

Article

Device-Free Crowd Counting Using Multi-Link Wi-Fi CSI Descriptors in Doppler Spectrum

Ramon F. Brena ^{1,*}, Edgar Escudero ^{1,2}, Cesar Vargas-Rosales ¹, Carlos E. Galvan-Tejada ³
and David Munoz ¹

¹ Tecnológico de Monterrey, School of Engineering and Sciences, Av. Eugenio Garza Sada 2501 Sur, Monterrey 64849, Nuevo León, Mexico; A00777396@itesm.mx (E.E.); cvargas@tec.mx (C.V.-R.); dmunoz@itesm.mx (D.M.)

² Aerobit Technologies, Av. Eugenio Garza Sada 3820, Monterrey 64780, Nuevo León, Mexico

³ Unidad Académica de Ingeniería Eléctrica y Comunicaciones, Universidad Autónoma de Zacatecas, Jardín Juárez 147, Zacatecas Centro 98000, Zacatecas, Mexico; ericgalvan@uaz.edu.mx

* Correspondence: ramon.brena@tec.mx

Abstract: Measuring the quantity of people in a given space has many applications, ranging from marketing to safety. A family of novel approaches to measuring crowd size relies on inexpensive Wi-Fi equipment, taking advantage of the fact that Wi-Fi signals get distorted by people's presence, so by identifying these distortion patterns, we can estimate the number of people in such a given space. In this work, we refine methods that leverage Channel State Information (CSI), which is used to train a classifier that estimates the number of people placed between a Wi-Fi transmitter and a receiver, and we show that the available multi-link information allows us to achieve substantially better results than state-of-the-art single link or averaging approaches, that is, those that take the average of the information of all channels instead of taking them individually. We show experimentally how the addition of each of the multiple links information helps to improve the accuracy of the prediction from 44% with one link to 99% with 6 links.

Keywords: Wi-Fi; CSI; crowd counting; Doppler spectrum



Citation: Brena, R.F.; Escudero, E.; Vargas-Rosales, C.; Galvan-Tejada, C.E.; Munoz, D. Device-Free Crowd Counting Using Multi-Link Wi-Fi CSI Descriptors in Doppler Spectrum. *Electronics* **2021**, *10*, 315. <https://doi.org/10.3390/electronics10030315>

Academic Editor: Matjaz Gams
Received: 19 December 2020
Accepted: 22 January 2021
Published: 29 January 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In recent years, mainly due to the COVID-19 health crisis in 2020 and beyond, the importance of technology capable of providing assistance to assess safety in crowds [1–4] has been brought to mainstream awareness [5]. However, crowd assessment applications are not limited to those that provide support for safety, and a new set of applications have been envisioned in businesses [6,7], and in other practical scenarios [8,9]. Of particular interest to the scientific community is the passive and device-free (meaning that the people who are monitored do not need to carry a device such as a cellular phone) estimation of the number of people in a given area. It is important to know the number of people in a room, to monitor human queues or to track the volume of customers in a commercial location, to provide valuable information in the context of smart space design, consumer marketing and venue security [10–12].

Though some recent and some decades-old developments have used computer vision for crowd measurement [1,13,14], nowadays visible light sensors are used with limitations due to the need of a line-of-sight which is subject to variable lighting conditions and coverage, as well as privacy concerns.

Recently, the increasing availability and descending costs of Wi-Fi equipment has promoted its use even in applications other than digital communications, such as indoor location [15,16]. In recent years it has begun to be the case of crowd measurement, given that popular Machine Learning techniques [17] can be used to recognize the disturbance patterns that human bodies produce when placed between a Wi-Fi transmitter and a

receiver [10]. Notwithstanding the wide range of potential applications that Wi-Fi sensing crowd analysis may reach, there are fundamental aspects of the subject (such as accuracy, reproducibility, and scale) that remain as limitations to overcome the boundaries of the current body of knowledge.

The aim of this work is to improve the results obtained from Machine Learning analysis of the disturbance patterns produced by human bodies to the signal propagation of individual channels of a Wi-Fi connection using the Doppler spectrum experienced in a crowd [18]. The original contribution of this paper is the systematical use of the Channel State Information (CSI) [19] of all the available channel links of a Wi-Fi communication (we refer to this as the 'multi-link' approach) rather than using indicators of just one link and discard the rest or to apply summary operations on all the links to reduce them to a single value (we refer to this as the 'single-link' approach, which is the one that has been used so far) to count the people present in one room, using a classification technique based on supervised Machine Learning. We demonstrate in this paper that the use of our multi-link approach improves accuracy in a dynamic environment with multiple wireless signals, multi-path components in the signal propagation through the channel that cause fading and absorption.

Many of recent works on this subject have used the Received Signal Strength Indicator (RSSI) as an index of the channel quality. RSSI is processed for feature extraction and the counting estimation is obtained after carrying out a learning phase [20–24]. For the estimation of the exact number of people in a place, the best RSSI-based reported results come from the work of Yoshida et al. [24], with a 77% of accuracy for up to 7 people. A major drawback of this technique is that RSSI-based algorithms tend to ignore the multi-path effects of the RF propagation, and as a matter of fact, its performance is greatly affected by channel disturbances. A more recent approach uses the Channel State Information (CSI) [19] that provides channel response information for Multiple Input Multiple Output (MIMO) Wi-Fi systems. As a result, CSI offers better measurements of people activity by capturing the disturbances the crowd cause in the channel.

Even with the use of CSI-based techniques, the performance of crowd-counting solutions documented in the literature present accuracy challenges that typically worsen with the group scale. The research carried out by Di Domenico et al. [25], which is commonly used as an indicator of the state of the art, reports an accuracy of about 80% for counting up to 7 people. The work from Xi et al. [26], is another common reference in the field, they achieved a probability of 80% of having either a perfect count or failing by one person when counting up to 30 people.

In this paper, we present a data driven work that takes advantage of the advances in Machine Learning (ML) techniques and apply it to multi-link Wi-Fi CSI information to produce better results than those reported in the literature for the recognition of the characteristics of a crowd, and specifically its counting.

The proposed method takes the information extracted from a CSI pattern of commercial off-the-shelf (COTS) Wi-Fi and translates it into the Doppler Spectrum where a set of features that capture information provided by the multi-link nature of the MIMO system is extracted to achieve high accuracy counting predictions. Furthermore, our approach can be potentially useful to identify dynamic characteristics of the crowd, such as its size, growth, dispersion and mobility, that could be applied to many relevant scenarios such as offering services based on the occupation detected in an environment, trends of influx in public spaces, occupancy predictions, mobility trends by region, and many more.

The method here presented reports several advantages with respect to other works such as:

- Fewer features derived from the signal are required for an accurate counting estimation. This results in reduced processing time since feature extraction requires less computing power.
- It works seamlessly with COTS Wi-Fi access points.

- Increased accuracy and other performance metrics as a result of using multiple links instead of one or an average.

The remaining of this paper is structured as follows. In Section 2 we present the background concepts for the sake of self-containment, as well as a review of the related work. Then, our method is presented in Section 3, together with the experimental setup and results description. Finally, in Section 4 we discuss the relevance of our contributions and provide some ideas for future work.

2. Background and Related Work

The field of crowd dynamics refers to the analysis of the motion of people within a defined group and its changes over a period of time. The topic has attracted an increasing interest from the research community due to many potential applications, and more recently the COVID-19 health crisis under way, makes clear that it is imperative to avoid crowd concentrations, especially in indoor spaces, in order to avoid further contagion [27]. Crowd applications are not limited to health issues though, and among other ones, we find crowd security and management for emergency handling, where the ability to recognize patterns in the crowd behavior allows better and faster responses or improvements to the space design [3,28,29]. Hence, several frameworks coming from different disciplines have been proposed in order to model the motion of a crowd.

The work of Helbing and Johansson [30], explores the analogies between the patterns of a crowd and the properties of a fluid of particles. Their study provides a framework to model the interactions of individuals in crowds, and the study of self-organized patterns of motion they generate as a result of the emerged collective intelligence and the social forces involved in the process. In a prior work, Helbing et al. [31], introduced the concepts of attraction and repulsion forces to simulate the dynamics of crowds in panic or evacuation situations.

In the following sections, it is also discussed how different authors put different levels of emphasis in one or more attributes of the crowd in order to model, describe and predict its dynamics; and each of them uses a set of metrics, either quantitative or categorical, as a basis for their work.

2.1. Quantitative Characterization of a Crowd

Still [32] defines crowd speed as ‘the emergent speed of a group of individuals’ that is a result of the non-linear interactions within the crowd in the local geometry. In his study, the author describes how crowd speed is modulated by crowd density (number of people by unit of area), being the flow volume a function of both.

Helbing et al. [33], analyze the relationships among numerical properties of crowds to describe a motion model. In this study, Helbing defines key crowd parameters such as density, speed, pressure and flow vectors. The research argues that even at highly dense crowds the motion of the crowd continues, which in turn causes dangerous ‘turbulence’ spots where crowd pressure is beyond a critical threshold.

The work of Pathan et al. introduces a novel approach for crowd behavior analysis and anomaly detection in coherent and incoherent crowded scenes [34]. The authors explore the crowd problem from a data-driven perspective and propose a method to calculate social entropy. The introduced metric is used as a descriptor of crowd behavior. Support Vector Machines are used to train and classify the flow feature vectors as normal and abnormal.

2.2. Categorical Metrics of Crowd Dynamics

Vicsek et al. [35], provide a general classification of collective motion. They proposed that any group of individuals can be categorized in five possible motion states: (i) disordered state (individuals moving randomly); (ii) fully ordered state (individuals moving at pace in the same direction); (iii) rotational (individuals moving in well-defined patterns); (iv) critical (state very sensitive to perturbations); (v) velocity correlated (individuals

behave as elongated particles). Saleem et al. proposed a simplified approach to these categories by grouping them into coherent and incoherent crowds [34].

Interesting to our own work are crowd analysis studies that have been proposed in the image processing and computer vision fields. In this context, Xu et al. proposed an algorithm to detect the gathering and dispersing stages of a crowd in video recordings. To achieve this, the authors combine techniques of crowd counting and group entropy to estimate the spatial distribution of the individuals. These parameters are then used as features for the classification process [36].

2.3. Sensing Crowds with Wi-Fi

Human sensing based on commodity Wi-Fi has gained attention in recent years mainly because of the pervasive availability of Wi-Fi signals, that can be re-purposed sometimes with little cost. Advances in device-free sensing, where neither special equipment nor cooperation is required from the sensed subject, have been documented [37–39].

Amplitude and phase of Wi-Fi signals are very sensitive to the surrounded environment, and it has been shown that it is possible to extract patterns from such variations to identify human-related activity [23,40–42]. This way of acquiring human or crowd data is known as passive sensing. It refers to the family of sensing techniques that requires no cooperation from the sensed subject (as opposite to active sensing that is based, for example, in wearables or mobile apps used by the sensing target).

There are in the literature two different approaches to passive sensing with Wi-Fi signals: the first one is *passive device-present*, here the sensing signals from client-to-Access Point (AP) communication are exploited; the second one is *Passive device-free*, where sensing signals from AP-to-receiver communication are used. Passive device-free sensing based on commodity Wi-Fi has gained special attention due to the advantages of achieving crowd sensing without requiring the collaboration of the sensed group. Also, in contrast to device-present sensing, a device-free approach protects subjects privacy inherently.

2.4. Human-Centric vs. Crowd-Centric Sensing

Research work may also be classified according to the type of inputs it tries to detect: while human-centric aims to detect events and activities at the person level, crowd-centric aims to modeling properties of a group of people. Although a plethora of applications have been subject of study, in practical terms, sensing research may be binary classified in one of these two approaches.

As opposite to the human-centric perspective, crowd-centric sensing is not interested on predicting or estimating properties from single individuals, but those that arise from a group of people as an entity. Models in this category are designed to define crowd variables such as the size, density, speed and direction.

While human-centric sensing has found applications especially in health care [40,43–45], crowd-centric sensing has an important range of potential applications in public safety, mobility planning and marketing [6,46–50].

2.5. Sensing Crowd Properties

As discussed above in this section, crowds present several properties of interest that can be useful to prevent unwanted situations or stimulate desired behaviors. In this context, research work on Wi-Fi-based sensing models that can deliver accurate results is gaining attraction.

Wi-Fi-based crowd counting is the task of estimating the number of people gathered in a specific area. An increasing amount of research work has been dedicated to crowd counting with Wi-Fi signals in recent years. Xi et al. published a method to estimate the size of a crowd of up to 30 people using a Grey Verhulst model [26]. Another important crowd indicator is its density (i.e., the number of people per unit of area). While there is few documented research on this subject, a recent work by Tang et al. uses a device-present approach to calculate crowd density by capturing device's request probes and RSSI

signal [51]. Depatla et al. [52] proposed a framework to sense speed of a crowd in both outdoor and indoor locations.

2.6. Rssi and Csi: The Sensing Signals

Wi-Fi is a commercial denomination for the IEEE 802.11 standard, which is the predominant technology used for Wireless Local Area Networks (WLAN) that operates in the 2.4 GHz or 5 GHz frequency bands. The first version of the standard was released in 1997 [53]. A decade later, the idea of using Wi-Fi signals to identify, analyze or predict human activity started to be increasingly present in researchers' work desks around the world. Since then, several techniques have been proposed for capturing, denoising, processing and classifying the hints of human activity that are intrinsically carried by the wireless signals the Wi-Fi standard uses.

Estimation techniques are in general based on some kind of measurement of signal parameters on the RF propagation channel as a function of the variables of interest. There are two specific types of signal-related parameters that are commonly used in current Wi-Fi sensing research, which are: RSSI and Channel State Information (CSI).

2.6.1. Rssi

Due to the simplicity of field strength measuring, RSSI has been commonly used in many applications as those concerned to indoor localization. As its name describes, RSSI is a measure of the power present in the signal at the time it arrives to the receiver. According to signal propagation models and measurements, signal energy decreases with distance. Because of this, RSSI is often used with multilateration methods to estimate position [15,16]. This is a device-assisted approach, as it requires that the subject to be localized carry a device with a Wi-Fi receiver. Human tracking with Wi-Fi can, however, also be accomplished by device-free methods. This is achieved by analyzing the variance and other statistical properties of the RSSI signal data [54,55]. Youssef et al. for example, used the moving average of the signal strength values to track the location of a person [56].

Although RSSI can be easily implemented, its suitability as sensing vehicle in non-controlled environments is limited, as RSSI presents several impairments when obstacles are present in the area of interest. The strength of a single received signal is greatly affected by multipath and shadowing effects that yield estimation errors. As a result of this, CSI has been proposed as an alternative [57].

2.6.2. Csi

Modern Wi-Fi standards, powered with Orthogonal Frequency Division Multiplexing (OFDM) modulation and MIMO capabilities, utilize CSI as an indicator of the properties of the channel to dynamically optimize and adapt the transmission parameters to improve performance. Therefore, CSI is an overall representation of the channel state that sums up the signal's propagation effects including scattering, fading and multipath [58]. Although RSSI also offers an overall picture of the channel by providing time averaged total power of the signal envelope at the receiver, CSI is an estimation of the channel coefficients that represent either the impulse or the frequency response at the sub-carrier level [59]. Because of its granularity of sub-carrier frequencies and its vector representation, CSI data provide more information of the channel impairment effects that the signal experiences in contrast to the received power strength given by RSSI.

The proposed research work makes use of the advantages mentioned above and employ CSI as the sensing signal. Hence, it is worth providing a more detailed description of the nature of the CSI data and the context in which it is produced.

2.6.3. Mimo Systems

In wireless communications the information signals are transmitted through a channel, ideally in line-of-sight (LOS) obstacle-free conditions, but in practical scenarios, the LOS condition is not met and the transmission paths are not unique, so the beam uniqueness

does not hold. The objects and subjects located in the surroundings of the channel may reflect, refract, diffract and scatter the signal, producing multiple new paths the signal traverse to the receiver, an effect called multipath propagation that exhibits deep fades with high variance [60].

From the receiver point of view, multiple copies of the signal arrive with different time delays, amplitudes and phase shifts. This aggregation of signal replicas produces fading, a multipath-induced interference that results in variations of Signal-to-Noise Ratio (SNR) of the emerging signal [61]. Such a behavior is traditionally considered as a problem the wireless communication system has to solve.

A Multiple Input Multiple Output (MIMO) system uses multiple antennas at the transmitter or the receiver to improve the overall communication performance. In contrast to single-antenna systems (SISO) that gather a single signal for reception, MIMO systems are able to use simultaneously multiple signals carrying information. Then, signal processing techniques such as space-time coding, beamforming, channel estimation and symbol detection are used to take advantage of the various signals that are available at the receiver. This results in significant improvements of spectral efficiency, data rate and system capacity [62].

A typical configuration of a MIMO channel consisting of M number of antennas at the transmitter and N antennas at the receiver is shown in Figure 1. The transmission is expressed in terms of the received vector \mathbf{y} , the transmitted vector \mathbf{x} , the channel coefficient matrix \mathbf{H} and noise \mathbf{n} , as follows:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \quad (1)$$

Similarly, each of the N antennas at the receiver, receives signals from all the M transmitting antennas, creating S number of communications links according to:

$$S = MN \quad (2)$$

The general expression for (1) in its matrix representation is

$$\begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_N \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & \dots & h_{1M} \\ h_{21} & h_{22} & \dots & h_{2M} \\ \dots & \dots & \dots & \dots \\ h_{N1} & h_{N2} & \dots & h_{NM} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_M \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \\ \dots \\ n_N \end{bmatrix} \quad (3)$$

For example, a 3×2 MIMO system will have 3 transmitting antennas and 2 receiving antennas, conforming a total of 6 communications links. Each link's channel coefficient $h_{m,n}$ represents the channel effects that the signal that travels to the m -th receiving antenna from the n -th transmitting antenna undergoes. Thus, for the 3×2 example, the received vector is represented as follows:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \\ n_3 \end{bmatrix} \quad (4)$$

Notice that, as previously mentioned, a CSI packet is a complex number representing the amplitude and phase of the channel state; so in MIMO-enabled Wi-Fi, CSI signals provide the estimated values of the channel coefficient matrix \mathbf{H} .

2.6.4. Ofdm Transmission

The IEEE 802.11 standard adopted Orthogonal Frequency Division Multiplexing, or OFDM, as part of its transmission technique to achieve higher data rates [62]. The idea behind this technology is to increase data rate using parallelism by dividing the assigned spectrum into several narrowband sub-carriers that are used for simultaneous transmission. The OFDM sub-carriers are orthogonal in the mathematical sense, which implies that sub-

carrier frequencies are selected to cancel out inter-symbol interference (ISI). Therefore, additionally to delivering higher rates, OFDM offers immunity to the ISI effect caused by multipath fading and it also requires relatively simple receivers to reconstruct the transmitted data, since signal processing is done using the Fast Fourier Transform (FFT) and the inverse FFT algorithms instead of hardware.

According to standard IEEE 802.11n, MIMO data are modulated into 52 sub-carriers using Inverse Fast Fourier Transformation (IFFT) and transmitted as OFDM symbols in discrete packets. The receiver measures the CSI for each packet and adapts parameters to channel variations, the received signal is then demodulated by applying the direct FFT. The tool provided by Halperin et al. [59], which was used to obtain our experimental data, allows the extraction of CSI information from an Intel 5300 wireless card, exposing channel information of 30 of the 52 Wi-Fi sub-carriers. The CSI for each sub-carrier is defined as

$$h = |h| e^{j\theta} \quad (5)$$

where $|h|$ and θ represent the magnitude and phase of the communication channel, respectively.

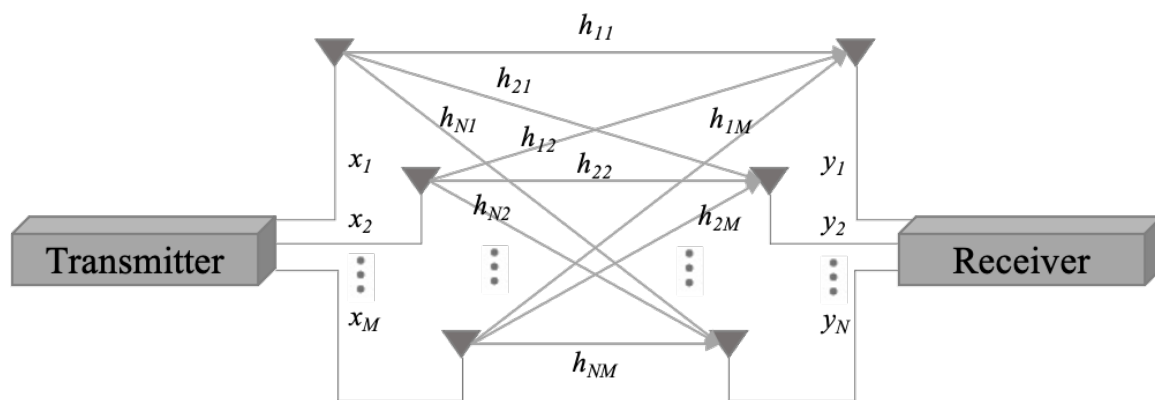


Figure 1. Multiple Input Multiple Output System.

2.7. Related Work

In this section we briefly review relevant literature that documents the current state-of-the-art of device-free Wi-Fi-based crowd counting. Each subsection corresponds to a published article in the field. In the last subsection we provide a summary table with the reported accuracy of each reviewed method for further reference and benchmark.

2.7.1. Trained-Once Device-Free Crowd Counting and Occupancy Estimation Using Wi-Fi: A Doppler Spectrum Based Approach

The work from Di Domenico et al. [25], performs Doppler Spectrum transformation to a CSI stream. The authors carried out a series of experiments where groups of people with several participants (from 0 to 8) were sensed in 3 different locations. Due to the scope of the experiments, the resultant dataset is of enormous value for the research community and it is the one we use for the preliminary work of the present research.

Di Domenico also introduces a long list of features that can be extracted from the Doppler spectrum matrices. Among all the possibilities the authors selected Spectral Kurtosis as the unique descriptor for their model. The performed series of arithmetical mean to the sub-carriers and links in order to get to a simpler parameter to work with.

Also, for the learning stage, the authors used a Naive Bayes classifier, because of its simplicity as a probabilistic algorithm. With this setup, they achieve about 80% of accuracy for crowd counting estimation. It is worth noticing that the main purpose of the work by Di Domenico et al. is to show the advantages of their method for a 'training once' scenario where there is no need for dedicated training in every different location.

2.7.2. Frog Eye: Counting Crowd Using Wi-Fi

The article from Xi et al. [26], is another often cited work in the Wi-Fi-based sensing field. In this research the authors documented a sensing framework based on CSI measurements from off-the-shelf Wi-Fi equipment.

Xi's model introduces a feature called Percentage of non-zero elements (PEM), which is a measurement technique based on the non-zero counting of the CSI dilated matrix. The resulting dataset is classified with the help of a grey Verlhust model factor.

To measure the performance of their method, the authors utilized the probability that an error equal or less than a defined threshold occurs for a particular counting estimation. This indicator was reported to be 98% for a threshold of 2 or fewer person and about 80% for a threshold error of 1 person.

2.7.3. Wicount: A Deep Learning Approach for Crowd Counting Using Wi-Fi Signals

The work from Liu et al. [63,64], explored the capabilities of a deep learning-based method for crowd counting. It uses a fully connected neuronal network with two hidden layers. It also implements regularization and exponential decay to improve performance. The experimental results show that the introduced deep learning model is able to estimate the number of crowd up to 5 with the accuracy of 82.3%.

The key contribution of the article for the Wi-Fi sensing field is that it documents a Deep Learning model that is arguably the first in its kind to be applied for crowd counting. Even if one can claim that the amount of time and computing resources required to train a DL system are still very demanding and the outcome quality does not correspond to the effort, the authors clearly pointed out a direction for future work. Similar works with data coming from Wi-Fi CSI information, which use Deep-Learning, like Cheng et al. [65], achieve slightly higher accuracy, with a reported 88.66%.

2.7.4. Occupancy Estimation Using Only Wi-Fi Power Measurements

The article from Depatla et al. [20] introduced, to the best of our knowledge, the most cited RSSI-based method for crowd counting. Their framework is based in a model that incorporates the pattern of both blocking LOS and scattering that human bodies produces in the strength of the Wi-Fi signals.

The authors approached the problem from an analytical perspective to obtain a mathematical expression that relates the signal strength with the PDF of the number of people in a crowd. Then, the method uses Kullback-Leiber divergence to estimate the size of a crowd with up to 9 people.

As in other works in the literature, Depatla presented its results as probabilities of getting errors of certain number of people. Specifically, their method reported $P(e \leq 1) = 55\%$ and $P(e \leq 2) = 63\%$ for indoor experiments with off-the-shelf equipment. For the outdoor scenario $P(e \leq 1)$ was 64%, while achieving 96% for a threshold of 2 or less people of counting error.

2.7.5. Estimating the Number of People Using Existing Wi-Fi Access Point in Indoor Environment

Yoshida et al. [24] published a relevant work where the counting estimation is made using regression algorithms. The method uses RSSI data as feature and test both linear regression and SVM regression to make the classifications.

The experiment setup consisted in a single transmitter and four receivers, each of them working as independent measuring point. A notable contribution of this research is that it explores, with this kind of layout, additional crowd characteristics such as density and presence/absence of people.

The accuracy rate of Yoshida's method for estimating the number of people is 77%, for estimating the degree of congestion (crowd density) is 95%, and for estimating the presence/absence of people is 98%. This work was updated by Mabuchi [66] to achieve 93% in counting smalls groups of people.

2.7.6. Freecount: Device-Free Crowd Counting with Commodity Wi-Fi

FreeCount [67] is a high-accuracy system for crowd counting that uses a set of features of three kinds: statistics, frequency domain and shape. In their publication, Zou et al. focus in the problem of temporal variation of the channel conditions and the unpractical need of re-training the classifier every period of time.

By using a model based on SVM, the author reported a crowd counting accuracy of 99%, and $P(e \leq 1) = 97\%$. Moreover, FreeCount implements transfer kernel learning (TKL) to cope with the changes in the channel condition with time. With TKL as the SVM kernel FreeCount reported an accuracy of 96% two weeks after the actual training took place.

A downside of the FreeCount approach is that it requires to modify the Access Points in the location. This means the solution is not “commercial off-the shelf”, in the sense that it can not work seamless with currently installed infrastructure as the rest of the methods reviewed above can. For this reason we considered it a non-COTS solution.

Table 1 shows a summary of the performance of crowd counting methods that use COTS through Wi-Fi technology

Table 1. Performance summary of crowd counting methods using commercial off-the-shelf (COTS) Wi-Fi.

Author	Data Source	Classifier	Accuracy			Scale	
			Accuracy Rate	$P(e < 1)$	$P(e < 2)$	Max # of People	Sensing Area (m ²)
Di Domenico	CSI	Naive Bayes	81%	-	91%	8	30, 45, 70
Xi	CSI	Grey Verhulst	-	80%	98%	30	-
Liu	CSI	Deep Learning	82%	-	-	5	-
Depatla	RSSI	Math. Exp.	-	55%	63%	9	33
Yoshida	RSSI	SVR	77%	-	-	7	-

2.8. Theoretical Framework of Crowd Characterization with Wi-Fi Csi in the Doppler Spectrum

In a real propagation environment, a signal propagates along multiple paths, and the receiver experiences multiple time-delayed replicas of the transmitted signal. Furthermore, if the receiver is moving, a set of Doppler shifts occurs in the receiving end and a Doppler spread spectrum arises. In a MIMO-OFDM transmission, random variations on the sub-carriers frequency causes uncorrelated fading between the different received paths. If a simple correlation receiver is applied to the received signal, delayed versions of the transmitted signal will not correlate properly and thus cause self-interference [47].

For multipath communication, the radio signals arrive at the receiver device as the sum of all the contributions produced by the scattering process. When the scatter objects are static with respect to the radio source, the radio frequencies do not change in the propagation channel. However, in the case of scatterers or source motion, there will be a Doppler shift that depends on the speed and moving direction with respect to the signal propagation path. At a single frequency level this phenomena is known as Doppler shift, but in a time-varying scenario (as the scatter objects change direction and speed over time) a set of Doppler shifts or Doppler spread is also referred to as “Doppler spectrum”.

The work of Yang et al. [68], provides a conceptual framework to analyze the Doppler spectrum of a Wi-Fi transmission using the CSI signals. We will briefly summarize Yang’s analytical model in the following lines for convenience. The scenario is illustrated in Figure 2 and is described as follows: if a transmitter is at a distance d from a moving receiver with velocity v at a given instant, then the Wi-Fi signal that will arrive at the receiver is affected by a channel that is multipath and time-varying. Let’s suppose there are a total of L independent paths l , each of them with different angles of arrival θ_l to the receiver’s moving direction.

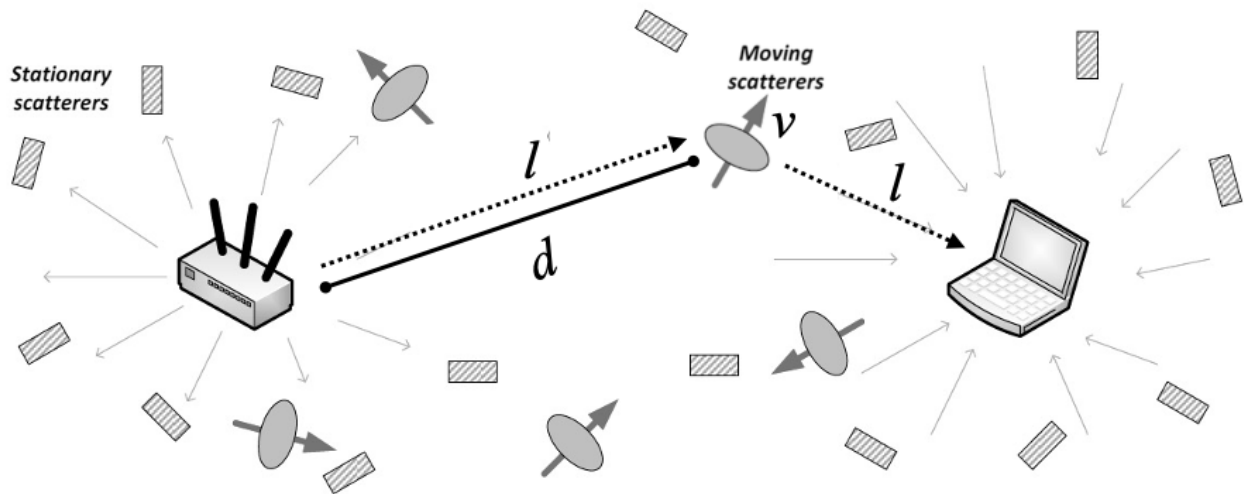


Figure 2. Doppler shifts produced by moving scatterers in a Wi-Fi link.

In MIMO-OFDM Wi-Fi there are k sub-carriers and multiple links that will be affected by the same multipath process, the channel response can be expressed as a function of the time instant nT_s

$$h_k(nT_s) = \sum_{l=1}^L \alpha_l e^{-j\omega_k \frac{d_l}{c}} e^{j\omega_k \frac{v_l}{c} nT_s} + \epsilon_k \tag{6}$$

where α_l is the amplitude of the l -th path, c is the propagation velocity of electromagnetic wave, ϵ_k is the measurement error, and $\omega_k \frac{v_l}{c}$ is the Doppler shift.

From the Wi-Fi IEEE802.11n standard [69], we know that the center frequency for each OFDM sub-carrier k in GHz is given by

$$f_k = 2.4 + k\Delta f \quad k = 1, 2, \dots, 30 \tag{7}$$

where Δf is the frequency difference between sub-carriers, and it is extremely small compared with the 2.4 GHz. The Doppler shift can be approximated as

$$\omega_l = 2\pi \frac{v_l}{c} \tag{8}$$

Hence, a good approximation of the channel response can be given by the average of the available individual channel responses of the subcarriers, as follows:

$$\bar{h}(nT_s) = \frac{1}{30} \sum_{i=1}^L \sum_{k=1}^{30} \alpha_i e^{-j\omega_k \frac{d_i}{c}} e^{j\omega_k nT_s} + \bar{\epsilon}_k \tag{9}$$

Next, we obtain the frequency domain channel response by using the discrete Fourier transform as follows:

$$\begin{aligned} \bar{H}(m) &= \sum_{n=1}^{N-1} \bar{h}(nT_s) e^{-j(\frac{2\pi}{N})nm} \\ &= \frac{1}{30} \sum_{i=1}^L \sum_{k=1}^{30} \alpha_i e^{-j\omega_k \frac{d_i}{c}} \frac{1}{30} \sum_{k=1}^{30} e^{j2\pi(f_i T_s - \frac{m}{N})n} + \bar{E}_k \end{aligned} \tag{10}$$

Equation (10) is an analytical representation for Doppler Spectrum of WiFi CSI. From this foundation, different authors take different approaches for crowd counting.

For instance, Zou et al. [67], use a set of features coming from statistics of magnitude and phase, Fourier transformation and shaped metrics, all combined to achieve predictors of human motion. Di Domenico et al. [25], extract Average Spectral Kurtosis as the unique feature they use for their model, while Yang et al. [68], use only the first link of the MIMO grid for feature extraction.

3. Proposed Method and Results

The method here presented is based on the hypothesis that the diversity in channel response information that multiple communication links of a MIMO system carry could provide better descriptors of the number of people in a crowd than a single channel or a channel average. While other authors average or discard the multi-link information [25,68], our method exploits such information to produce high-quality features for the estimation model.

The results presented in this paper follow a data-driven, quantitative approach. We show how the data can be processed to get estimations of the crowd characteristics with acceptable accuracy. The methodology followed in this paper includes:

- The use of a reference dataset (see Section 3). On using a public dataset instead of collecting our own experimental data, we give up the possibility of adding information we might find useful. However, there is the opportunity to directly compare our results to the ones obtained by other researchers.
- Implementation of our method in MATLAB, and the further training of the model with a set of learning algorithms.
- Performance assessment of our method in terms of accuracy and other results providing quality indices, and a comparison with state-of-the-art approaches by re-implementing them in order to have a reliable comparison.

A high-level view of our method is illustrated in Figure 3. The data collection (which we took from the public dataset of Di Domenico et al. referenced before) is on the dotted box to the left, and our process appears inside the right dotted box. Our data-driven method comprises the steps:

1. Doppler spectrum estimation: in this step we created a MATLAB script that transforms the available CSI data points to the spectrum domain in order to obtain Doppler spectrum data as described in Equation (10).
2. Feature extraction: From the CSI readings together with the Doppler spectrum estimated parameters, a vector of signal features is derived, which is supposed to be a good compact representation of the signal characteristics, at least for the classification purposes that we have, that is, the estimation of the number of people in the room. We used mainly statistical descriptors of the signal, which could be a good or bad idea, and we can only assess this later on, when we obtain the classification performance figures. The output of the feature extraction phase is a *dataset*, which is a table in which rows are individual observations and columns are the features. The dataset is the starting point of the Machine Learning process itself.
3. Train-test split: In order to assess the prediction quality of the trained Machine Learning classifier in a fair way, we need that the data used for testing its performance has not already been seen by the classifier; so we make a separation or *split* of the dataset into two subsets: the *training* dataset and the *testing* one. The relative size of each one is critical for a good performance, and this will be discussed below in the corresponding subsection.
4. Classifier training: Using the training part of the dataset, we adjust the parameters of a standard classifier (like Random Forest and others, described below). All of those classifiers are readily available in programming libraries, for the different platforms (MATLAB in our case), so the real work is not to construct the classifiers but to choose and configure them properly; whether or not it has been well done is only seen later, when classification performance results are obtained.

5. Prediction assessment: The already trained classifiers are used to obtain, for each row in the testing part of the dataset, a predicted class (in this case a number of people in the room). Once the prediction is done, its quality can be assessed by a number of well-known metrics such as accuracy, precision, recall and others, which will be discussed below when we present experimental results.

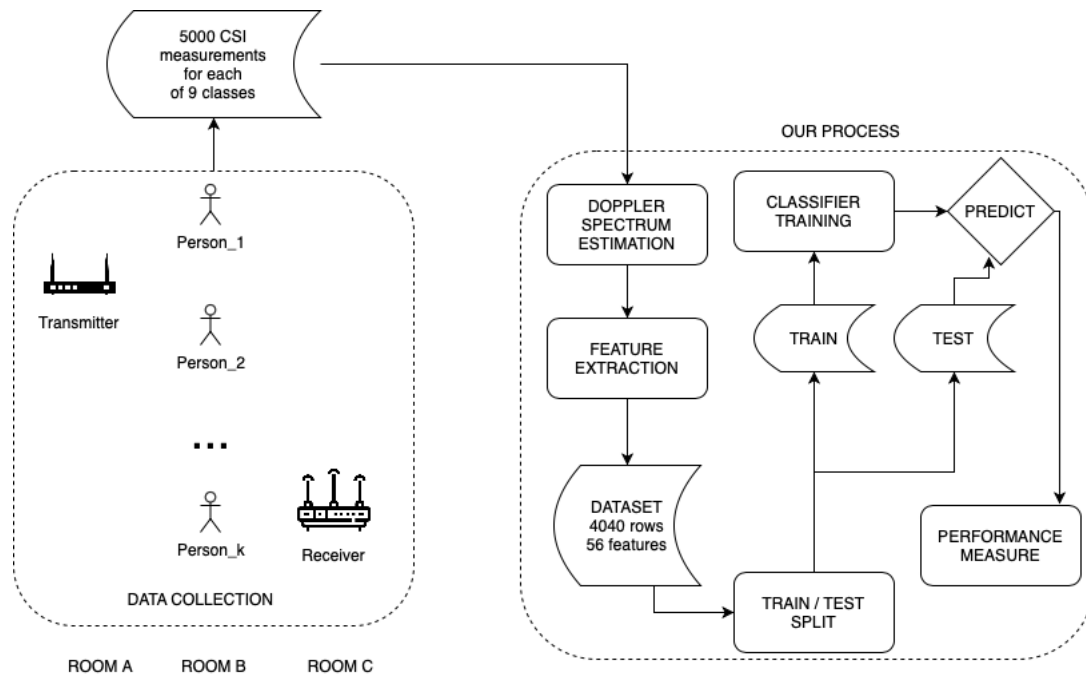


Figure 3. High-level diagram illustrating our method.

From the 4040-row features dataset to the classification assessment, everything is a mostly standard Machine Learning process (for which there are even free ML software platforms), so we do not claim to make any contribution there. Our contribution is the process that goes from the instrumentation readings, available from the Di Domenico et al. public data base [70] to the dataset construction from which the ML process is done, but of course the usefulness of the Doppler spectrum information, as well as the features we proposed can only be seen once the predictive power of the trained classifier is fully assessed.

3.1. Multi-Link Based Csi Crowd Counting Estimation

Our model calculates a set FS of p feature vectors F , each one of them being a combination function g_i of a descriptor $d_i \in D_i$ applied to each one of the T_{lk} available links \bar{H}_{lk} in the MIMO Wi-Fi transmission. From the set of feature vectors our goal is to estimate, for a given one not previously seen, the number of people in the room with an accuracy as high as possible.

This set of feature vectors can be represented as follows:

$$FS = \{F_0, F_1, \dots, F_p\}, \quad \text{where } F_i = g_i(D_i) \quad \text{for } i = 0, 1, 2, \dots, p \quad (11)$$

$$\text{and } D_i = \{d_i(\bar{H}_{lk_1}), d_i(\bar{H}_{lk_2}), \dots, d_i(\bar{H}_{lk_{T_{lk}}})\}$$

The exact way of defining the features for Machine Learning prediction is entirely domain-dependent, and some argue that it is as much an art as is a science. In this paper, we explore the prediction performance of the model when d_i are standard statistical measures.

3.2. Dataset

In order to experimentally test the performance of our multi-link method, we used a publicly available dataset. The dataset from Di Domenico et al. [19] provides great opportunities to explore and test our Wi-Fi-based crowd counting hypothesis in a data-driven way. This dataset consists of the following:

- A 2-antenna Wi-Fi transmitter (an off-the shelf AP) and a 3-antenna Wi-Fi receiver (a computer with an Intel 5300 NIC) are set in the experiment location.
- Groups from 1 to up to 8 people were sensed using the mentioned setup, additionally to the 'empty room' case.
- The volunteers are allowed to move freely, but the only meta-data being labeled is crowd counting.
- A CSI trace is extracted every 20 ms, with a round lasting about 2 min for every counting case.
- The whole experiment was repeated in 3 different locations, as follows: Room A is a small size office room (5 m × 6 m), Room B is a medium size meeting room (5 m × 9 m), and Room C is a large size meeting room (6 m × 12.5 m).

The Di Domenico's dataset includes at least 5000 CSI measurements for each counting class, for each type of Room. Also, each CSI trace consists of a channel response representation for each of the 6 resulting links, and every link includes 30 RF sub-carriers.

3.3. Machine Learning Process for Crowd Counting

For crowd counting estimation, we used standard Machine Learning classifiers so that each predicted number of people is considered as a class, so that the empty room is one class, 1 person in the room is another class and so on. Obviously in most future practical applications, classes would be numeric ranges, like 1-10 persons for one class, 11 to 50 for another one, etc.

The Machine Learning process follows the following steps:

- Step 1: Shuffle randomly the dataset rows.
- Step 2: Use feature selection criteria depending on the experiment variant.
- Step 3: Train the model with 5-folds cross-validation using the following classifiers:
 - Random Forest
 - Weighted KNN
 - Linear Discriminant
 - SVM
 - SVM with Gaussian Kernel
- Step 4: Test the model and report performance results (accuracy, AUC, etc.).

In the k -fold cross-validation process of step 3 we used a $k = 5$ instead of the more popular $k = 10$ because of the relative abundance of data, and the absence of improvements in more intensive computations resulting from increasing the k value.

3.4. Feature Extraction

A first selection of descriptor functions was made from a set of statistics commonly used in signal processing [71]. Our first objective was to test our multi-link model with all the descriptors listed in Table 2, and then to proceed to perform feature selection in order to reduce dimensionality. A more complete list of these kinds of features is provided by Di Domenico et al. [25].

Table 2. Features and descriptor functions used in the experiment.

Feature Count	Function g	Multi-Link Descriptor d
1	Mean	Mean
2	Mean	Standard deviation (SD)
3	Mean	Mean/SD
4	Mean	Spectral Energy (E)
5	Mean	Spectral Centroid (SC)
6	Mean	2nd Order Spectral Moment (SOSM)
7	Mean	2nd Order Spectral Central Moment (SOSCM)
8	Mean	Spectral Kurtosis
9–14	NOP	Mean
15–20	NOP	Standard deviation (SD)
21–26	NOP	Mean/SD
27–32	NOP	Spectral Energy (E)
33–38	NOP	Spectral Centroid (SC)
39–44	NOP	2nd Order Spectral Moment (SOSM)
45–50	NOP	2nd Order Spectral Central Moment (SOSCM)
51–56	NOP	Spectral Kurtosis

All the descriptors are relative to the magnitude of \bar{H}_{lk} . Notice that in the last 8 rows of Table 2 NOP refers to multi-link features without any combination function. As we have 6 links in our setup, there are 6 instances of every descriptor. It is our aim to demonstrate that this technique provides valuable information to the classification stage.

A MATLAB code was implemented to, first, obtain the Doppler spectrum from the CSI data as given by Equation (10) and then, process the feature extraction from the frequency domain function. At this time all the features were loaded into the model. The processed dataset for each of the 3 reported locations have a total of 4040 vectors of 56 features each; it also includes a class column with a labeled metadata specifying the number of people in the crowd that correspond to the row.

3.5. Feature Selection

Our experiment had four variations relative to the features taken into account, each with different criteria for feature selection, as listed below:

- Variant 1: All features—our first iteration was a ‘brute force’ approach to get a first estimate of the classification power of the model. The purpose of this variant is to get a “baseline” against which all other options should be compared: any subset of all features must either perform better than this one, or else perform very similarly but with less computing effort.
- Variant 2: Multicollinearity feature selection—a common approach to feature selection is to find a pair of features that are highly-correlated (i.e., above a correlation threshold) and drop one of them. We implemented this algorithm in MATLAB and applied to our dataset. The selected features using a correlation threshold of 0.85 are listed in Table 3.

Table 3. Features selected by multicollinearity method.

Feature Count	Link #	Multi-Link Descriptor d
1	Mean	Mean
2	Mean	Standard deviation (SD)
3	Mean	Spectral Energy (E)
4	Mean	Spectral Centroid (SC)
5	Mean	Spectral Kurtosis
6	1	Standard deviation (SD)
7	2	Standard deviation (SD)
8	3	Standard deviation (SD)
9	4	Standard deviation (SD)
10	5	Standard deviation (SD)
11	6	Standard deviation (SD)
12	4	Mean/SD
13	5	Spectral Energy (E)
14	1	Spectral Centroid (SC)
15	3	Spectral Centroid (SC)
16	6	Spectral Centroid (SC)
17	2	2nd Order Spectral Moment (SOSM)

- Variant 3: Mean descriptors vs multi-link descriptors-in this experiment we tested our hypothesis about the quality of multi-link descriptors (those features with multi-link descriptors without combination function). To achieve this, we compare the classification performance when only multi-link descriptors are used vs the scenario in which only multi-link mean descriptors feed into the model.
- Variant 4: Single descriptor analysis-At the opposite extreme of variant 1 we would have the use of only one feature, which is not really of practical interest, but we find it useful as another baseline. It answers the question of how well a single feature (per channel) model can perform using a multi-link approach with respect to the accuracy.

3.5.1. Variant 1: All Features

As shown in Table 4, our model delivered an outstanding accuracy performance when all the 56 selected features are used. Four out of the five classifiers were able to correctly estimate the number of people in experiment location with 100% of accuracy. Only Random forest performed just below perfect.

Table 4. Accuracy rate for model with all features used.

	All Features	
	Training Accuracy	Testing Accuracy
Random Forest	99.5%	99.7%
Weighted KNN	100.0%	100.0%
Linear Discriminant	100.0%	100.0%
SVM	100.0%	100.0%
SVM Gaussian	100.0%	100.0%

3.5.2. Variant 2: Multicollinearity Feature Selection

With the correlation criteria our set of features decreased from 56 to 17. As expected, many multi-link NOP descriptors were eliminated by the algorithm since of the strong correlation among them. However, the complete set of multi-link Standard Deviation descriptors remained. The results are shown in Table 5. All classifiers performed above 90% of accuracy.

Table 5. Accuracy rate for model with multicollinearity feature selection.

	Collinear Selection	
	Training Accuracy	Testing Accuracy
Random Forest	99.8%	100.0%
Weighted KNN	99.9%	100.0%
Linear Discriminant	92.8%	92.5%
SVM	98.9%	98.5%
SVM Gaussian	100.0%	100.0%

3.5.3. Variant 3: Mean Descriptors vs. Multi-Link Descriptors

As mentioned before, other authors have disregarded multi-link descriptors in favor of either single-link descriptors [68], or mean descriptors [25]. In this experiment, we compared both kinds of descriptors face to face. As shown in Tables 6 and 7, while using mean descriptors yield fairly good results, remarkably multi-link vectors provide perfect accuracy for all the classifiers.

Table 6. Accuracy rate for model with mean descriptors only.

	Mean Descriptors	
	Training Accuracy	Testing Accuracy
Random Forest	98.1%	98.8%
Weighted KNN	97.5%	97.6%
Linear Discriminant	86.8%	87.1%
SVM	96.5%	96.8%
SVM Gaussian	99.4%	99.7%

Table 7. Accuracy rate for model with multi-link descriptors only.

	Multi-Link Descriptors	
	Training Accuracy	Testing Accuracy
Random Forest	100.0%	100.0%
Weighted KNN	100.0%	100.0%
Linear Discriminant	100.0%	100.0%
SVM	100.0%	100.0%
SVM Gaussian	100.0%	100.0%

3.5.4. Variant 4: Single Descriptor Analysis

Now that we have empirically demonstrated that multi-link descriptors have better performance than single mean descriptors, our next step was focused on reducing the dimensionality of the model.

A hint for this task was provided by variant 2, where we applied multicollinearity feature selection. That process outlined the quality of multi-link standard deviation descriptor. As show in Table 8 multi-link SD provides 100% accuracy for the SVM Gaussian classifier.

Table 8. Accuracy rate for every individual multi-link descriptor.

Classifiers	SD	Mean	E	SC	SOSM	SOSCM	Kurtosis
Random Forest	99.7%	96.5%	98.7%	98.6%	97.8%	98.2%	98.5%
Weighted KNN	99.8%	98.5%	99.6%	99.1%	99.3%	99.2%	99.3%
Linear Discriminant	85.2%	67.6%	87.1%	77.0%	73.1%	73.1%	80.1%
SVM	90.5%	72.4%	91.8%	81.7%	79.3%	79.3%	83.6%
SVM Gaussian	100.0%	99.4%	99.9%	99.8%	99.4%	99.2%	99.8%

In order to further investigate the performance of each individual descriptor, we implemented Neighborhood Component Analysis (NCA), a multi-class, high-dimensional feature selection method initially proposed by Yang et al. [2]. This method maximizes the expected leave-one-out classification accuracy using the gradient ascent technique.

An examination of the results of NCA shown in Figure 4 indicates that: (1) Multi-link approach has higher prediction power than single-link based methods, and (2) Multi-link SD is the best single-statistic feature vector among those under review. These observations are in line with the outcomes of variants 2 and 3.

Finally, we were interested on knowing how many Wi-Fi links deliver an optimal trade-off for accuracy in our multi-link model using SD as descriptor function. To accomplish this, we ran several iterations of the model, including one additional link at each iteration and repeating the process for every possible link combination. Results in Figures 5 and 6 show that the accuracy of our model increases logarithmically with the number of available links, and this metric is near-to-perfect with a set of 6 available links (The specific numbers in this figure are too small to be read; this figure is intended to have a bird's eye view comparing the quantity of green squares (good classification) against the pink ones, as well as the way the AUC, marked in blue color, gets more and more of the area as it upper-left side grows).

3.5.5. Summary of Results by Number of Features

Table 9 shows a summary of the results from the experiment variants sorted in ascending order by the number of features for the model involved. It is worth noticing that SVM with Gaussing kernel provides perfect accuracy in all scenarios. Hence, all scenarios have at least one classifier with perfect accuracy.

Table 9. Accuracy rates for models with multi-link descriptors.

Classifiers	Multi-Link SD 6 Features	Collinear Selection 17 Features	All Multi-Link 48 Features	All Features 56 Features
Random Forest	99.7%	100.0%	100.0%	99.7%
Weighted KNN	99.8%	100.0%	100.0%	100.0%
Linear Discriminant	85.2%	92.5%	100.0%	100.0%
SVM	90.5%	98.5%	100.0%	100.0%
SVM Gaussian	100.0%	100.0%	100.0%	100.0%

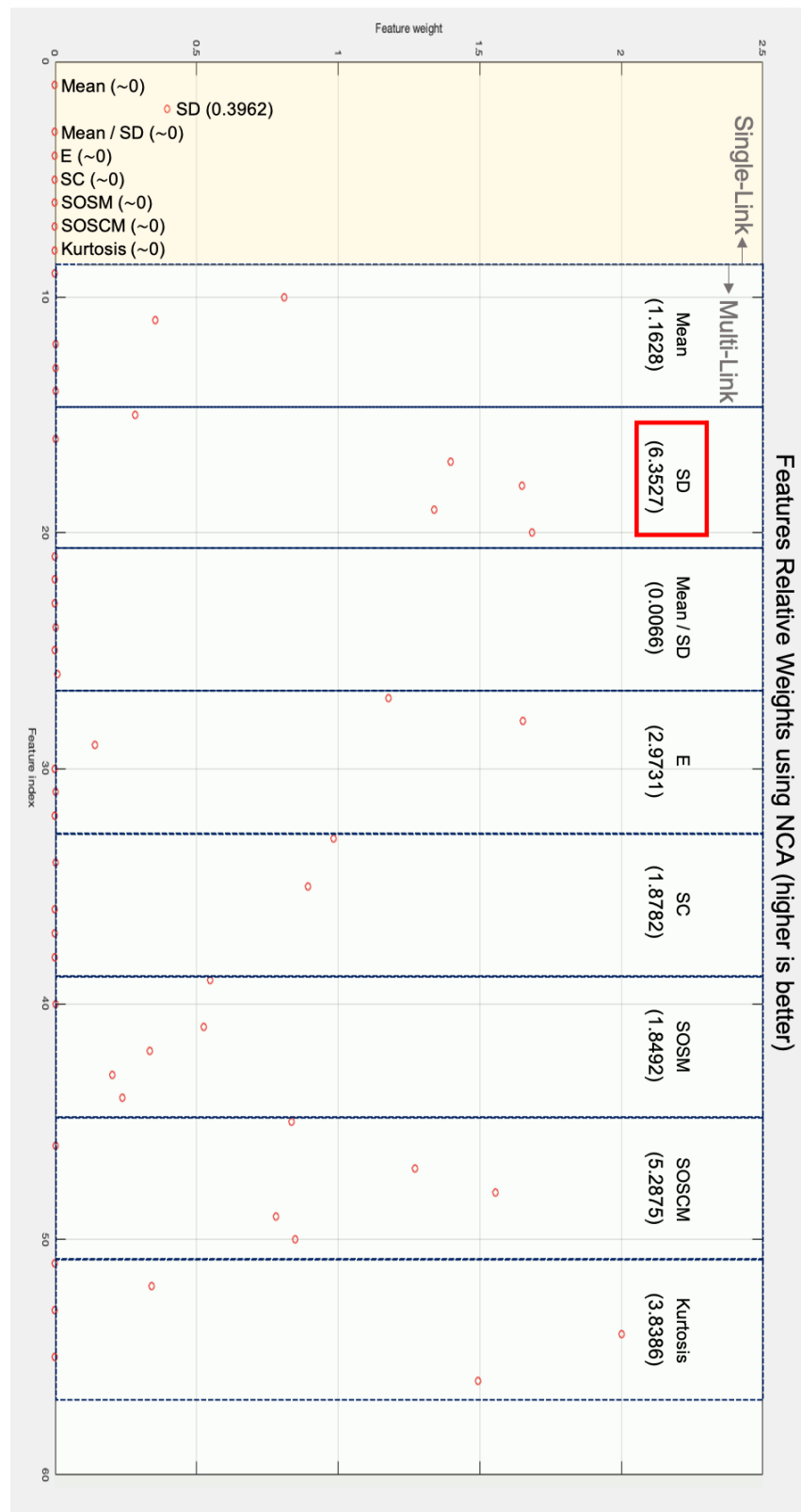


Figure 4. Relative Weights of Features using Neighborhood Component Analysis (NCA) (higher is better).

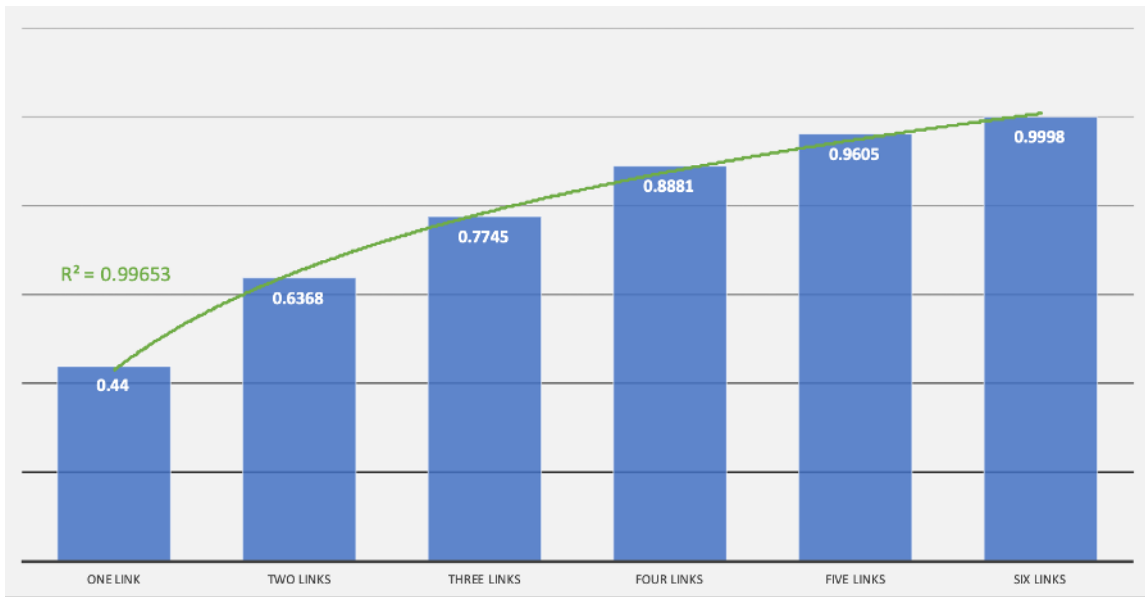


Figure 5. Accuracy of the model at incremental links.

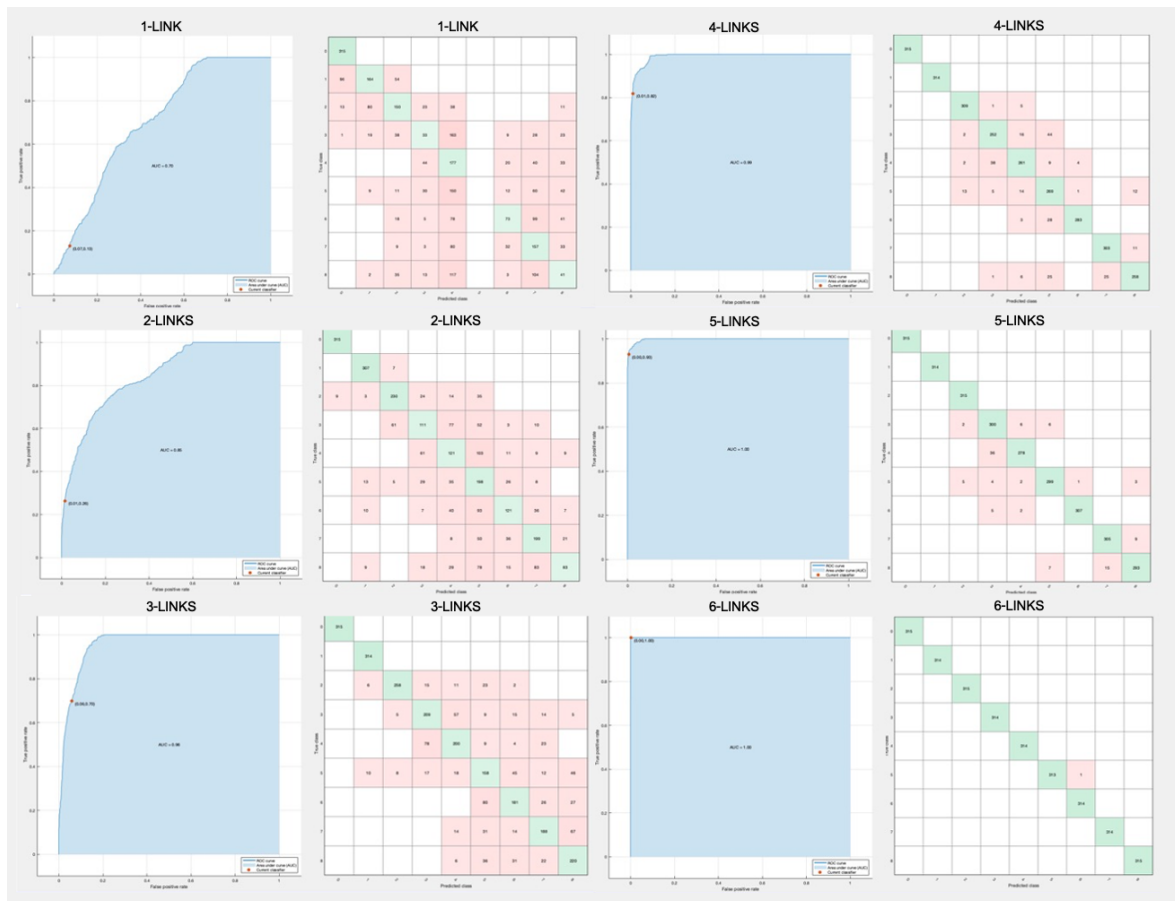


Figure 6. ROC and Confusion Matrix for incremental links.

3.5.6. Results in All Rooms

The four initial scenarios were studied using the dataset of Room A. Now, our interest was to investigate if the results in the other two rooms are similar to those we have obtained so far.

Of special interest was to validate the quality of the multi-link SD descriptor and whether the high performance of SVM-Gaussian was also observed in the rest of the locations.

Table 10 shows the results for Room A, Room B & Room C and provides evidence that the results obtained in the first iterations with dataset of Room A extend well to the other available datasets. Multi-link SD descriptor produces accuracies of more than 97% for Random Forest, Weighted KNN and SVM-Gaussian classifiers in all rooms. We can see that for SVM-Gaussian, the model delivered perfect accuracy in Room A and nearly perfect in Room B (99.7%) and Room C (99.9%). The confusion matrices are shown in Figure 7.

Table 10. Accuracy rates for multi-link approach in all rooms.

Multi-Link SD			
	Room A	Room B	Room C
Random Forest	99.7%	97.0%	98.2%
Weighted KNN	99.8%	98.8%	99.7%
Linear Discriminant	85.2%	69.2%	64.8%
SVM	90.5%	74.2%	71.3%
SVM Gaussian	100.0%	99.7%	99.9%
Collinear Selection			
	Room A	Room B	Room C
Random Forest	100.0%	99.2%	100.0%
Weighted KNN	100.0%	100.0%	100.0%
Linear Discriminant	92.5%	87.2%	87.2%
SVM	98.5%	97.3%	99.4%
SVM Gaussian	100.0%	100.0%	100.0%
All multi-link			
	Room A	Room B	Room C
Random Forest	100.0%	100.0%	100.0%
Weighted KNN	100.0%	100.0%	100.0%
Linear Discriminant	100.0%	99.2%	98.4%
SVM	100.0%	100.0%	100.0%
SVM Gaussian	100.0%	100.0%	100.0%
All features			
	Room A	Room B	Room C
Random Forest	99.7%	99.7%	100.0%
Weighted KNN	100.0%	100.0%	100.0%
Linear Discriminant	100.0%	99.1%	98.2%
SVM	100.0%	100.0%	100.0%
SVM Gaussian	100.0%	100.0%	100.0%

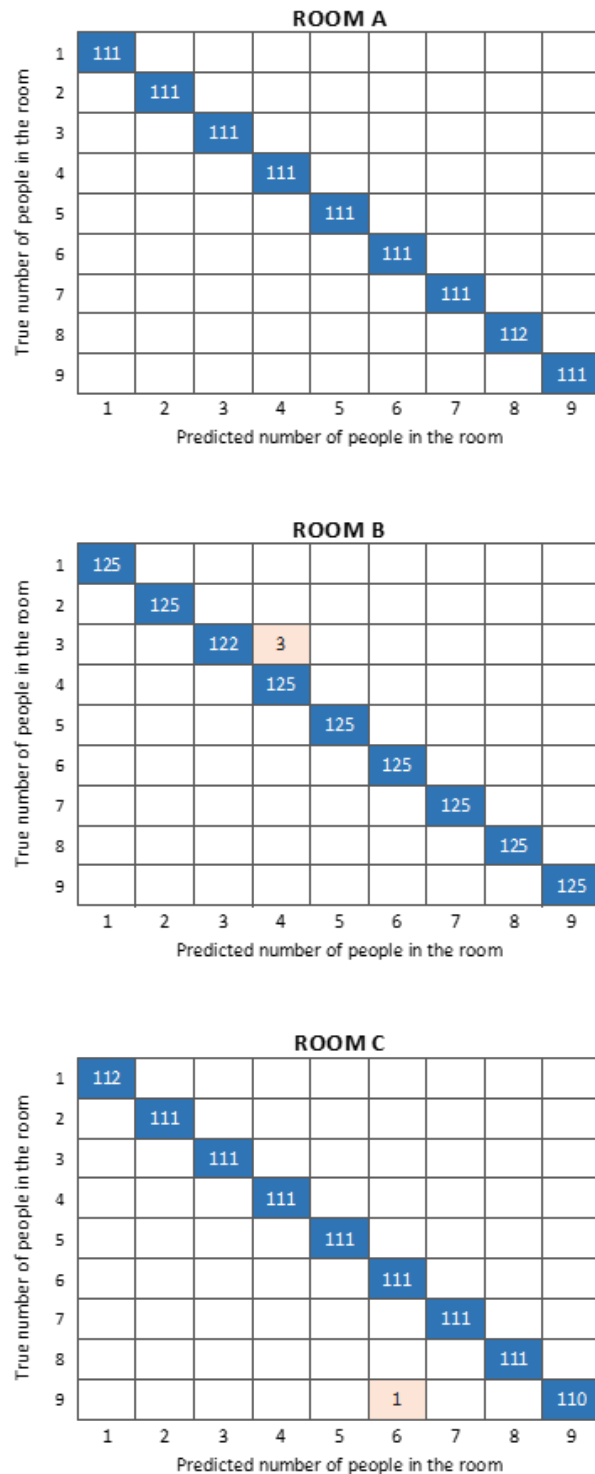


Figure 7. Confusion matrices for SVM-Gaussian with multi-link SD descriptors in all Rooms.

‘All multi-link’ feature set show an excellent performance since four out of five classifiers estimated the size of the crowd present in the Room with perfect accuracy. This was true for all the test cases in the available dataset.

4. Conclusions

In this paper, we have presented a novel method for crowd measuring (counting, in particular) using recognition of patterns in the Channel State Information over multiple

links, and showed that the use of multiple links, instead of a single one –or the aggregation of several ones in an average– can be translated into an improved performance at least for the people counting scenario considered in the dataset we used. This was the main contribution of this work.

Using our method, based on data-driven Machine Learning supervised classifiers, we empirically demonstrated that multi-link predictors yield better performance in terms of accuracy than those that use the mean value for multi-link, or single-value of one link.

Another contributions of our work was to show that even reducing the number of features used for training and predicting with the classifiers, the performance could be maintained above that of other state-of-the-art methods. Also, we showed the prediction power of the Standard Deviation when used over the channel response data given by the Doppler Spectrum.

In Table 11, we summarize the comparison of our approach with other state-of-the-art methods.

Table 11. Benchmark of accuracy rates for state-of-the-art Wi-Fi-based crowd counting.

Author	Wi-Fi APs	Features	Classifier	Accuracy Rate	Max # of People
Di Domenico	COTS	Average Spectral Kurtosis	Naive Bayes	0.81	8
Liu	COTS	-	Deep Learning	0.82	5
Zuo	Custom	Statistics, FFT-based, Shape-based	SVM-TKL	0.96	7
Kianoush	COTS	Statistics in Space Domain	LSTM	0.99	5
Our Work	COTS	Multi-link SD in Doppler Spectrum	SVM-Gaussian	>0.99	8

For future work, we are interested in exploring the application of our method to less restricted scenarios (for instance, by increasing the maximum number of people in the crowd), and to take measurements in real-life situations. We also want to explore other crowd properties like the direction of movement and cope with limitations imposed by scale and a dynamic environment.

Author Contributions: conceptualization, R.F.B. and E.E.; investigation and methodology, E.E. and R.F.B.; supervision, R.F.B.; project administration, R.F.B.; experimental validation, E.E.; writing—original draft preparation, E.E.; writing—review and editing, R.F.B. and C.V.-R. and C.E.G.-T. and D.M.; funding acquisition, C.V.-R. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the SEP-CONACyT Research Project under Grant 256237, the School of Engineering and Sciences at Tecnológico de Monterrey and the Telecommunications Research Group. The Ph.D. studies of E.E. were supported by the CONACyT and the Tecnológico de Monterrey.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Zhang, X.; Yu, Q.; Wang, Y. Fuzzy Evaluation of Crowd Safety Based on Pedestrians' Number and Distribution Entropy. *Entropy* **2020**, *22*, 832. [[CrossRef](#)] [[PubMed](#)]
- Yang, J.; Cheng, J.; Chen, Y. Mobile Sensing Enabled Robust Detection of Security Threats in Urban Environments. In *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST*; Springer: Berlin/Heidelberg, Germany, 2012; Volume 74, pp. 88–104. [[CrossRef](#)]
- Radianti, J.; Granmo, O.C.; Bouhmala, N.; Sarshar, P.; Yazidi, A.; Gonzalez, J. Crowd models for emergency evacuation: A review targeting human-centered sensing. In Proceedings of the Annual Hawaii International Conference on System Sciences, Maui, HI, USA, 7–10 January 2013; pp. 156–165. [[CrossRef](#)]
- Fortino, G.; Russo, W.; Savaglio, C.; Viroli, M.; Zhou, M. Modeling Opportunistic IoT Services in Open IoT Ecosystems. In Proceedings of the XVIII Workshop from Objects to Agents (WOA 2017), Scila, Italy, 2017; pp. 90–95.
- Bouchnita, A.; Jebrane, A. A hybrid multi-scale model of COVID-19 transmission dynamics to assess the potential of non-pharmaceutical interventions. *Chaos Sol. Fractals* **2020**, *138*, 109941. [[CrossRef](#)] [[PubMed](#)]

6. Perera, C.; Zaslavsky, A.; Christen, P.; Georgakopoulos, D. Sensing as a Service Model for Smart Cities Supported by Internet of Things. *Trans. Emerg. Tel. Tech.* **2014**, *25*, 81–93. [[CrossRef](#)]
7. Luo, T.; Kanhere, S.S.; Huang, J.; Das, S.K.; Wu, F. Sustainable incentives for mobile crowdsensing: Auctions, lotteries, and trust and reputation systems. *IEEE Commun. Mag.* **2017**, *55*, 68–74. [[CrossRef](#)]
8. Joglekar, P.; Kulkarni, V. Mobile crowd sensing for urban computing. *Int. J. Latest Trends Eng. Technol.* **2017**, *7*. [[CrossRef](#)]
9. Calabrese, F.; Pereira, F.C.; Di Lorenzo, G.; Liu, L.; Ratti, C. The geography of taste: Analyzing cell-phone mobility and social events. In *International Conference on Pervasive Computing*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 22–37.
10. Djenouri, D.; Laidi, R.; Djenouri, Y.; Balasingham, I. Machine Learning for Smart Building Applications: Review and Taxonomy. *Mach. Learn. Smart Build. Appl. Rev. Taxon. Acm Comput. Surv.* **2018**, *1*, 1–42. [[CrossRef](#)]
11. Bisio, I.; Lavagetto, F.; Marchese, M.; Sciarrone, A. Smartphone-centric ambient assisted living platform for patients suffering from co-morbidities monitoring. *IEEE Commun. Mag.* **2015**, *53*, 34–41. [[CrossRef](#)]
12. Wang, Y.; Yang, J.; Chen, Y.; Liu, H.; Gruteser, M.; Martin, R.P. Tracking human queues using single-point signal monitoring. In *Proceedings of the 12th Annual International Conference on Mobile Systems, Applications, and Services*, Bretton Woods, NH, USA, 16–19 June 2014; pp. 42–54.
13. Zhan, B.; Monekosso, D.N.; Remagnino, P.; Velastin, S.A.; Xu, L.Q. Crowd analysis: A survey. *Mach. Vis. Appl.* **2008**, *19*, 345–357. [[CrossRef](#)]
14. Wang, C.; Zhang, H.; Yang, L.; Liu, S.; Cao, X. Deep people counting in extremely dense crowds. In *Proceedings of the 23rd ACM International Conference on Multimedia*, Brisbane, Australia, 26–30 October 2015; pp. 1299–1302.
15. Nuño-Maganda, M.A.; Herrera-Rivas, H.; Torres-Huitzil, C.; Marisol Marin-Castro, H.; Coronado-Pérez, Y. On-Device learning of indoor location for WiFi fingerprint approach. *Sensors* **2018**, *18*, 2202.
16. Brena, R.F.; García-Vázquez, J.P.; Galván-Tejada, C.E.; Muñoz-Rodríguez, D.; Vargas-Rosales, C.; Fangmeyer, J. Evolution of Indoor Positioning Technologies: A Survey. *J. Sens.* **2017**, *2017*. [[CrossRef](#)]
17. Bishop, C.M. *Pattern Recognition and Machine Learning*; Springer: Berlin/Heidelberg, Germany, 2006.
18. Domazetovic, A.; Greenstein, L.J.; Mandayam, N.B.; Seskar, I. Estimating the Doppler spectrum of a short-range fixed wireless channel. *IEEE Commun. Lett.* **2003**, *7*, 227–229. [[CrossRef](#)]
19. Ma, Y.; Zhou, G.; Wang, S. WiFi sensing with channel state information: A survey. *ACM Comput. Surv. (CSUR)* **2019**, *52*, 1–36. [[CrossRef](#)]
20. Depatla, S.; Muralidharan, A.; Mostofi, Y. Occupancy estimation using only WiFi power measurements. *IEEE J. Sel. Areas Commun.* **2015**, *33*, 1381–1393. [[CrossRef](#)]
21. Xu, C.; Firner, B.; Moore, R.S.; Zhang, Y.; Trappe, W.; Howard, R.; Zhang, F.; An, N. SCPL: Indoor Device-Free Multi-Subject Counting and Localization Using Radio Signal Strength. In *Proceedings of the 12th international conference on Information processing in sensor networks-IPSN '13*, Philadelphia, PA, USA, 8–11 April 2013; ACM Press: New York, NY, USA, 2013; p. 79. [[CrossRef](#)]
22. Yuan, Y.; Zhao, J.; Qiu, C.; Xi, W. Estimating crowd density in an RF-based dynamic environment. *IEEE Sens. J.* **2013**, *13*, 3837–3845. [[CrossRef](#)]
23. Doong, S.H. Spectral human flow counting with rssi in wireless sensor networks. In *Proceedings of the 2016 international conference on distributed computing in sensor systems (DCOSS)*, Washington, DC, USA, 26–28 May 2016; pp. 110–112.
24. Yoshida, T.; Taniguchi, Y. Estimating the number of people using existing WiFi access point in indoor environment. In *Proceedings of the 6th European Conference of Computer Science (ECCS '15)*, Rome, Italy, 7–9 November 2015; pp. 46–53.
25. Di Domenico, S.; Pecoraro, G.; Cianca, E.; De Sanctis, M. Trained-once device-free crowd counting and occupancy estimation using WiFi: A Doppler spectrum based approach. In *Proceedings of the International Conference on Wireless and Mobile Computing, Networking and Communications*, Dubai, United Arab Emirates, 27–28 August 2016; pp. 1–8. [[CrossRef](#)]
26. Xi, W.; Zhao, J.; Li, X.Y.; Zhao, K.; Tang, S.; Liu, X.; Jiang, Z. Electronic frog eye: Counting crowd using WiFi. In *Proceedings of the IEEE INFOCOM 2014-IEEE Conference on Computer Communications*, Toronto, ON, Canada, 27 April–2 May 2014; pp. 361–369.
27. Rocklöv, J.; Sjödin, H. High population densities catalyse the spread of COVID-19. *J. Travel Med.* **2020**, *27*, taaa038. [[CrossRef](#)]
28. Yamin, M.; Ades, Y. Crowd management with RFID & wireless technologies. In *Proceedings of the 1st International Conference on Networks and Communications, NetCoM*, Toronto, ON, Canada, 27 April–2 May 2014; pp. 439–442. [[CrossRef](#)]
29. Wijermans, N.; Conrado, C.; van Steen, M.; Martella, C.; Li, J. A landscape of crowd-management support: An integrative approach. *Saf. Sci.* **2016**, *86*, 142–164. [[CrossRef](#)]
30. Helbing, D.; Johansson, A. Pedestrian, crowd, and evacuation dynamics. *arXiv* **2013**, arXiv:1309.1609.
31. Helbing, D.; Buzna, L.; Johansson, A.; Werner, T. Self-Organized Pedestrian Crowd Dynamics: Experiments, Simulations, and Design Solutions. *Transp. Sci.* **2005**, *39*, 1–24. [[CrossRef](#)]
32. Still, G.K. *Introduction to Crowd Science*; CRC Press: Boca Raton, FL, USA, 2014.
33. Helbing, D.; Johansson, A.; Al-Abideen, H.Z. Dynamics of crowd disasters: An empirical study. *Phys. Rev. Stat. Nonlinear Soft Matter Phys.* **2007**, *75*. [[CrossRef](#)]
34. Pathan, S.S.; Al-Hamadi, A.; Michaelis, B. Incorporating social entropy for crowd behavior detection using SVM. *Lect. Notes Comput. Sci. (Incl. Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinform.)* **2010**, *6453*, 153–162. [[CrossRef](#)]
35. Vicsek, T.; Zafeiris, A. Collective motion. *Phys. Rep.* **2012**, *517*, 71–140. [[CrossRef](#)]

36. Xu, F.; Rao, Y.; Wang, Q. An unsupervised abnormal crowd behavior detection algorithm. In Proceedings of the 2017 International Conference on Security, Pattern Analysis, and Cybernetics, SPAC, Shenzhen, China, 15–17 December 2017; pp. 219–223. [\[CrossRef\]](#)
37. Yousefi, S.; Narui, H.; Dayal, S.; Ermon, S.; Valaee, S. A Survey on Behavior Recognition Using WiFi Channel State Information. *IEEE Commun. Mag.* **2017**, *55*, 98–104. [\[CrossRef\]](#)
38. Sobron, I.; Del Ser, J.; Eizmendi, I.; Velez, M. Device-Free People Counting in IoT Environments: New Insights, Results, and Open Challenges. *IEEE Internet Things J.* **2018**, *5*, 4396–4408. [\[CrossRef\]](#)
39. Wang, W.; Liu, A.X.; Shahzad, M.; Ling, K.; Lu, S. Device-Free Human Activity Recognition Using Commercial WiFi Devices. *IEEE J. Sel. Areas Commun.* **2017**, *35*, 1118–1131. [\[CrossRef\]](#)
40. Wu, C.; Yang, Z.; Zhou, Z.; Liu, X.; Liu, Y.; Cao, J. Non-invasive detection of moving and stationary human with WiFi. *IEEE J. Sel. Areas Commun.* **2015**, *33*, 2329–2342. [\[CrossRef\]](#)
41. Jiang, H.; Cai, C.; Ma, X.; Yang, Y.; Liu, J. Smart Home Based on WiFi Sensing: A Survey. *IEEE Access* **2018**, *6*, 13317–13325. [\[CrossRef\]](#)
42. Guo, L.; Wang, L.; Liu, J.; Zhou, W. A Survey on Motion Detection Using WiFi Signals. In Proceedings of the 12th International Conference on Mobile Ad-Hoc and Sensor Networks, MSN 2016, Hefei, China, 16–18 December 2016; pp. 202–206. [\[CrossRef\]](#)
43. Wang, X.; Yang, C.; Mao, S. PhaseBeat: Exploiting CSI phase data for vital sign monitoring with commodity WiFi devices. In Proceedings of the 2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS), Atlanta, GA, USA, 5–8 June 2017; pp. 1230–1239.
44. Han, C.; Wu, K.; Wang, Y.; Ni, L.M. WiFall: Device-free fall detection by wireless networks. *IEEE Trans. Mob. Comput.* **2014**, *16*, 271–279. [\[CrossRef\]](#)
45. Chowdhury, T.Z. Using Wi-Fi Channel State Information (CSI) for Human Activity Recognition and Fall Detection. Ph.D. Thesis, University of British Columbia, Vancouver, BC, USA, 2018.
46. Myrvoll, T.A.; Håkegård, J.E.; Matsui, T.; Septier, F. Counting public transport passenger using WiFi signatures of mobile devices. In Proceedings of the 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), Yokohama, Japan, 16–19 October 2017; pp. 1–6.
47. Andersen, J.B.; Nielsen, J.O.; Pedersen, G.F.; Bauch, G.; Dietl, G. Doppler spectrum from moving scatterers in a random environment. *IEEE Trans. Wirel. Commun.* **2009**, *8*, 3270–3277. [\[CrossRef\]](#)
48. Khan, M.A.; Khan, S.F. IoT based framework for Vehicle Over-speed detection. In Proceedings of the 2018 1st International Conference on Computer Applications & Information Security (ICCAIS), Riyadh, Saudi Arabia, 4–5 April 2018; pp. 1–4. [\[CrossRef\]](#)
49. Depatla, S.; Mostofi, Y. Crowd counting through walls using WiFi. In Proceedings of the 2018 IEEE International Conference on Pervasive Computing and Communications (PerCom), Athens, Greece, 19–23 March 2018; pp. 1–10.
50. Shen, J.; Cao, J.; Liu, X.; Tang, S. SNOW: Detecting Shopping Groups Using WiFi. *IEEE Internet Things J.* **2018**, *5*, 1. [\[CrossRef\]](#)
51. Tang, X.; Xiao, B.; Li, K. Indoor Crowd Density Estimation Through Mobile Smartphone Wi-Fi Probes. In Proceedings of the IEEE Transactions on Systems, Man, and Cybernetics: Systems, Miyazaki, Japan, 7–10 October 2018; pp. 1–12. [\[CrossRef\]](#)
52. Depatla, S.; Mostofi, Y. Passive crowd speed estimation and head counting using WiFi. In Proceedings of the 15th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON), Hong Kong, China, 11–13 June 2018; pp. 1–9.
53. Sidhu, B.; Singh, H.; Chhabra, A. Emerging Wireless Standards-WiFi, ZigBee and WiMAX. *World Acad. Sci. Eng. Technol. Int. J. Electr. Comput. Energ. Electron. Commun. Eng.* **2007**, *1*, 42–48.
54. Wilson, J.; Patwari, N. See-through walls: Motion tracking using variance-based radio tomography networks. *IEEE Trans. Mob. Comput.* **2011**, *10*, 612–621. [\[CrossRef\]](#)
55. Kosba, A.E.; Saeed, A.; Youssef, M. Rasid: A robust wlan device-free passive motion detection system. In Proceedings of the 2012 IEEE International Conference on Pervasive Computing and Communications, Lugano, Switzerland, 19–23 March 2012; pp. 180–189.
56. Youssef, M.; Mah, M. Challenges: Device-free Passive Localization for Wireless. *MobiCom* **2007**, *7*. [\[CrossRef\]](#)
57. Yang, Z.; Zhou, Z.; Liu, Y. From RSSI to CSI. *ACM Comput. Surv.* **2013**, *46*, 1–32. [\[CrossRef\]](#)
58. Wu, K.; Xiao, J.; Yi, Y.; Chen, D.; Luo, X.; Ni, L.M. CSI-based indoor localization. *IEEE Trans. Parallel Distrib. Syst.* **2013**, *24*, 1300–1309. [\[CrossRef\]](#)
59. Halperin, D.; Hu, W.; Sheth, A.; Wetherall, D. Predictable 802.11 packet delivery from wireless channel measurements. *ACM SIGCOMM Comput. Commun. Rev.* **2012**, *40*, 159. [\[CrossRef\]](#)
60. Goldsmith, A. *Wireless Communications*; Cambridge University Press: Cambridge, UK, 2005. [\[CrossRef\]](#)
61. Ahmed, B.; Abdul Matin, M. *Coding for MIMO-OFDM in Future Wireless Systems*; Springer Briefs in Electrical and Computer Engineering, Springer International Publishing: Cham, Switzerland, 2015; pp. 11–21. [\[CrossRef\]](#)
62. Dubuc, C.; Starks, D.; Creasy, T.; Hou, Y. A MIMO-OFDM prototype for next-generation wireless WANs. *IEEE Commun. Mag.* **2004**, *42*, 82–87. [\[CrossRef\]](#)
63. Liu, S.; Zhao, Y.; Chen, B. WiCount: A deep learning approach for crowd counting using wifi signals. In Proceedings of the 15th IEEE International Symposium on Parallel and Distributed Processing with Applications, Guangzhou, China, 12–15 December 2017; pp. 967–974. [\[CrossRef\]](#)
64. Liu, S.; Zhao, Y.; Xue, F.; Chen, B.; Chen, X. DeepCount: Crowd counting with WiFi via deep learning. *arXiv* **2019**, arXiv:1903.05316
65. Cheng, Y.K.; Chang, R.Y. Device-free indoor people counting using Wi-Fi channel state information for Internet of Things. In Proceedings of the GLOBECOM 2017–2017 IEEE Global Communications Conference, Singapore, 4–8 December 2017; pp. 1–6.

66. Mabuchi, T.; Taniguchi, Y.; Shirahama, K. Person recognition using Wi-Fi channel state information in an indoor environment. In Proceedings of the 2020 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-Taiwan), Taoyuan, Taiwan, 28–30 September 2020; pp. 1–2.
67. Zou, H.; Zhou, Y.; Yang, J.; Gu, W.; Xie, L.; Spanos, C. Freecount: Device-free crowd counting with commodity wifi. In Proceedings of the GLOBECOM 2017-2017 IEEE Global Communications Conference, Singapore, 4–8 December 2017; pp. 1–6.
68. Yang, D.; Wang, T.; Sun, Y.; Wu, Y. Doppler Shift Measurement Using Complex-Valued CSI of WiFi in Corridors. In Proceedings of the 2018 3rd International Conference on Computer and Communication Systems, ICCCS 2018, Nagoya, Japan, 27–30 April 2018; pp. 497–501. [[CrossRef](#)]
69. Xiao, Y. IEEE 802.11 n: Enhancements for higher throughput in wireless LANs. *IEEE Wirel. Commun.* **2005**, *12*, 82–91. [[CrossRef](#)]
70. Di Domenico, S.; De Sanctis, M.; Cianca, E.; Bianchi, G. A trained-once crowd counting method using differential wifi channel state information. In Proceedings of the 3rd International on Workshop on Physical Analytics, Singapore, 26–30 June 2016; pp. 37–42.
71. Nita, G.M.; Gary, D.E. Statistics of the Spectral Kurtosis Estimator. *Publ. Astron. Soc. Pac.* **2010**, *122*, 595–607. [[CrossRef](#)]