Machine Learning Enabled Performance Prediction Model for Massive-MIMO HetNet System

Shuvabrata Bandopadhyaya ¹, Soumya Ranjan Samal ² and Vladimir Poulkov ^{2,*} ¹ School of Engineering & Technology, BML Munjal University, Gurugram 122414, India; shuvabrata.bandopadhyaya@bmu.edu.in² Faculty of Telecommunications, Technical University of Sofia, 1756 Sofia, Bulgaria; sranjansamal@tu-sofia.bg

* Correspondence: vkp@tu-sofia.bg

Abstract: To support upcoming novel applications, fifth generation (5G) and beyond 5G (B5G) wireless networks are being propelled to deploy an ultra-dense network with an ultra-high spectral efficiency using the combination of heterogeneous network (HetNet) solutions and massive Multiple Input Multiple Output (MIMO). As the deployment of massive MIMO HetNet systems involves a high capital expenditure, network service providers need a precise performance analysis before investment. The performance of such networks is limited because of presence of inter-cell and inter-tier interferences. The conventional analytic approach to model the performance of such networks is not trivial, as the performance is a stochastic function of many network parameters. This paper proposes a machine learning (ML) approach to predict the network performance of a massive MIMO HetNet system considering a multi-cell scenario. This paper considers a two-tier network in which the base stations of each tier are equipped with massive MIMO systems working in a sub 6-GHz band. The coverage probability (CP) and area spectral efficiency (ASE) are considered to be the network performance metrics that quantify the reliability and achievable rate in the network, respectively. Here, an ML model is inferred to predict the numerical values of the performance metrics for an arbitrary network configuration. In the process of practical deployments of future networks, the use of this model could be very valuable.

Keywords: 5G; B5G wireless networks; massive MIMO; HetNet; machine learning; coverage probability; area spectral density

Citation: Bandopadhyaya, S.; Samal, S.R.; Poulkov, V. Machine Learning Enabled Performance Prediction Model for Massive-MIMO HetNet System. *Sensors* **2021**, *21*, 800. <https://doi.org/10.3390/s21030800>

Academic Editors: Adrian Kliks, Paweł Sroka, Cynthia Hood and Nikos Dimitrou

Received: 15 December 2020

Accepted: 21 January 2021

Published: 26 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

To support the upcoming novel applications, such as IoT, self-driving cars, Industry 4.0, smart healthcare systems, AR/VR services, fifth generation (5G) and beyond fifth generation (B5G) wireless networks aim to achieve ultra-low latency and ultra-reliability with a multi-gigabit transmission rate [1]. The combination of heterogeneous network (HetNet) solutions and massive MIMO promises significant improvement in the physical layer performance by deploying an ultra-dense network with an ultra-high spectral efficiency [2,3]. Massive MIMO is the next generation MIMO system that significantly enhances the spectral efficiency of the communication link compared with its conventional counterpart. With this technology, a few hundred antennas are deployed at the base-station (BS) to serve a few tens of active users in same time–frequency grid [4,5]. Massive MIMO is considered a key technology for upcoming wireless generations to support $100\times$ data rates per user and per cell by implementing adaptive beamforming and spatial multiplexing technologies with large antenna arrays [6]. HetNet is a network densification technique in which several classes of low-powered transmitters are deployed with the existing macro cells sharing the same spectrum. The low-powered small-/pico-cells are deployed to target highly concentrated user groups. The network densification process significantly improves the coverage penetration and area spectral efficiency of the network with non-uniformly

distributed users [7]. By optimizing resource utilization and network performance, HetNets are going to be a principal candidate for the implementation of 5G and B5G networks [8].

For achieving customer satisfaction, which is the aim of network service providers, the latter continuously upgrade the network for maximizing the coverage probability and achievable rate for the end users. The deployment of massive MIMO HetNet systems for the upcoming network generation needs precise planning and high capital expenditure. Hence, before deployment, network service providers must optimize network parameters in order to meet the desired goal, which requires precise performance analysis. The analysis and estimation of massive MIMO system performance before practical deployment has been a major concern of research for the past few years. An asymptotic analysis of the coverage probability and sum-rate of a single-cell massive MIMO system has been presented in the literature [9]. The performance of single-cell downlink massive MIMO in terms of spectral efficiencies and link reliability using various precoding techniques has been analyzed and compared by the authors of [10]. Gao et al. have proposed a performance evaluation technique for massive a MIMO system based on propagation data [11]. Feng et al. have modeled the user-interference power distribution in a single-cell multi-user massive MIMO system using Gamma function, and formulated asymptotic deterministic equivalences for sum-rate and outage probability in terms of tight-form approximation of the model [12].

The performance of a massive MIMO system in a multi-cell scenario is limited because of the presence of inter-cell interference. A typical user in a multi-cell scenario associated with a given BS receives signals from other BSs as interference. Liang et al. have done a statistical analysis of the interference present in a massive MIMO system, considering inter-cell interference as the dominant component [13]. Li et al. have derived a large-scale approximation of the downlink signal power to interference plus noise power ratio (SINR) in a multi-cell massive MIMO system with their proposed MMSE precoder [14]. Adhikary et al. have done a tractable analysis of the interference in uplink for a large-scale antenna system [15]. The closed form outage probability of a typical user has been derived in terms of BS density and the maximum number of users served by a BS for a multi-cell massive MIMO system, assuming the BSs are distributed randomly following a Poisson point process (PPP) [16].

Similarly, in HetNet, the network performance is limited by inter-tier interference. The interference experienced by a typical user in a k -tier HetNet has been accurately modeled by the authors of [17], where each tier differs by transmitting power and cell density. Closed-form approximation has been made for the coverage probability, traffic off-loading, and sum-rate considering inter-tier interference. The analysis of HetNet with a multiple antenna system is relatively complex because of the random matrix channel. The expressions for success probability and area spectral efficiency have been formulated for MIMO heterogeneous cellular networks using a Toeplitz matrix representation [18]. With proper analysis of the interference and its cancellation, the association of massive MIMO with HetNet is a promising physical layer solution for next generation wireless networks with several-fold increases in area spectral efficiency (ASE) and coverage probability (CP) [19]. The detailed system model and performance analysis of massive MIMO with HetNet is discussed in the literature [20] and the references therein. The impact of different network settings on the virtual coverage areas for massive MIMO-enabled HetNet has been explicitly studied in the literature [20]. With a theoretic framework for massive MIMO HetNets, tractable expressions have been obtained [21] for evaluating the average achievable rate and ASE. The performance of massive MIMO HetNet systems with limited channel information is discussed in the literature [22]. A wireless backhaul-based downlink rate enhancement technique for multi-antenna HetNet is presented in the literature [23]. A tractable approach is developed for evaluation of the spectral and energy efficiency in a massive MIMO-enabled three-tier network considering the presence of eavesdroppers [24].

The performance of massive MIMO-enabled HetNet in a multi-cell scenario is limited because of the presence of both inter-cell and inter-tier interferences. Therefore, the conventional analytic approach to model the statistical behavior of overall interference is

not trivial, as it needs to include many stochastic network parameters. To overcome this challenge, this paper has proposed an ML-based approach to model the massive MIMO enabled HetNet system for predicting the network performance. To the best knowledge of the authors, currently, in the literature, such an approach to predict the performance of such networks has not yet been considered.

The main focus of the work is to predict the performance of a two-tier network, in which the base stations of each tier are equipped with massive MIMO systems working in a sub 6-GHz band. This paper assumes perfect channel state information (CSI) at the transmitter. The analysis for imperfect/outdated CSI is beyond the scope of the work and may be considered in future. In this work, the coverage probability (CP) and area spectral efficiency (ASE) are considered to be the network performance metrics that quantify the reliability and achievable rate of the network, respectively. Both metrics are stochastically related with the network parameters. Two separate supervised ML models are inferred to predict the numerical values of CP and ASE from a given set of network parameters of an arbitrary network configuration. The rest of the paper is organized as follows: the system model for a multi-cell massive MIMO HetNet system is given in Section 2, the ML enabled performance prediction model in Section 3, and Section 4 concludes the paper.

2. System Model

2.1. Network Topology

A HetNet system deployed in a bounded area, $A \subset \mathbb{R}^2$, with randomly placed BSs forming K tiers is considered. Each tier is distinguished by a unique transmit power, density (the average number of BSs per unit area), and antenna configuration providing service to same coverage area in the same frequency band. The spatial locations of the base stations of each tier are represented with a stochastic two-dimensional point process. For the tractable analysis, the spatial locations of the BSs of each tier are modeled with an independent Poisson point process (PPP). $\Phi_k(A) \sim \text{PPP}(\lambda_k)$, $k = 1, 2, K$, where λ_k is the density of the k -th tier, which captures the worst-case scenario [25]. In the given bounded area, L_k numbers of k -th tier BSs are being deployed with a uniform transmitting power, P_k . This work considers $K = 2$, representing macro-base stations (MBS) acting as an umbrella cell, under which many low-powered pico-base stations (PBS) are being deployed in the vicinity of user hotspots in order to provide uniform service. The BSs of each tier are equipped with massive-MIMO transmission systems capable of serving multiple users in same time–frequency grid, considering intra-cell SDMA. Each BS of the k -th tier is equipped with N_k transmit antennas, serving R_k single antenna active users in the same time–frequency grid ($R_k < N_k$). The schematic representation of the given model is presented in Figure 1.

Figure 1. Schematic representation of a typical massive MIMO-enabled two-tier heterogeneous network (HetNet).

2.2. Channel Model

In this work, both tiers are working in a sub 6-GHz band. An Orthogonal Frequency Division Multiplex (OFDM) communication system is assumed, resulting in a flat-fading channel for each sub-carrier; the channel between the i -th antenna in the BS of the m -th cell of k -th tier, and the j -th user associated with the n -th cell of the l -th tier, has been modeled with a single-tap channel coefficient [15].

$$g_{(mi)k}^{(nj)l} = \sqrt{\gamma_{(m)k}^{(nj)l}} h_{(mi)k}^{(nj)l}. \quad (1)$$

The subscript $(mi)k$ and superscript $(nj)l$ represent the index of the transmitting antenna at the BS and user, respectively. $\gamma_{(m)k}^{(nj)l}$ and $h_{(mi)k}^{(nj)l}$ are the coefficients capturing the large-scale and small-scale fading effect of the channel, respectively. $\gamma_{(m)k}^{(nj)l} \in \mathbb{R}^+$ is the combination of the path attenuation, depending on the distance between user and the BS, and the shadowing effect due to the propagation environment. The value is modeled with $10\log_{10}(\gamma_m^n) = (-127.8 - 35\log d_m^n + X_{\sigma^2})\text{dB}$, where d_m^n is the Euclidean distance between the corresponding n -th user and the m -th BS, and X_{σ^2} is a log-normally distributed random variable with zero mean and σ^2 dB variance [26]. $h_{(mi)k}^{(nj)l} \in \mathbb{C}$ is a random variable drawn from an independent Rayleigh distribution, $h_{(mi)k}^{(nj)l} \sim \mathcal{CN}(0, 1)$. Considering the block-fading channel, the expected value of the channel coefficient within the coherence time is given by $E[g_{(mi)k}^{(nj)l}] = \sqrt{\gamma_{(m)k}^{(nj)l}}$. The channel coefficient vector between the BS of the m -th cell in the k -th tier, and the j -th user of the n -th cell in l -th tier is as follows

$$g_{(m)k}^{(nj)l} = \sqrt{\gamma_{(m)k}^{(nj)l}} h_{(m)k}^{(nj)l} \in \mathbb{C}^{N_k \times 1} = \sqrt{\gamma_{(m)k}^{(nj)l}} \left(h_{(m1)k}^{(nj)l}, h_{(m2)k}^{(nj)l}, \dots, h_{(mN_k)k}^{(nj)l} \right)^T. \quad (2)$$

2.3. Interference Analysis

Using a matched filter beamformer, the transmit vector of the BS of m -th cell in k -th tier is as follows

$$x_{(m)k} = \sum_{j=1}^{R_k} \frac{[g_{(m)k}^{(mj)k}]^H}{\|g_{(m)k}^{(mj)k}\|} s^{(mj)k}. \quad (3)$$

Here, $s^{(mj)k}$ is the information symbol intended for the j -th user in same cell. The signal received at the j -th associated user of the n -th cell in the l -th tier is as follows

$$y^{(nj)l} = \sum_{k=1}^K \sum_{m=1}^{L_k} P_k g_{(m)k}^{(nj)k} x_{(m)k} + z \quad (4)$$

Here, z is the additive Gaussian noise, $z \sim \mathcal{CN}(0, \sigma_z^2)$.

$$y^{(nj)l} = \sum_{m=1}^{L_l} P_l g_{(m)l}^{(nj)l} x_{(m)l} + \sum_{k=1, k \neq l}^K \sum_{m=1}^{L_k} P_k g_{(m)k}^{(nj)k} x_{(m)k} + z. \quad (5)$$

The expression of inter-tier interference is as follows

$$I_{inter-tier} = \sum_{k=1, k \neq l}^K \sum_{m=1}^{L_k} P_k g_{(m)k}^{(nj)k} x_{(m)k}. \quad (6)$$

Putting Equation (6) in Equation (5),

$$y^{(nj)l} = P_l g_{(n)l}^{(nj)l} x_{(n)l} + \sum_{m=1, m \neq n}^{L_l} P_l g_{(m)l}^{(nj)l} x_{(m)l} + I_{inter-tier} + z. \quad (7)$$

The expression of inter-cell interference is as follows

$$I_{inter-cell} = \sum_{m=1, m \neq n}^{L_l} P_l g_{(m)l}^{(nj)l} x_{(m)l}. \quad (8)$$

Putting Equations (3) and (8) in Equation (7),

$$y^{(nj)l} = P_l g_{(n)l}^{(nj)l} \sum_{t=1}^{R_l} \frac{[g_{(n)l}^{(nt)l}]^H}{\|g_{(n)l}^{(nt)l}\|} s^{(nj)l} + I_{inter-cell} + I_{inter-tier} + z. \quad (9)$$

The intra-cell interference is as follows

$$I_{intra-cell} = P_l g_{(n)l}^{(nj)l} \sum_{t=1, t \neq j}^{R_l} \frac{[g_{(n)l}^{(nt)l}]^H}{\|g_{(n)l}^{(nt)l}\|} s^{(nj)l} \quad (10)$$

Hence, the received signal,

$$y^{(nj)l} = P_l \|g_{(n)l}^{(nj)l}\| s^{(nj)l} + I_{intra-cell} + I_{inter-cell} + I_{inter-tier} + z. \quad (11)$$

The first term of the r.h.s. of Equation (11) is the weighted value of the intended signal. The signal power to interference plus noise power ratio (SINR) experienced at the receiver is given by the following:

$$SINR = \frac{(P_l \|g_{(n)l}^{(nj)l}\| s^{(nj)l})^2}{I_{total}^2 + \sigma_z^2}. \quad (12)$$

Here, $I_{total} = I_{intra-cell} + I_{inter-cell} + I_{inter-tier}$.

2.4. Performance Metrics

In this work, two fundamental performance metrics are considered as evaluation parameters of the network performance [27].

2.4.1. Coverage Probability (CP)

CP is the measure of the reliability of a typical transmission link, and is defined as the probability that a typical mobile user is able to achieve some threshold SINR (Th), given by the following

$$CP = Pr[SINR > Th]. \quad (13)$$

For successfully running any given application, the user needs to have an SINR value of more than a minimum value. If the SINR experienced by any user drops below the desired minimum value, the customer satisfaction would be compromised. Hence, a higher value of CP implies a better quality of experience (QoE).

2.4.2. Area Spectral Efficiency (ASE)

ASE is a measure of spectral reuse efficiency in the network and is defined as the sum of the average data rates per unit bandwidth normalized with the total service area (bits/sec/Hz/Km²), given by the following,

$$ASE = \frac{\sum_k L_k R_k}{A} E[\log_2(1 + SINR^{(nj)k})]. \quad (14)$$

A higher value of ASE implies a higher achievable sum rate for the network, which allows a greater number of users to get service from the network.

3. Performance Prediction Model

Before deploying a complex HetNet with each tier supporting massive MIMO, network providers are interested in predicting the overall network performance. The analysis provided in the previous section shows that the network performance metrics are functions of various network parameters, such as the number of transmitting antennas at BSs, the number of active users associated per cell, and the transmitted power of each tier. However, the numerical values of the parameters are stochastic variables because of the stochastic network topology and user location. Moreover, for a given topology, the instantaneous performances of the network are also stochastic processes that depend on the stochastic behavior of multipath channel fading coefficients.

In this work, a two-tier network ($K = 2$) is considered—microcell and picocell. In the given scenario, the numerical values of the network performance metrics (PMs), either CP or ASE , are considered to be stochastically influenced by the number of the antennas present in the BS of both tiers, i.e., N_{macro} and N_{pico} ; the number of active users served in same time–frequency grid, i.e., R_{macro} and R_{pico} , for both tiers; and the difference in the transmitted powers of the tiers in dB, i.e., $P_{TD} = 10 \log\left(\frac{P_{macro}}{P_{pico}}\right)$. Thus, the performance metrics are given by the following

$$PM = f(N_{macro}, R_{macro}, N_{pico}, R_{pico}, P_{TD}) + \epsilon. \quad (15)$$

Here, $f(\cdot)$ represents an unknown stochastic function and ϵ represents random noise terms that capture the contributions of the unknown parameters that influence the performance metrics. The objective of the present work is to infer supervised learning models to approximate the unknown stochastic functions for predicting the numerical values of the network performance metrics (PMs). The process is implemented with the following steps, listed below.

3.1. Step 1: Data Preparation

In order to accurately predict the performance metrics for an arbitrary network configuration, the supervised learning model requires a quality dataset that is a collection of instances with input attributes and a leveled output. For the given configuration, the input attributes are N_{macro} , R_{macro} , N_{pico} , R_{pico} , P_{TD} and the output is (PMs; either CP or ASE). The data set is created by running a realistic simulation for a massive MIMO HetNet system with various combinations of network parameters. For a given set of network parameters (input attributes), the simulation is carried out to evaluate the numerical values of the performance metrics given in Equations (13) and (14). The values of the network parameters and the simulated result of the performance metrics constitute a single instance of the dataset.

In each round of simulation with new input attributes, two-tier cellular networks ($K = 2$) were simulated, comprising macro-cells and pico-cells, where the BSs of each tier were massive-MIMO enabled. The threshold (Th) of the SINR value in the CP calculation was taken as 10dB. To capture the stochastic nature of the network topology, for each network configuration, the network was simulated 200 times and the results were averaged out. Each time, the cellular network was deployed on a torus of $(10 \times 10) \text{ km}^2$. To capture the worst-case scenario, the spatial locations of the BSs stations were modeled using two independent PPPs, with the node densities of two layers related as $\lambda_{pico} = 3\lambda_{macro}$. The topology of a typical tow-tier simulated network where the spatial locations of BSs are modeled with two independent PPPs is shown in Figure 2.

For every network topology, the test user was considered to be static, located randomly in the torus. The user was associated to a single BS of any tier that promised the highest average received signal strength as the servicing BS. The signals from all other remaining base stations were considered to be interference. The stochastic channel fading property was captured by simulating each network topology for 1000 coherence time periods with uncorrelated channel coefficients, and the expected values of the outputs were considered.

The simulation was carried for 1450 combinations of these network parameters. The ranges of values of all parameters are given in Table 1.

Figure 2. Topology of a simulated two-tier typical network.

Table 1. Range of values of all input parameters used in the simulation.

Parameters	Range of Values
N_{macro}	50–200
R_{macro}	10–40
N_{pico}	8–20
R_{pico}	4–8
P_{TD}	5–20 dB

The outputs of the simulation were coverage probability (CP) and area spectral efficiency (ASE) in bits/sec/Hz/km². The output of each combination was recorded as a single instance in the dataset. The prepared dataset was used for model learning and testing.

3.2. Step 2: Hypothesis Testing and Model Selection

In the prepared dataset, the initial five columns (N_{macro} , N_{pico} , R_{macro} , R_{pico} , P_{TD}), as shown in Table 2, represent the input attributes, and the remaining two are the model target variables (CP and ASE). The dataset was analyzed to evaluate the best fit hypothesis for our problem, stating the statistical relation between the input attributes and the target variables. The null hypothesis for the given problem was set as, “The input attributions are not linearly related with the target”. The existence of a null hypothesis was tested using correlation analysis. Correlation is a measure of the strength and direction of the relationship between variables ranging between -1 to 1 . The null hypothesis was established if the correlation coefficient was closely bound to zero, indicating no relation between the variables. Table 3 shows the correlation of the input attributes with the targets.

Table 2. Initial five rows of the prepared dataset.

N_{macro}	R_{macro}	N_{pico}	R_{pico}	P_{TD}	CP	ASE
50	40	20	8	10	0.5786	46.8879
50	40	20	8	20	0.8596	42.4943
100	10	8	4	5	0.4207	60.0825
100	10	8	4	15	0.5017	73.2641
150	10	16	4	5	0.4763	74.8983

Table 3. Correlation table.

Attributes	CP	ASE
N_{macro}	0.51	0.59
R_{macro}	−0.58	0.48
N_{pico}	0.15	0.12
R_{pico}	−0.12	−0.11
P_{TD}	0.41	0.18

As the correlation coefficients were not close to zero for any of the inputs, there was sufficient evidence to reject the null hypothesis and to suggest that the PMs (either CP or ASE) showed a linear relationship with the network parameters. Hence, two multivariate linear regression models were inferred to approximate the relations of the network performance metrics (PMs) with the network parameters that will predict the network performance of an arbitrary set of network parameter values outside the training set. Hence, the hypothesis to approximate to an unknown stochastic function is given by the following: [28],

$$\hat{f} = h_{\beta}(m) = \beta^T m. \quad (16)$$

Here, m is the independent input attribute vector given by $[1, N_{macro}, R_{macro}, N_{pico}, R_{pico}, P_{TD}]^T$ and $\beta = [\beta_0, \beta_1, \dots, \beta_5]^T$ is the tunable regression parameter vector.

3.3. Step 3: Training for Best-Fit Model

The model was trained using the dataset with Q i.i.d. instances, given by $\chi = \{m^{(q)}, PM^{(q)}\}_{(q)=1}^Q$, where the q -th instance is represented by the superscript $^{(q)}$. The loss function for optimization is based on the difference between the simulated output, and the hypothesis evaluation is [20] as follows

$$J(\beta|m) = \frac{1}{2Q} \sum_{q=1}^Q [h_{\beta}(m^{(q)}) - PM^{(q)}]^2. \quad (17)$$

The optimal solution corresponds to the values of the regression parameters that minimize the loss function:

$$\hat{\beta} = \arg \min_{\beta} J(\beta|m). \quad (18)$$

The optimum solution is obtained using the gradient-descent algorithm, which starts with a random β , and updates it with the given iterative process:

$$\begin{aligned} & \beta_0 \beta_0 - \frac{\alpha}{Q} \sum_{q=1}^Q [h_{\beta}(m^{(q)}) - PM^{(q)}], \\ & \beta_1 \beta_1 - \frac{\alpha}{Q} \sum_{q=1}^Q [h_{\beta}(m^{(q)}) - PM^{(q)}] N_{macro}^{(q)}, \\ \text{Repeat until convergence} & \{ \beta_2 \beta_2 - \frac{\alpha}{Q} \sum_{q=1}^Q [h_{\beta}(m^{(q)}) - PM^{(q)}] R_{macro}^{(q)}, \\ & \beta_3 \beta_3 - \frac{\alpha}{Q} \sum_{q=1}^Q [h_{\beta}(m^{(q)}) - PM^{(q)}] N_{pico}^{(q)}, \\ & \beta_4 \beta_4 - \frac{\alpha}{Q} \sum_{q=1}^Q [h_{\beta}(m^{(q)}) - PM^{(q)}] R_{pico}^{(q)}, \\ & \beta_5 \beta_5 - \frac{\alpha}{Q} \sum_{q=1}^Q [h_{\beta}(m^{(q)}) - PM^{(q)}] P_{TD}^{(q)}, \} \end{aligned} \quad (19)$$

Here, α is the step size of the algorithm, taken as 0.01. On convergence, CP and ASE are estimated using Equations (20) and (21), respectively,

$$\hat{CP} = \hat{\beta}_{CP}^T m, \quad (20)$$

$$\hat{ASE} = \hat{\beta}_{ASE}^T m. \quad (21)$$

Here, $\hat{\beta}_{CP}$ and $\hat{\beta}_{ASE}$ are the regression parameter vectors after convergence for ASE and CP, respectively.

3.4. Step 4: Model Validation

The acceptability of the hypothesized relationships between variables was evaluated based on the residual errors. The model validation process split the dataset randomly into two sets—the first set consisted of 80% of instances that were used in training the hypothesis, and the second set had the remaining 20% of instances which were used for validating the hypothesis. The generalization ability of the trained regression models is measured using the percentage of error margin and the coefficient of determination, i.e., R^2 values. To ensure the stability of the trained model, k-fold cross validation was implemented, where the model was evaluated k times, with a unique subset for validation. In this work, it was considered to be $k = 5$. Table 4 provides the five-fold validation evaluation results for both models. The consistency in the evaluation results ensures stability in the models.

Table 4. 5-fold cross validation evaluation results.

Model	Evaluation Parameters	k = 1	k = 2	k = 3	k = 4	k = 5
CP	% of error margin	13.18	13.46	13.62	12.08	12.90
	R^2 score	0.7355	0.749	0.7272	0.7797	0.7624
ASE	% of error margin	13.47	12.84	12.35	13.19	13.31
	R^2 score	0.6782	0.692	0.7274	0.6917	0.7531

3.5. Step 5: Final Model Preparation with Complete Dataset

The final model was made to predict the probable output on the new data. The models were finalized by training with the complete available dataset, which was saved for operational use in later cases. In Table 5, the parameter values of the finalized models are tabulated.

Table 5. Parameter values of the finalized models.

Coverage Probability	Area Spectral Efficiency
$\hat{\beta}_0^{CP} = +0.355020$	$\hat{\beta}_0^{ASE} = -12.3185$
$\hat{\beta}_1^{CP} = +0.001934$	$\hat{\beta}_1^{ASE} = +0.4574$
$\hat{\beta}_2^{CP} = -0.010107$	$\hat{\beta}_2^{ASE} = +1.8500$
$\hat{\beta}_3^{CP} = +0.008203$	$\hat{\beta}_3^{ASE} = +1.0308$
$\hat{\beta}_4^{CP} = -0.016797$	$\hat{\beta}_4^{ASE} = -2.5567$
$\hat{\beta}_5^{CP} = +0.013265$	$\hat{\beta}_5^{ASE} = +1.1783$

To visually inspect the performance of the regression process, the actual values vs. predicted values of the finalized model were plotted. Figures 3 and 4 show the actual values vs. values predicted by the finalized model for ASE and CP, respectively, where it could be seen that the model fits linearly with the trend line. The performance of the finalized models has been quantitatively evaluated based on the residue over the predicted output of the dataset, and the results are given in Table 6.

Table 6. Evaluation of the finalized models.

Evaluation Parameters	CP	ASE
% of error margin	12.06	11.32
R^2 score	0.754	0.749

Figure 3. Simulated vs. predicted values of the area spectral density (ASE) in bits/sec/Hz/km² with a trend line.

Figure 4. Simulated vs. predicted values of the coverage probability with a trend line.

4. Conclusions

This paper introduced an ML approach to predict the performance of a MIMO HetNet system considering a multi-cell scenario. The performance metrics considered in this paper are *CP* and *ASE*, which are stochastic functions of the network parameters. Two separate multivariate linear regression models have been trained for the network performance metrics, with network parameters as the input attributes. The generalization ability trained models have been evaluated numerically based on the percentage of the error margin and R^2 score. The error margin for *CP* and *ASE* are found to be 12.06% and 11.32%, respectively, which are within the tolerable range for practical application. The R^2 scores for *CP* and *ASE* are 0.754 and 0.749, respectively, which are closed to 1, and are found to be satisfactorily high. The visual inspection-based performance evaluation is done using actual vs. predicted output plots. For both models, the scatter points linearly match with the respective trend lines, showing the worth of fitness. During practical deployments of 5G and B5G networks, the application of this model could be very valuable in the precise planning of network and capital expenditures. This model would help the network provider to estimate the quality of service for a given network configuration. The study of other parameters that influence the network performance and their inclusion in the model may be considered as a future direction of research.

Author Contributions: S.B., concept and setup preparation, design of system model, text and plot preparation, and review; S.R.S., methodology creation, performance prediction model selection, analysis and simulations supervision, and text editing; V.P., overview of model validation, final model preparation, data preparation, text editing, and review. All authors have read and agreed to the published version of the manuscript.

Funding: The APC was funded by the Bulgarian Science Fund research project KP06-N27/3.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: Authors from Technical University of Sofia were supported by research project KP06-N27/3 “Resource self-configuration and management in ultra-dense networks with user centric wireless access” of the Bulgarian Science Fund.

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

References

1. Yang, P.; Xiao, Y.; Xiao, M.; Li, S. 6G Wireless Communications: Vision and Potential Techniques. *IEEE Netw.* **2019**, *33*, 70–75. [CrossRef]
2. Jungnickel, V.; Manolakis, K.; Zirwas, W.; Panzner, B.; Braun, V.; Lossow, M.; Sternad, M.; Apelfröjd, R.; Svensson, T. The role of small cells, coordinated multipoint, and massive MIMO in 5G. *IEEE Commun. Mag.* **2014**, *52*, 44–51. [CrossRef]
3. Rajoria, S.; Trivedi, A.; Godfrey, W. A comprehensive survey: Small cell meets massive MIMO. *Elsevier J. Phys. Commun.* **2018**, *26*, 40–49. [CrossRef]
4. Larsson, E.; Edfors, O.; Tufvesson, F.; Marzetta, T. Massive MIMO for next generation wireless systems. *IEEE Commun. Mag.* **2014**, *52*, 86–195. [CrossRef]
5. Chataut, R.; Robert, A. Massive MIMO Systems for 5G and beyond Networks—Overview, Recent Trends, Challenges, and Future. *Sensors* **2020**, *20*, 2753. [CrossRef] [PubMed]
6. Zhang, J.; Björnson, E.; Matthaiou, M.; Ng, D.; Yang, H.; Love, D. Prospective Multiple Antenna Technologies for Beyond 5G. *IEEE J. Sel. Areas Commun.* **2020**, *38*, 1637–1660. [CrossRef]
7. SajidHaroon, M.; HaqAbbas, Z.; Muhammad, F.; Abbas, G. Analysis of coverage-oriented small base station deployment in heterogeneous cellular networks. *Elsevier J. Phys. Commun.* **2020**, *38*. [CrossRef]
8. Chien, W.-C.; Cho, H.-H.; Lai, C.-F.; Tseng, F.-H.; Chao, H.-C.; Hassan, M.M.; Alelaiwi, A. Intelligent Architecture for Mobile HetNet in B5G. *IEEE Netw.* **2019**, *33*, 34–41. [CrossRef]
9. Bai, T.; Heath, R.W., Jr. Asymptotic Coverage Probability and Rate in Massive MIMO Networks. 2013. Available online: <http://arxiv.org/abs/1305.2233> (accessed on 30 September 2020).
10. Lim, Y.; Chae, C.; Caire, G. Performance Analysis of Massive MIMO for Cell-Boundary Users. *IEEE Trans. Wirel. Commun.* **2015**, *14*, 6827–6842. [CrossRef]
11. Gao, X.; Edfors, O.; Rusek, F.; Tufvesson, F. Massive MIMO Performance Evaluation Based on Measured Propagation Data. *IEEE Trans. Wirel. Commun.* **2015**, *14*, 3899–3911. [CrossRef]
12. Feng, C.; Jing, Y.; Jin, S. Interference and outage probability analysis for massive MIMO downlink with MF precoding. *IEEE Signal Process. Lett.* **2016**, *23*, 366–370. [CrossRef]
13. Liang, N.; Zhang, W.; Shen, C. An uplink interference analysis for massive MIMO systems with MRC and ZF receivers. In Proceedings of the IEEE Wireless Communications and Networking Conference, New Orleans, LA, USA, 9–12 March 2015. [CrossRef]
14. Li, X.; Björnson, E.; Larsson, E.G.; Zhou, S.; Wang, J. A Multi-Cell MMSE Precoder for Massive MIMO Systems and New Large System Analysis. In Proceedings of the IEEE Global Communications Conference (GLOBECOM), San Diego, CA, USA, 6–10 December 2015; pp. 1–6.
15. Adhikary, A.; Ashikhmin, A.; Marzetta, T. Uplink Interference Reduction in Large-Scale Antenna Systems. *IEEE Trans. Commun.* **2017**, *65*, 2194–2206. [CrossRef]
16. Kusaladharma, S.; Zhu, W.; Ajib, W. Exact Outage Analysis for Stochastic Cellular Networks under Multi-User MIMO. In Proceedings of the IEEE 17th Annual Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, USA, 10–13 January 2020; pp. 1–6.
17. Dhillon, H.S.; Ganti, R.K.; Baccelli, F.; Andrews, J.G. Modeling, and analysis of K-tier downlink heterogeneous cellular networks. *IEEE J. Sel. Areas Commun.* **2012**, *30*, 550–560. [CrossRef]
18. Li, C.; Zhang, J.; Andrews, J.; Letaief, K. Success Probability and Area Spectral Efficiency in Multiuser MIMO HetNets. *IEEE Trans. Commun.* **2016**, *64*, 1544–1556. [CrossRef]

19. Adhikary, A.; Dhillon, H.S.; Caire, G. Massive-MIMO Meets HetNet: Interference Coordination through Spatial Blanking. *IEEE J. Sel. Areas Commun.* **2015**, *33*, 1171–1186. [[CrossRef](#)]
20. Hattab, G.; Cabric, D. Rate-based cell range expansion for downlink massive MIMO heterogeneous networks. *IEEE Wirel. Commun. Lett.* **2018**, *7*, 296–299. [[CrossRef](#)]
21. Nie, X.; Zhang, J.; Zhou, T.; Li, X.; Yao, Y.; Wang, Y. Location-Aware Cross-Tier Cooperation for Massive MIMO Heterogeneous Networks. *IEEE Wirel. Commun. Lett.* **2020**. [[CrossRef](#)]
22. Li, H.; Wang, Z.; Wang, H. Joint user association and power allocation for massive MIMO HetNets with imperfect CSI. *Elsevier J. Signal Process.* **2020**, *173*. [[CrossRef](#)]
23. Ni, S.; Zhao, J.; Yang, H.; Gong, Y. Enhancing Downlink Transmission in MIMO HetNet with Wireless Backhaul. *IEEE Trans. Veh. Technol.* **2019**, *68*, 6817–6832. [[CrossRef](#)]
24. Umer, A.; Hassan, S.; Pervaiz, H.; Musavian, L.; Ni, Q.; Imran, M. Secrecy Spectrum and Energy Efficiency Analysis in Massive MIMO-Enabled Multi-Tier Hybrid HetNets. *IEEE Trans. Green Commun. Netw.* **2020**, *4*, 246–262. [[CrossRef](#)]
25. Stoyan, D.; Kendall, W.S.; Mecke, J. *Stochastic Geometry and Its Applications*, 2nd ed.; Wiley: Hoboken, NJ, USA, 1996.
26. 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Spatial Channel Model for MIMO Simulations (Release 10), Document 3GPP TR 25.996 V10.0.0. April 2011. Available online: https://www.etsi.org/deliver/etsi_tr/125900_125999/125996/10.00.00_60/tr_125996v100000p.pdf (accessed on 30 September 2020).
27. Andrews, J.G.; Baccelli, F.; Ganti, R.K. A tractable approach to coverage and rate in cellular networks. *IEEE Trans. Commun.* **2011**, *59*, 3122–3134. [[CrossRef](#)]
28. Alpaydin, E. *Introduction to Machine Learning*, 3rd ed.; PHI Learning pvt Ltd., The MIT Press: Cambridge, MA, USA; London, UK, 2016.