; & \text{if } l_1\le x<p_1\\ \dfrac{r_2-x}{r_2-p_2}; & \text{if } \dfrac{r_1(p_2-l_2)+l_2(r_1-p_1)}{(p_2-l_2)+(r_1-p_1)}<x\le r_1\\ 0; & \text{if } x\ge r_2\end{cases} \tag{3}$$

where, I_1 and I_2, represent the left end point of both UMF and LMF, respectively and r_1 and r_2, represent the right end point of both LMF and UMF, respectively. The domain intervals for the study were then partitioned according to their lower and upper values, conditioned towards standard WLAN regulatory estimates (Mazar, 2016; Xue *et al*., 2017). Using the fuzzy RSSI datasets, a simulation of the IT2FL system was performed, to generate the SQ (c.f. Ekpenyong *et al*., 2018). A PCA was then used to reduce the dimension of the RSSI and site-specific parameters. In this study, we concentrate on the principal components with eigen values of at least unity, selected in Ekpenyong *et al*. (2017). These parameters now form the inputs to our PCA Model formulated in the next section. Six site specific parameters were accepted as major principal components out of the 25 parameters used in the survey. The accepted parameters and their respective eigen values include: BHeight (3.721831), BType (2.709880), NOR (2.455971), BSize (1.338041), Floor (1.246454) and BPurp (1.062904).

We also include the following RSS related parameters (DFNOC, Pathoss and SQ). The SQ represents the refined RSS information obtained from the application of IT2FL on selected RSSI data (RSSI, Nchannel and MBR) captured during the field survey. This refinement process was necessary to eliminate uncertainties inherent in the field data and will aid better location service estimation. The Pathloss and DFNOC parameters with eigen values 0.949341 and 0.841402, respectively were also selected on the grounds that they represent crucial Reference Point (RP) information for computing the quality of service offered by the WLAN. The additional features therefore extend the number of extracted features to nine.

Principal Component Analysis (PCA): PCA is a dimension reduction technique that transforms original set of fields into smaller set that accounts for most of the variance (or information) in the data. The new fields are called factors or Principal Components (PCs). It is a technique used to emphasize variation and bring out strong patterns in a dataset. The PCs are extracted sequentially with the first principal component accounting for the most variance in the data. Intuitively the first principal component is a vector that points in the direction in which the data are most "spread out". The second principal component is set up similarly but with the additional constraint that it must be uncorrelated with the first. Each subsequent principal component captures an increasingly lower percentage of variation in the data and is uncorrelated with the previously extracted principal components. Principal components can be used instead of the original fields in predictive models, to avoid the problems that can occur when highly correlated variables are used but at the expense of making model interpretation more difficult. Also, the method can be used to determine which groups of fields are likely to be jointly highly related to one another and help guide decisions in which fields to omit from a predictive model.

Model formulation: Suppose a set of indoor parameters for mobile and fixed infrastructure within an academic environment, here in referred to as Fingerprints (FPs) have been recorded at time instance (t_m), m = 1, 2, 3, ..., M with FP magnitudes $\left(f_j^i(t_1), f_j^i(t_2), ..., f_j^i(t_M)\right)$ at each Reference Point (RP) where i represents the AP index from a known set of Aps, S = {AP^i, AP^2, ..., AP^s. It is common practice to take advantage of same number of training samples, M at each RP. The FPs from all APs at time t_m at a reference position p_j are organised as vector $f_j(t_m) = [f^1{}_j(t_m), f^2{}_j(t_m), ..., f^s{}_j]^T$. Hence, the entire radio map at recorded time instance t_m is:

$$R(t_m) = f_1(t_m), f_2(t_m), ..., f_N(t_m) = \begin{pmatrix} f_1^1(t_m)...f_N^1(t_m) \\ \vdots \quad \ddots \quad \vdots \\ f_1^s(t_m)...f_N^s(t_m) \end{pmatrix}; m = 1, 2, ..., M \tag{4}$$

Also let, $f^i_j = [f^i_j(t_1), f^i_j(t_2), ..., f^i_j(t_M)]^T$, $f^i(t_m) = [f^i{}_1(t_1), f^i{}_2(t_2), ..., f^i{}_N(t_m)]^T$ and $f_j(t_m) = [f^1{}_j(t_m), f^2{}_j(t_2), ..., f^s{}_j(t_m)]^T$ indicate (at different time instances) a vector of FPs, at different RPs and for different APs, respectively. If the time sequence of the radio maps, R (t_m) is averaged over the recorded time interval, then the average radio map time is:

$$\varphi = (\varphi_1, \varphi_2, ..., \varphi_N) = \begin{pmatrix} \varphi_1^1 & \cdots & \varphi_N^1 \\ \vdots & \ddots & \vdots \\ \varphi_1^s & \cdots & \varphi_N^s \end{pmatrix} \tag{5}$$

Where:

$$\varphi_j = \left[\varphi_j^1, \varphi_j^2, ..., \varphi_j^s\right]^T$$

And:

$$\varphi_j^i = \frac{1}{M}\sum_{m=1}^{M} f_j^i(t_m)$$

Next, we map the recorded measurements to a domain of its PCs (Fang and Lin, 2012) and compute the sample covariance of the FP (indoor and outdoor characteristic) patterns at a RP, j, as:

$$\text{Cov}(i, i') = \frac{1}{M}\sum_{i=1}^{m}\left(f_j^i(t_m) - \varphi_j^i\right)\left(f_j^{i'}(t_m) - \varphi_j^{i'}\right)^T; \quad i, i' = 1, 2, ..., S;\ j = 1, 2, ..., N \tag{6}$$

and the global covariance matrix contains:

$$\text{Cov}_g(i, i') = \frac{1}{MN}\sum_{j=1}^{N}\sum_{m=1}^{M}\left(f_j^i(t_m) - \varphi_j^i\right)\left(f_j^{i'}(t_m) - \varphi_j^{i'}\right)^T; \quad i, i' = 1, 2, ..., S \tag{7}$$

The eigenvectors of the global covariance matrix, becomes:

$$\text{Cov}_g . v_s = \lambda_s . v_s;\ s = 1, 2, ..., S \tag{8}$$

The data transformation into its PCs is obtained by grouping the eigenvectors that correspond to the eigenvalues and arranged in decreasing order of magnitude as:

$$A = [v_1, v_2, ..., v_s]; \lambda_1 \geq \lambda_2 \geq, ..., \lambda_s \tag{9}$$

Feature selection and dimension reduction: A PCA-based unsupervised selection algorithm (Luo *et al*., 2008) is adopted in this study to select the dominant principal component features and eliminate redundant finger print patterns that may contribute to (poor) selection. PCA is useful in this research because it is a powerful tool for visualising high dimensional data. It shows quantified difference among observations and is useful for assessing data quality and the discovery of relationships between data points. Now, given an input space R^D and target space R^d; d<<D, let $X \epsilon R^{N \times D}$ be an input dataset of N samples and D features and $X \epsilon R^{N \times d}$ its low-dimensional representation. A dimension reduction technique is the mapping ϕ: $R^D \rightarrow R^d$ that optimizes a cost function ϵ: $R^d \rightarrow R$ on the target space. This problem can often be reduced to an eigen value problem whose eigen vectors defines the embedding Y. Assuming a training set with N samples $\{x_i\}_i^N=1$, each sample represented by an n-dimensional vector $x_i = [x_{1i}, x_{2i}, ..., x_{ni}]^T$, PCA can be considered as a linear transformation that maps data from the original measurement space to a new space populated by a set of new variables. Suppose the linear transform is denoted by matrix L, then pattern x in the new space is represented as:

$$y = L^T x \tag{10}$$

Where:

$$y = [y_1, y_2, ..., x_d]^T, \ L = [q_1, q_2, ..., q_d]^T$$

And:

$$q_j^T = [q_{1j}, x_{2j}, ..., x_{nj}], j = 1, 2, ..., d \tag{11}$$

where, d≤n but most often d»n. The new variables $y_j = j = 1, 2, ..., d$ are called Principal Components (PCs). Consider the projection of x_i on the k principal axis, then:

$$y_{ki} = q_k^T x_i = \sum_{j=1}^{n} q_{jk} x_{ik} \tag{12}$$

As observed in Eq. 9, the projection of a sample on the principal axis is a linear combination of all variables. However, some of the variables might be redundant, irrelevant or insignificant which indicates that feature selection can only be achieved through the identification of subset of variables whose roles are critical in determining data projections on the principal axes. We observe that the significance of a single variable x_j can be evaluated based on the value of the corresponding coefficient q_{jk}. An approximate method (Dash *et al*., 1997) is introduced here for feature selection in two inter-related steps: select a subset of relevant features and select critical features from the relevant features. A recurrence definition of Principal Component Dominant Feature (PCDF) about y_k can be thus defined as follows: for a specific principal component y, a variance with the largest coefficient in the component is a PCDF, for the remaining features, if x_j is a relevant feature about y_k, i.e., $\rho(x_j, y_k) > \omega$ and there exists no PCDF x_p which is subject to:

$$\frac{|\rho(x_j, x_p)|}{|\rho(x_j, x_p)|} \geq \theta(0 < \theta \leq 1)$$

then features x_j is a PCDF about y_k.

Pattern classification: Classification is the task of assigning an object to one or more categories. More precisely, it is the task of learning a target function f that maps each attribute set x to one of the predefined class labels y. The target function is also known informally as a classification model. A classification model is useful for both descriptive and predictive modelling. In this section we compare the effectiveness of four state-of-art and commonly used classification techniques with the aim of recommending which classification technique is best for WLAN problems in academic environments. The classification techniques employ machine learning and include: SVM, KNN, decision tree and ANFIS.

SVM (Cortes and Vapnik, 1995) is a discriminative classifier capable of deciphering subtle patterns in noisy and complex datasets, i.e., given labelled training data, SVM outputs an optimal hyperplane which categorises new exemplars. In SVM, a decision boundary that maximises the "Margin" separating the positive from the negative training data points. To find this, we minimise $1/2\|\vec{w}\|^2$ subject to the constraint $y_i(\vec{w}.\vec{x}_i+b) \geq 1$. The resulting lagrange multiplier equation that requires optimisation is:

$$L = \frac{1}{2}\|\vec{w}\|^2 - \sum_i \alpha_i \left(y_i\left(\vec{w}.\vec{x}_i + b\right) - 1\right) \tag{13}$$

Solving the Lagrangian optimization problem in Eq. 10 yields, w, b and α_i parameters that determines a unique maximal margin solution. On the maximum margin, the positive and negative points that stride the separating lines are called support vectors. The decision boundary lies at the middle of the margin. The definition of the separator is dependent only on the support vectors, hence, changing (adding deleting) non-support vector points will not change the solution. In this study we examine SVMs with the following kernel types: fine Gaussian, coarse Gaussian, quadratic and linear kernels.

k-NN classification (Friel and Pettitte, 2011) is one of the most fundamental and simple classification methods often preferred when there is little or no prior knowledge about the distribution of data. k-NN learns the input samples and predicts the response for a new sample by analysing a certain number (k) of the nearest neighbours of the sample using voting weight computation and so on. It is commonly based on the Euclidean distance between a test sample and the specific training samples. The kNN algorithm is first implemented by introducing some notations: Let $S = (x_i, y_i)$: $i = 1, 2, ..., n$ be the training set, where x_i is a d-dimensional feature vector and $y_i \in \{+1, -1\}$ is associated with the observed class labels. For simplicity, we consider a binary classification. Suppose that all training data are iid samples of random variables with unknown distribution. With previously labelled samples as the training set S, the k-NN algorithm constructs a local subregion $R(x) \subseteq R_d$ of the input space which is situated at the estimation point x. The predicting region R(x) contains the closest training points to x which is expressed as follows:

$$R(x)=\left\{\hat{x}\middle|D(x,\hat{x})\leq d_{(k)}\right\} \tag{14}$$

where $d_{(k)}$ is the k th order statistic of $\{D(x,\hat{x})\}_1^N$ and $D(x,\hat{x})$ is the distance metric. k[y] denotes the number of samples in region which is labelled. The k-NN algorithm is statistically designed for the estimation of posterior probability $p(y|x)$ of the observation point x:

$$p(y|x)=\frac{p(x|y)p(y)}{p(x)}\cong\frac{k[y]}{k} \tag{15}$$

For a given observation, x, the decision (x) is formulated by evaluating the values of k[y] and selecting the class with the highest value:

$$g(x)=\begin{Bmatrix}1,k[y=1]\geq k[y=-1]\\ -1,k[y=-1]\geq k[y=1]\end{Bmatrix} \tag{16}$$

Finally, the decision that maximises the associated posterior probability is used in the k-NN algorithm. For a binary classification problem where $y_i \in \{+1, -1\}$, the kNN algorithm produces the following decision rule:

$$g(x)=\text{sgn}\left(\text{ave}_{x_i\in R(x)}y_i\right) \tag{17}$$

In this study, we examine k-NNs with the following kernel types: fine, coarse, cosine and weighted kernels. Decision trees are a simple and widely used classification technique (Tan *et al*., 2006). Moving from one level to another in a decision tree requires a test condition to decide which branch to follow. This process is continued until a leaf node is reached. Decision trees research through recursive partitioning of the training set in order to obtain subsets that are as pure as possible to a given target class. Each node of the tree is associated to a particular set of records T that is split by a specific test on a feature. For instance, a split on a continuous attribute A can be induced by the test $A \leq x$. The set of samples T is then partitioned into two subsets that yield the left and right branches of the tree:

$$T_l=\{t\in T:t(A)\leq x\}$$

And:

$$T_r=\{t\in T:t(A)>x\}$$

Similarly, a categorical feature B can be used to induce splits according to its values. For instance, if $B = \{b_1, b_2, ..., b_k\}$ each branch i can be induced by the test $B = b_i$ The division step of the recursive algorithm that induces the decision tree observes all possible splits for each feature and tries to find the best one based on a chosen quality measure: the splitting criterion. If a dataset is induced on the following scheme: $A_1, ..., A_m$, C where A_j are attributes and C is the target class and all candidates splits are generated and evaluated by the splitting criterion. Splits on continuous and categorical attributes are generated as described above. The selection of the best split is usually carried out by impurity measures. The impurity of the parent node has to be decreased by the split. Let $(E_1, E_2, ..., E_k)$ be a split induced on the set of records E, then, a splitting criterion that makes used of the impurity measure $I(\cdot)$ is:

$$\Delta=I(E)-\sum_{i=1}^{k}\frac{|E_i|}{E}I(E_i) \tag{18}$$

Standard impurity measures are the Shannon entropy or the Gini index. Classification and Regression Tree (CART) uses the Gini index that is defined for a set E as follows: Let p_j be the fraction of samples in E of class c_j:

$$pj=\frac{|\{t\in E:t|C|=c_j\}|}{E} \tag{19}$$

Then:

$$\text{Gini}(E)=1-\sum_{j=1}^{Q}p_j^2 \tag{20}$$

where, Q is the number of classes. We examine three types of decision trees: simple, complex and medium decision trees.

ANFIS (Suparta and Alhasa, 2016; Yang *et al*., 2014) is a hybrid classifier that combines the capabilities of neural network and fuzzy logic, to learn features in a dataset and adjust the system parameters according to a given error criterion. It is trained with the back propagation gradient decent method in combination with the least squares method. The ANFIS architecture has five layers. The first and fourth layers contain an adaptive node while the other layers contain a fixed node. A brief description of each layer is as follows:

Layer 1: Every node in this layer adapts to a function parameter. The output from each node represents a degree of membership value that is given by the input of the membership functions. The membership function may be a Gaussian membership function (Eq. 18), a generalised bell membership function (Eq. 19) or any other type of membership function:

$$\mu_{Ai}(x) = \exp\left[-\left(\frac{x-c_i}{2a_i}\right)^2\right] \tag{21}$$

$$\mu_{Ai}(x) = \frac{1}{1+\left|\frac{x-c_i}{a_i}\right|^{2b}} \tag{22}$$

$$O_{1,i} = \mu_{Ai}(x), i=1,2 \tag{23}$$

$$O_{1,i} = \mu_{Bi-2}(y), i=3,4 \tag{24}$$

where, μ_{Ai} and μ_{Bi-2} are the degree of membership functions for fuzzy sets A_i and B_i, respectively and $\{a_i, b_i, c_i\}$ are the parameters of a membership function that can adjust the shape of the membership function. The parameters in this layer are typically referred to as the premise parameters.

Layer 2: Every node in this layer is fixed or Non-adaptive. The output node is the result of multiplying of signal coming into the node and delivered to the next node. Each node in this layer represents the firing strength for each rule. In the second layer, the T-norm operator with general performance, such as the and is applied to obtain the output:

$$O_{2i} = w_i = \mu_{Ai}(x) * \mu_{Bi}(y), i=1,2 \tag{25}$$

where, w_i is an output that represents the firing strength of each rule.

Layer 3: Every node in this layer is fixed or non-adaptive. Each node is a computation of the ratio between the ith rules firing strength and the sum of all rule's firing strengths. This result is known as the normalized firing strength:

$$O_{3i} = \bar{w}_i = \frac{w_i}{\sum_i w_i} \tag{26}$$

Layer 4: Every node in this layer is an adaptive node to an output with a node function defined as:

$$O_{4i} = \bar{w}_i f_i = \bar{w}_i (p_i x + p_i y + r_i) \tag{27}$$

where, $\bar{w}_i$ is the normalised firing strength from the previous layer (third layer) $(P_i x+p_i y+r_i)$ and is a parameter in the node. The parameters in this layer are referred to as consequent parameters.

Layer 5: The single node in this layer is a fixed or non-adaptive node that computes the overall output as the summation of all incoming signals from the previous node:

$$O_{5i} = \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \tag{28}$$

RESULTS AND DISCUSSION

Average PC analysis: An analysis of the average pathloss and SQ, computed for the three campuses under investigation (town, annex and main campuses) is presented in Fig. 4-6, respectively. Figure 4 and 5 both the town and annex campuses showed poor service quality with SQ and associated (pathloss) of 54.30% (84.77 dB) and 50.52% (82.24 dB) while the main campus showed good service quality SQ and associated (pathloss) of 60.21% (90.08 dB). The poor SQ at the town and annex campuses as observed in Ekpenyong *et al*. (2018) was due to poor terrain and congested or dense environment (at town campus) and widely spaced APs leading to poor signal reception (at annex campus). The main campus had few structures but the wide distance between the buildings and the Network Operating Centre (NOC) did not affect the signal quality because there was clearer line of sight, compared to the other campuses. The defect in the SQ is evident in the trend exhibited by the average pathloss and service quality plots as that of the main campus appeared more stable than the other two campuses.

PC parameterisation: A distribution of the data set with Sample×Feature (S×F) dimension for the three campuses

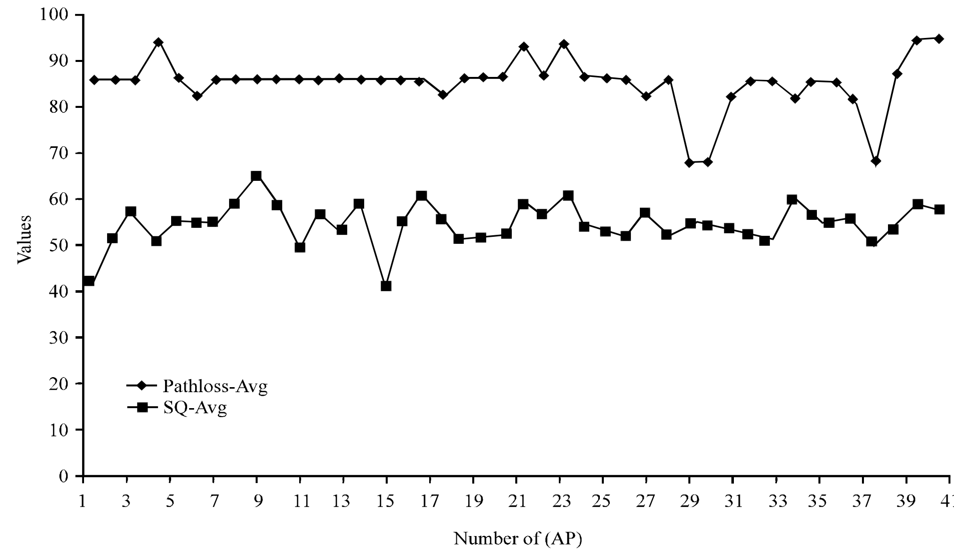

Fig. 4: Plot of average pathloss and service quality at the town campus

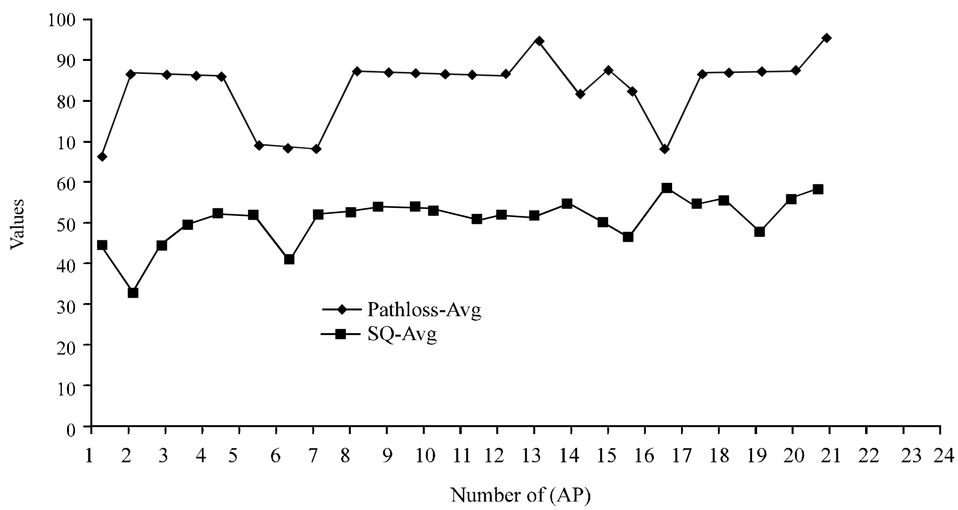

Fig. 5: Plot of average pathloss and service quality at the annex campus

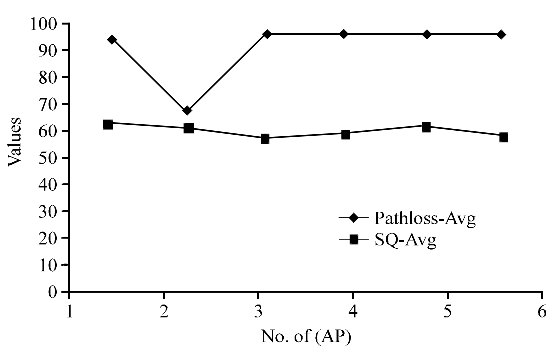

Fig. 6: Plot of average pathloss and service quality at the main campus

Table 3: Distribution of dataset for PCA parameterisation

Variables	Town campus		Annex campus		Main campus	
Category	S×F	Type	S×F	Type	S×F	Type
Inputs	2067×9	Double	925×9	Double	291×9	Double

and used for the PC parameterisation is shown in Table 3. Table 3, the town campus has the highest data points, followed by the annex campus and then the main campus. Also, the data points were populated from direct measurements obtained from the nine location features selected for the study. In Table 4 a PC parameterisation showing the first three PCs obtained from the town campus data for the selected location features is presented. Observe that only building size was most

Table 4: PCA parameterisation of town campus data

	Principal components		
Indoor/Outdoor parameters	1	2	3
Building type	-0.3567	-0.5740	-0.5856
Building size	2.6650	-0.0933	-0.0098
Purpose	-0.3549	-0.5435	-0.5287
Height	-0.3539	-0.5462	-0.4771
Distance from NOC	-0.2486	2.5199	-0.8149
Floor	-0.3570	-0.5731	-0.5948
Number of rooms	-0.3512	-0.5290	-0.4356
Pathloss	-0.3134	0.3314	1.7627
Service quality	-0.3293	0.0077	1.6838

Table 5: PCA parameterisation of town annex campus data

	Principal components		
Indoor/Outdoor parameters	1	2	3
Building type	-0.3539	-0.5936	0.5852
Building size	2.6658	-0.0621	0.0232
Purpose	-0.3525	-0.5607	0.4969
Height	-0.3508	-0.5694	0.4959
Distance from NOC	-0.2846	2.4564	0.9803
Floor	-0.3542	-0.5902	0.5864
Number of rooms	-0.3436	-0.6250	0.2024
Pathloss	-0.3038	0.4773	-1.9489
Service quality	-0.3225	0.0672	-1.4213

Table 6: PCA parameterisation of main campus data

	Principal components		
Indoor/Outdoor parameters	1	2	3
Building type	-0.3390	-0.5716	-0.7573
Building size	2.6665	-0.0263	0.0018
Purpose	-0.3387	-0.5353	-0.6160
Height	-0.3384	-0.5002	-0.5701
Distance from NOC	-0.3075	2.5542	-0.6891
Floor	-0.3391	-0.5814	-0.7751
Number of rooms	-0.3353	-0.4814	0.1313
Pathloss	-0.3332	0.1975	2.0031
Service quality	-0.3354	-0.0555	1.2715

dominant (i.e., had Eigen value of at least unity) for the first principal component while distance from NOC and (pathloss and service quality) were least dominant for the second and third principal components, respectively. These features captured the least variances when compared with the first PC. A PC parameterisation showing the first three PCs obtained from the annex campus data for the selected location features is presented in Table 5. From the table, all the location features showed no major differences, except building size distance from NOC and (pathloss and service quality) which showed significant differences for the first and second PCs, respectively. We observe a similar trend for the first two PCs in the PC parameterisation of the main campus data but with pathloss and service quality contributing the least variance for the third PC (Table 6). Hence, at the three campuses, building size (first principal component) appeared to capture the most variance while the remaining (or least) variances were captured by the second and third PCs, i.e., distance from NOC and

Table 7: PCA parameterisation of all campuses data

	Principal components		
Indoor/outdoor parameters	1	2	3
Building type	-0.3390	-0.5716	-0.7573
Building size	2.6665	-0.0263	0.0018
Purpose	-0.3387	-0.5353	-0.6160
Height	-0.3384	-0.5002	-0.5701
Distance from NOC	-0.3075	2.5542	-0.6891
Floor	-0.3391	-0.5814	-0.7751
Number of rooms	-0.3353	-0.4814	0.1313
Pathloss	-0.3332	0.1975	2.0031
Service quality	-0.3354	-0.0555	1.2715

Table 8: Classification results of selected classifiers

	Overall classification accuracy (%)			
Classifier	Town campus	Annex campus	Main campus	All campuses
Fine Gaussian SVM	88.10	70.80	66.70	83.30
Coarse Gaussian SVM	88.10	70.80	66.70	81.90
Quadratic SVM	90.50	79.20	33.30	88.90
Linear SVM	92.90	91.70	50.00	93.10
Fine k-NN	83.30	79.20	50.00	87.50
Coarse k-NN	88.10	70.80	66.70	87.60
Cosine k-NN	88.10	79.20	66.70	86.10
Weighted KNN	92.90	75.00	50.00	87.50
Simple decision tree	97.60	95.80	66.70	95.80
Complex decision tree	95.80	97.60	95.80	66.70
Medium decision tree	97.60	95.80	66.70	95.80
ANFIS	80.48	77.76	83.42	79.58

(pathloss and service quality), respectively. A PC parameterisation of all the data points (i.e., all the three campuses data) is presented in Table 7. We observe that building size showed major difference for the first principal component, distance from NOC and (pathloss and service quality) showed the least variances for the second and third principal components, respectively. We deduce here that the high number of data at the town campus may have influenced the overall dataset as the third PC in the annex campus showed no significant variance when parameterised alone (Fig. 5).

PC classification: In Table 8 we present the results of the selected classifiers examined in this study. In the table, columns with best classification accuracies are shaded with olive colour while worse classifiers are shaded with red colour. A graph plotting the classification accuracy vs. the classifier types is presented in Fig. 7. At the town campus, the decision tree classifiers gave the highest classification accuracy of 97.60%, followed by linear SVM and weighted k-NN classifiers which gave same classification accuracy of 92.90% and then the quadratic SVM which produced a classification accuracy of 90.50%. The least classification accuracy came from ANFIS (80.48%). The result implies that for areas with poor terrain, interference-prone condition occasioned by

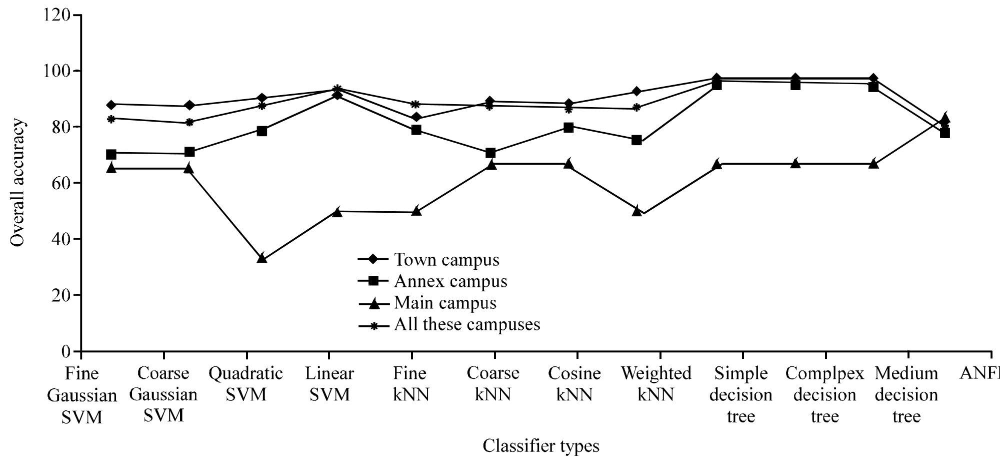

Fig. 7: Graph of overall classification accuracy vs. classifier type

congested structures and massive mobile users and more, decision tree, linear SVM and weighted k-NN classifiers are excellent at modelling WLAN FP patterns as impeding challenges such as interference between mobile and fixed infrastructure may have led to the wrong classification of related FP patterns. As regards the annex campus, decision tree classifiers and linear SVM classifier produced best classification accuracies of 95.80 and 91.70, respectively while all the other classifiers performed below 80% accuracy level with Coarse k-NN, Fine Gaussian and Coarse Gaussian SVM classifiers yielding the least classification accuracy of 70.80%. This result indicates that for areas with poor connections and spatial (uneven) distribution of WLAN infrastructure, decision tree and linear SVM classifiers can still maintain improved performance in terms of classification accuracy, compared to other classifiers which showed very poor performance. Concerning the main campus, the ANFIS classifier show promise of predicting well the FP patterns as it produced the highest classification accuracy of 83.42%, compared to other classification techniques which performed very poorly with the quadratic SVM producing the worse classification accuracy of 33.30%. Indeed, the result indicates that for areas with few population and WLAN infrastructure, ANFIS is best suitable for modelling the FP patterns. A combination of all campuses data reveals that decision tree classifiers gave the highest classification accuracy of 95.80%, followed by linear SVM which gave same classification accuracy of 93.10% while ANFIS gave the worse classification accuracy of 79.58%.

CONCLUSION

Location information of nodes in a wireless network is useful for a variety of reasons including, mobile nodes tracking within a network coverage area, periodic monitoring of the network coverage area, coverage area determination, support load and traffic management, node lifetime control, cluster formation and enhanced network routing. This reserach has provided a workable framework for fingerprint patterns parameterisation and classification, vital for indoor WLAN localisation. The parameterisation process offers useful insights into the degree of variances exhibited by dominant location features while the classification process predicts the fingerprints in a more robust manner. In the future, we shall exploit the use of sensor technology for information collection and signal identification, to provide real-time localisation, prompt service quality delivery and timely fault tracing and fixing in wireless communication environments which include academic environments.

ACKNOWLEDGEMENT

This research work is funded by the Tertiary Education Trust Fund (TETFund) National Research Fund (NRF) grant (Ref. No. TETFUND/NRF/UNI/UYO/STI/VOL.I/BE).

REFERENCES

Almuzaini, K.K. and T.A. Gulliver, 2012. Localization in wireless networks using decision trees and K-means clustering. Proceedings of the 2012 IEEE Conference on Vehicular Technology (VTC Fall'02), September 3-6, 2012, IEEE, Quebec, Canada, ISBN:978-1-4673-1880-8, pp: 1-5.

Azizyan, M., I. Constandache and R.R. Choudhury, 2009. SurroundSense: Mobile phone localization via. ambience fingerprinting. Proceedings of the 15th Annual International Conference on Mobile Computing and Networking, September 20-25, 2009, ACM, Beijing, China, ISBN:978-1-60558-702-8, pp: 261-272.

Bahl, P. and V.N. Padmanabhan, 2000. RADAR: An in-building RF-based user location and tracking system. Comput. Comm. Soc., 2: 775-784.

Dash, M., H. Liu and J. Yao, 1997. Dimensionality reduction of unsupervised data. Proceedings of the 9th IEEE International Conference on Tools with Artificial Intelligence, November 3-8, 1997, IEEE, Newport Beach, California, USA., pp: 532-539.

Deng, Z.A., Y.B. Xu and L. Ma, 2012. Indoor positioning via nonlinear discriminative feature extraction in wireless local area network. Comput. Commun., 35: 738-747.

Fang, S.H. and T. Lin, 2012. Principal component localization in indoor WLAN environments. IEEE. Trans. Mobile Comput., 11: 100-110.

Fang, S.H., T.N. Lin and K.C. Lee, 2008. A novel algorithm for multipath fingerprinting in indoor WLAN environments. IEEE. Trans. Wirel. Commun., 7: 3579-3588.

Farid, Z., R. Nordin and M. Ismail, 2013. Recent advances in wireless indoor localization techniques and system. J. Comput. Networks Commun., 2013: 1-10.

Gu, Y., A. Lo and I.G. Niemegeers, 2009. A survey of indoor positioning systems for wireless personal networks. IEEE Commun. Surv. Tutorials, 11: 13-32.

Han, D., S. Jung, M. Lee and G. Yoon, 2014. Building a practical Wi-Fi-based indoor navigation system. IEEE. Pervasive Comput., 13: 72-79.

He, S. and S.H.G. Chan, 2016. Wi-Fi fingerprint-based indoor positioning: Recent advances and comparisons. IEEE. Commun. Surv. Tutorials, 18: 466-490.

Hou, S., R.C. Qiu, Z. Chen and Z. Hu, 2011. SVM and dimensionality reduction in cognitive radio with experimental validation. Networking Internet Archit., 2011: 1-9.

Kupershtein, E., M. Wax and I. Cohen, 2013. Single-site emitter localization via multipath fingerprinting. IEEE. Trans. Signal Processing, 16: 10-21.

Kushki, A., K.N. Plataniotis and A.N. Venetsanopoulos, 2006. Location tracking in wireless local area networks with adaptive radio maps. Proceedings of the 2006 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'06) Vol. 5, May 14-19, 2006, IEEE, Toulouse, France, ISBN: 1-4244-0469-X, pp: V-741-V-744.

Li, D., L. Wang and S. Wu, 2014. Indoor positioning system using WIFI fingerprint. J. Inf. Technol., 2014: 1-5.

Longi, K., T. Pulkkinen and A. Klami, 2017. Semi-supervised convolutional neural networks for identifying Wi-Fi interference sources. Proceedings of the 9th Asian Conference on Machine Learning, November 15-17, 2017, PMLR, Florida, USA., pp: 391-406.

Lu, W.C., Y.C. Cheng and S.H. Fang, 2016. A study of singular value decomposition for wireless LAN location fingerprinting. Proceedings of the 2016 IEEE 2nd International Conference on Multimedia Big Data (BigMM'16), April 20-22, 2016, IEEE, Taipei, Taiwan, ISBN: 978-1-5090-2180-2, pp: 466-470.

Luo, J. and L. Fu, 2017. A smartphone indoor localization algorithm based on WLAN location fingerprinting with feature extraction and clustering. Sens., 17: 1339-1339.

Luo, Y., S. Xiong and S. Wang, 2008. A PCA based unsupervised feature selection algorithm. Proceedings of the 2nd International Conference on Genetic and Evolutionary Computing (WGEC'08), September 25-26, 2008, IEEE, Hubei, China, ISBN: 978-0-7695-3334-6, pp: 299-302.

Mirowski, P., D. Milioris, P. Whiting and T.K. Ho, 2014. Probabilistic radio-frequency fingerprinting and localization on the run. Bell Labs Tech. J., 18: 111-133.

Narayanan, S.J., R.B. Bhatt and I. Paramasivam, 2016. User localisation using wireless signal strength-an application for pattern classification using fuzzy decision tree. Intl. J. Internet Protoc. Technol., 9: 138-150.

Nuno-Barrau, G. and J.M. Paez-Borrallo, 2006. A new location estimation system for wireless networks based on linear discriminant functions and hidden Markov models. EURASIP. J., Adv. Sig. Process., 2006: 068154-068160.

Roy, R.A., J.P. Nair and E. Sherly, 2015. Decision tree based data classification for marine wireless communication. Proceedings of the 2015 International Conference on Computing and Network Communications (CoCoNet'15), December 16-19, 2015, IEEE, Trivandrum, India, ISBN: 978-1-4673- 7209-1, pp: 633-638.

Song, C. and J. Wang, 2017. WLAN fingerprint indoor positioning strategy based on implicit crowdsourcing and semi-supervised learning. ISPRS. Intl. J. Geo Inf., 6: 1-7.

Suparta, W. and K.M. Alhasa, 2016. Adaptive Neuro-Fuzzy Interference System. In: Modeling of Tropospheric Delays using ANFIS, Suparta, W. and K.M. Alhasa (Eds.). Springer, Cham, Switzerland, ISBN:978-3-319-28435-4, pp: 5-18.

Tan, P., M. Steinbach and V. Kumar, 2006. Introduction to Data Mining. 2nd Edn., Posts and Telecom Press, Beijing, China,.

Wu, C.L., L.C. Fu and F.L. Lian, 2004. WLAN location determination in e-home via support vector classification. Proceedings of the 2004 IEEE International Conference on Networking, Sensing and Control Vol. 2, March 21-23, 2004, IEEE, Taipei, Taiwan, pp: 1026-1031.

Xiang, Z., S. Song, J. Chen, H. Wang and J. Huang *et al*., 2004. A wireless LAN-based indoor positioning technology. IBM. J. Res. Dev., 48: 617-626.

Yim, J., 2008. Introducing a decision tree-based indoor positioning technique. Expert Syst. Appl., 34: 1296-1302.