Fusing Diverse Input Modalities for Path Loss Prediction: A Deep Learning Approach

SOTIRIOS P. SOTIROUDIS[1], PANAGIOTIS SARIGIANNIDIS[2], (Member, IEEE),
SOTIRIOS K. GOUDOS[1], (Senior Member, IEEE),
AND KATHERINE SIAKAVARA[1], (Member, IEEE)

[1]Radiocommunications Laboratory, Department of Physics, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece
[2]Department of Electrical and Computer Engineering, University of Western Macedonia, 50100 Kozani, Greece

Corresponding author: Sotirios K. Goudos (sgoudo@physics.auth.gr)

ABSTRACT Tabular data and images have been used from machine learning models as two diverse types of inputs, in order to perform path loss predictions in urban areas. Different types of models are applied on these distinct modes of input information. The work at hand tries to incorporate both modes of input data within a single prediction model. It therefore manipulates and transforms the vectors of tabular data into images. Each feature of the tabular data vector is spread into several pixels, corresponding to the calculated importance of the particular feature. The resulting synthetic images are then fused with images representing selected regions of the area's map. Compound pseudoimages, having channels of both map-based and tabular data-based images, are then being used as inputs for a Convolutional Neural Network (CNN), which predicts the path loss value at a specific point of the area of interest. The results are clearly better than those obtained from models based on a single mode of input data, as well as from the results produced by other bimodal-input approaches. This approach could be applied for path loss prediction in urban environments for several state-of-art wireless networks like 5G and Internet of Things (IoT).

INDEX TERMS Convolutional neural networks, data to image transformation, deep learning, path loss, pseudoimages, radio propagation.

I. INTRODUCTION

Current state-of-art wireless networks like 5G and IoT require the most accurate path loss calculation, especially in dense urban environments. Ray Tracing modeling of radio propagation has the potential to achieve high accuracy. However, it is prone to slow response time in real environments, due to its lack of computational efficiency. On the other hand, modeling approaches based on artificial intelligence (AI), and particularly on machine learning, provide fast and accurate results, being computationally intensive only at the stage of their initial training. While most of them use tabular data at their inputs, models based on images are also gaining momentum.

Those of the former category can be said to belong in the "classical" machine learning domain and implement techniques such as Support Vector Machines [1], Multilayer Perceptron Neural Networks [2], KNN [3], Random Forest [4],

The associate editor coordinating the review of this manuscript and approving it for publication was Mohammad Tariqul Islam.

ANFIS [5] etc. This category of models uses data of tabular format at their inputs. The second category of models, which delivers its predictions based on images, is generally being referred to as Deep Learning. CNNs [6]–[9] are used in order to extract information from images and perform predictions.

The work at hand presents a framework which facilitates the integration of both types of inputs (tabular data and images). The resulting dual-mode model reaps the benefits of both input types and delivers more reliable results.

More specifically, our work provides a method to transform a vector of tabular data into an image. The importances of the vector's features are used in order to effectively populate the pixels of the image. This artificial image, along with images from the area's map, is going to be integrated into a compound pseudoimage, which will facilitate the simultaneous introduction of tabular data and image data into a deep learning model.

The proposed bimodal approach delivers a MAE value of 3.07 dB, while the models that rely on a single source of input information reach errors of 3.32 dB (when using

images) or 3.26 dB (for tabular data). Moreover, our method outperforms other state-of-the-art fusion frameworks by at least 0.08 dB (3.07 dB versus 3.15 dB).

The contributions of our work can be summarized as follows:

1. The concept of compound pseudoimages, encapsulating tabular data and images of the urban area's built-up footprint, is introduced.

2. A CNN regressor, capable of processing the pseudoimages and producing better results than those obtained from the single-mode models, is being fully deployed.

3. A comparison with other dual-mode approaches, in terms of fusion implementation and performance gain, highlights the advantages of the proposed approach.

The rest of the paper has the following structure: Section II presents the current state of the art, while Section III analyzes the proposed method. The results are presented in Section IV and further discussed in Section V. The conclusions are contained in Section VI.

II. RELATED WORK

The need for accurate predictions of propagation path loss has driven the research among various machine learning algorithms [1]–[9]. Substantial efforts have been made in order to optimally define their internal configuration [10]–[12], in an attempt to achieve the best results possible.

However, it is the input of a machine learning algorithm that lies at the core of its implementation and mostly controls its performance [2], [13]. The manmade characteristics (such as buildings and roads) of an urban area determine heavily the propagation profile for a given wireless communications link [2]. It is therefore straightforward to claim that machine learning models need detailed information regarding the built-up profile of the area, in order to provide accurate predictions [13]. This information could be either provided through tabular data, or images.

A. TABULAR DATA-BASED MODELS

In the tabular data case, inputs describing the Line of Sight path [14], the area around the receiver [15], or the whole map of the area under investigation [16] can be found in the literature. Some of the models rely on mean values of the built-up characteristics [5], [10], while others use more detailed input information, corresponding to specific regions of the area [14], [15].

Each feature vector is then associated with the corresponding path loss value. Machine learning algorithms [1]–[5] are trained on sets of these pairs, in order to become able to predict the path loss values for vectors that they have not seen during their training.

Feature engineering determines the vector's input and consequently dictates the prediction's success for the tabular data-based models [17]. The manual extraction of features is nevertheless a time and effort consuming process, without any guarantee regarding its possible success [6], [8].

B. IMAGE-BASED MODELS

While the classical machine learning models that depend on tabular data have already proved their value, a new category of models, based on deep learning and using images as their inputs, has emerged [6]–[9], [18].

Their fundamental advantage is that the preprocessing step of producing handcrafted features can be skipped. That is, these models are able to directly extract the information they need from appropriate images of the area under investigation. Their internal architecture can therefore be broken down into two distinct and consecutive units that respectively perform the tasks of: a) feature extraction, and b) regression.

Satellite images [6], [7], as well as vector images representing the built-up area footprint [8], [18], have been used at the inputs of image-based prediction models. It has been concluded that both image types lead to similar results, with the benefit of reduced model complexity for the case of vector images [18]. Moreover, the images may differ with regard to the boundaries of the represented area: they could either depict the whole area between the transmitter and the receiver, or a particular area of predefined size around the receiver [8], [18]. Better results have been reported for the whole area depiction [18].

C. MODALITY FUSION TECHNIQUES

The co-existence of diverse modes of input information naturally raises the question of their combination. That is, a framework facilitating the collaboration of the multiple input representations is needed in order to profit from every type of information and provide an enhanced final prediction [19]. Two are the state-of-the-art techniques that can help towards this direction: a) model ensembling through stacked generalization and b) feature concatenation inside a CNN.

Stacked generalization [20] can be conceived as a general-purpose ensembling technique. It provides a way to combine heterogeneous regressors (base learners), through the proper combination of their individual predictions from a second-order meta-model. That is, the meta-regressor uses the predictions of the base learners as input features in order to make its own, final predictions. Within that context, stacked generalization can serve as a modality fusion technique, provided that the base learners are grown based on diverse input modes. It should however be mentioned that fusion takes place indirectly: the distinct modes are mixed through the combination of their corresponding models' results.

Feature concatenation [18], [19] on the other hand can be performed inside a CNN: that is, two distinct branches, respectively extract features from the tabular data and the images. These features are then concatenated and fed into the CNN's regressor part, which performs the predictions. Though fusion happens earlier as compared to stacked generalization, it must be noted that features are being independently extracted. This means that the influence of a modality's features might suppress the contribution of the
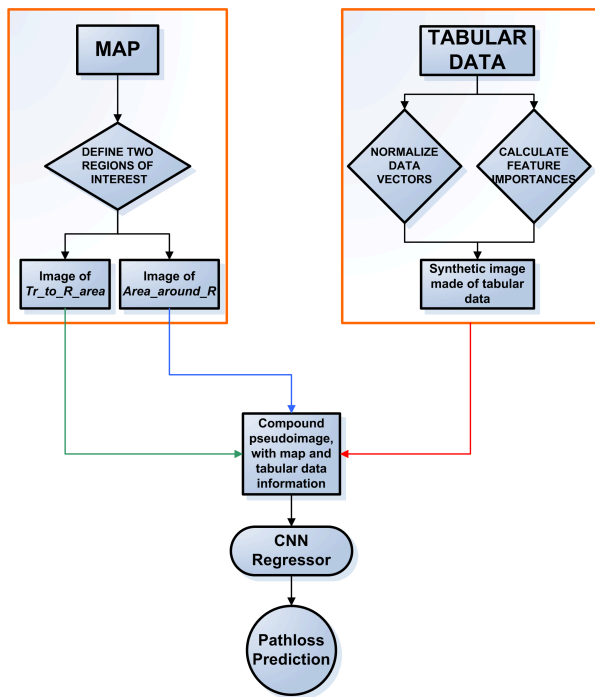
FIGURE 1. Three monochrome images, made from the map and the tabular data, act as the channels of a compound pseudoimage. The CNN performs its predictions based on both sources of information.

FIGURE 2. Creation of images corresponding to the two regions of interest: (a) The positions of the Transmitter, Tr, and the Receiver, R, within the map. (b) The orthogonal area between Tr and R. (c) The rectangular area around R. (d) Transformation of the orthogonal area into a rectangle. (e) Both rectangles have the same size.

other modality. Feature concatenation has been applied to path loss prediction problems in [7] and [18], respectively resulting to error rates of 1 dB (for 60000 training instances of a university campus) and 6 dB (for a set of 125000 instances from inherently different measuremet campaigns).

Our approach investigates modality fusion at the input level, through the creation of pseudoimages: concurrent processing of images and tabular data is facilitated through their integration into a compact, though bimodal, input.

III. PROPOSED METHOD

The workflow of our approach is presented in Figure 1. Two images, taken from two regions of interest from the raster map of the area, occupy two of the pseudoimage's channels. The raster map is a grayscale image which contains information regarding the footprints and the heights of the area's buildings. More precisely, the tallest buildings have the darkest colors, while the roads are white. The mapping between building height and grayscale value is linear. This information can be obtained through either a building database of the area, or through OpenStreetMap [21] (for the footprints) in conjunction with the open source software QGIS (for the building heights).

The third channel of the pseudoimage is being filled with a synthetic image, which acts as the carrier of the tabular data information. Feature importances are used in order to assign more space to the most dominant features.

The pseudoimage contains therefore information from both the area's map and the vectors of tabular data, acting as a single, though compound, source for the CNN regressor who then performs the path loss predictions.
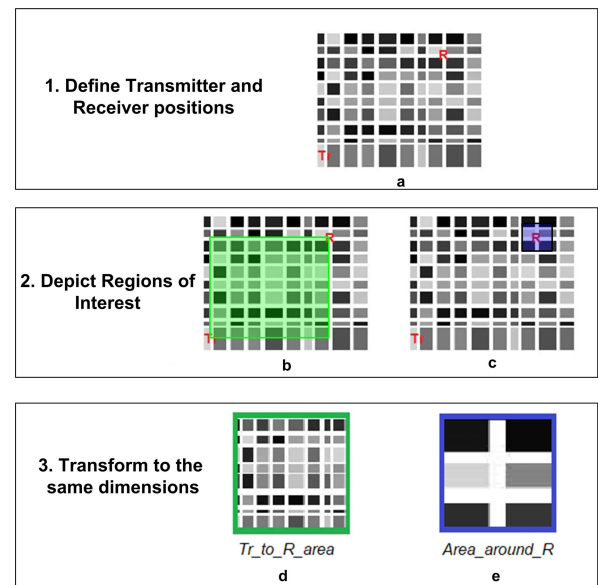
An analysis of all the individual steps is being carried out through the remainder of the current section.

A. RASTER IMAGES REPRESENTING MAP INFORMATION

Figure 2 shows a way to extract two equisized images from the map of the investigated area. The regions of interest are defined with regard to the positions of the transmitter and the receiver. The first region is the orthogonal area bounded from the transmitter at its down-left corner and the receiver at its top-right one [8] (2b) and includes the Line of Sight path, which has been shown to strongly influence the level of received power [4], [14]. This area is then transformed to a rectangle of predefined size (2d). The built-up profile of the area around the receiver is also influencing path loss [15]. Thus, the second region of interest is a 40m wide rectangular area having the receiver at its center (2c), depicting its closest roads and buildings. It is transformed to the same dimensions (2e), with the previously extracted image. The area near the transmitter, when he is placed well above the surrounding rooftops, is of negligible influence.

B. IMAGES MADE OF TABULAR DATA

Tabular data is organized in the form of vectors. Each vector contains values of a number of features (or predictors), as well as the corresponding target's value (which is the path loss value for our case). Our goal is to create a synthetic image from each data vector, which will still hold the information contained within the individual features.

An approach towards this direction is the one introduced in [22], where each data vector was multiplied by its own transposition, resulting to a data matrix. This matrix was then represented by a rectangular gray pixel image, whose dimension (in pixels) was equal to the number of features held by the data vector.
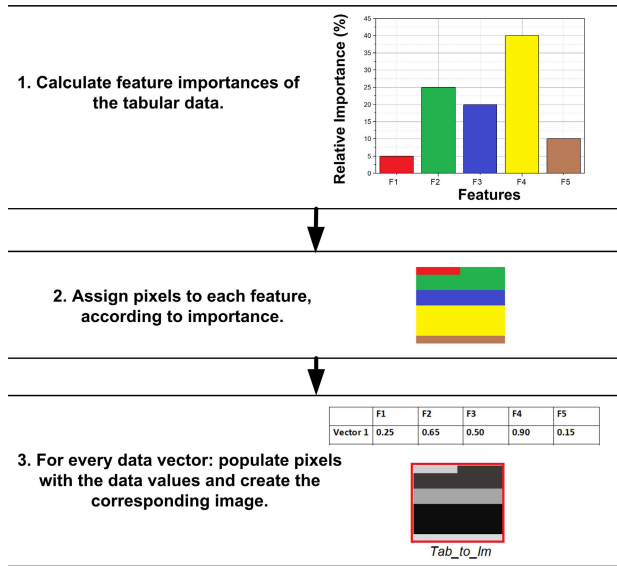
FIGURE 3. Transforming a tabular data vector, containing five features, to a synthetic image.
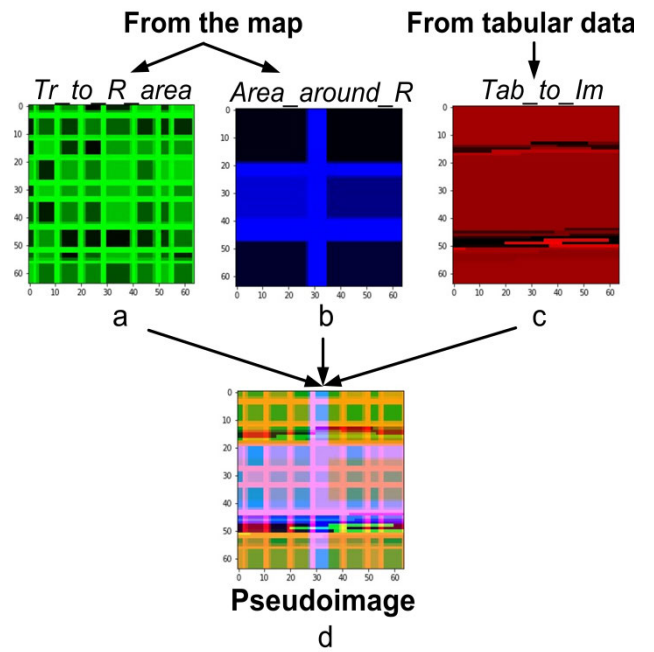

FIGURE 4. Assembling a pseudoimage (a,b,c) Three monochrome images occupy the three channels of the pseudoimage. (d) All types of input data have been fused within the pseudoimage.

Our approach is capable of creating an arbitrary sized image, making use of the feature importances that a wide variety of machine learning models are able to calculate. That is, the percentage of the synthesized image's pixels that correspond to a certain feature, is dictated through the importance of the particular feature. The values of the pixels are then populated from the normalized values of the corresponding features.

More precisely, the first step of our approach is the normalization of the raw data in the [0,1] range, in order to eliminate the differences stemming from the different scales that the numerical features use. Using the notation x_i for the *i-th* numerical feature of each raw data vector, we can produce the normalized values x_i of each feature according to the following equation:

$$x_i' = \frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)} \quad (1)$$

where $min(x_i)$ and $max(x_i)$ are respectively the minimum and the maximum values of each feature.

Besides the normalization of the data vectors, another step towards the data-to-image transformation is the calculation of the individual feature importances. An extensive analysis of the concept of feature importances can be found in [23]. Briefly speaking, the importance of each feature (when tree-based models are concerned) is being denoted from its participation in the process of node-splitting during the growth of the individual trees. As a result, features are being assigned a percentage, which depicts their relative importance. Our approach uses these importances in order to assign pixels for each feature when transforming tabular data to an image. An example is shown in Figure 3: the size of the image's patches corresponds to the importance of its features, while their greyscale values reflect the data vector's values.

C. PSEUDOIMAGES CONTAINING MAP AND TABULAR DATA

Our goal is to use the images taken from the map, as well as the image, which was produced from the tabular data, inside a standalone prediction model. We have therefore chosen to use the three aforementioned monochrome images as independent channels of a new image. This pseudoimage encapsulates all the previously deployed sources of information (Figure 4) and therefore paves the way for their concurrent processing.

D. CNN-BASED REGRESSION FOR PATH LOSS PREDICTION

CNNs have the ability to perform predictions, based on images of the area's as map [6]–[9]. The same holds true when these networks receive appropriate pseudoimages at their inputs.

The way that CNNs operate [24], could be broken down into two parts: the first part extracts features from the input images, while the second part uses the extracted features in order to perform predictions. Feature extraction is being operated through a series of convolutional and pooling layers, while predictions are conducted from a set of fully connected layers. Figure 5 presents a CNN regressor's architecture.

The evaluation of the prediction performance is quantitively reflected into the values of error functions, such as the mean absolute error (MAE), the mean absolute percentage error (MAPE) and the Root Mean Squared Error (RMSE).

Their mathematical expressions are given in the following equations:

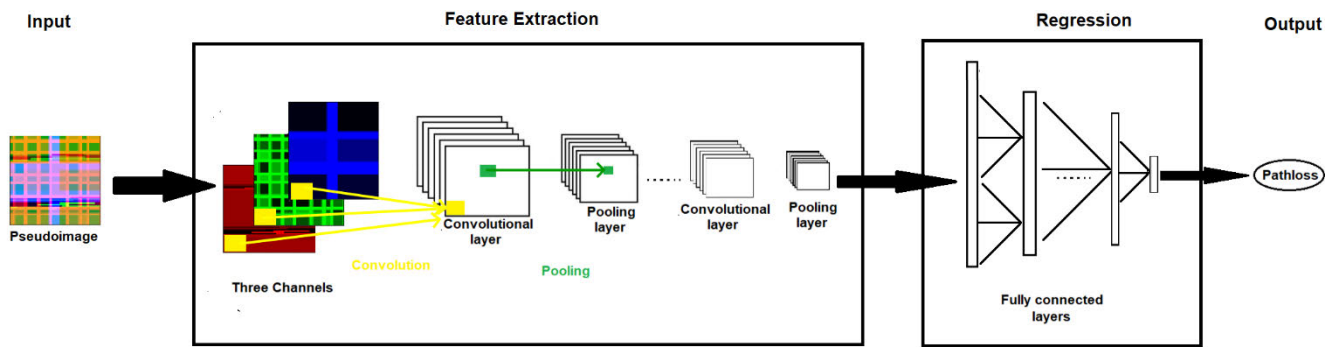$$MAE = \frac{1}{N} \sum_{p=1}^{N} |t(p) - y_o(p)| \quad (2)$$

FIGURE 5. The pseudoimage at the input of the CNN facilitates the concurrent extraction of features based on images and tabular data, thus performing the pathloss prediction based on both modes of input data.

$$MAPE = \frac{1}{N} \sum_{p=1}^{N} \left| \frac{t(p) - y_o(p)}{t(p)} \right| \cdot 100\% \qquad (3)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{p=1}^{N} [t(p) - y_o(p)]^2} \qquad (4)$$

N denotes the total number of test patterns, $t(p)$ is the calculated (from the Ray Tracing software) path loss value for the pth test pattern and $y_0(p)$ is the path loss value that the machine learning model predicts for the same pattern. According to the machine learning model used, the test pattern represents either a tabular data vector (when using a "classical" machine learning method) or an image (for the CNN case). The lower the values of the above indicators are, the better the precision of prediction is.

IV. NUMERICAL RESULTS
A. PROBLEM DESCRIPTION
A total of 146 areas, each consisting of 100 buildings, were created within a special software [25], implementing the Ray Tracing [26] technique for the calculation of path loss. The heights of the buildings were randomly distributed between 5 m and 29 m, while the transmitter was always placed at the rooftop of the first bottom-left building, at a height of 35m. The lengths and the widths of the roads were also randomly defined. The frequency of operation was set equal to 900 MHz.

The path loss was calculated in the roads and the crossroads of all the areas, producing a total of 35395 path loss values. Each of these values was associated to a specific point for each of the 146 maps, as well as to a vector of tabular data, consisting of 23 numerical features.

An analytical presentation of the numerical features lies beyond the scope of our current work. An interested reader could refer to our previous work [27] for more details about the features. In summary, ten features describe the Line of Sight path, while eight of them are gathered from the receiver's vicinity; finally, five features describe the positions of the transmitter and the receiver, along with the distance between them.

The 35395 values, along with the corresponding images taken from the map and created from the tabular data vectors,

TABLE 1. Input types and model parameters.

Input type	Machine Learning model	Parameters
Tabular data	XGBoost	n_estimators=700, learning_rate=0.1, colsample_bytree=0.9, subsample=0.9, max_depth=8
Images (*Tr_to_R_area*)	CNN	Convolutional layers=4, Filter numbers=16,32,64,128 Filter size=(3,3) Pooling fiter size= (2,2) Fully Connected layers=3 Nodes=32,8,1
Pseudoimages		

were randomly split into three disjoint sets, according to a 60/20/20 ratio, producing the training, validation and testing sets. The training and validation sets were used to fit and fine-tune the parameters of the corresponding machine learning models, while their final performance was measured through their predictions on the testing set.

B. BENCHMARKS AND PROPOSED MODEL
We first evaluate the performance of models that make their predictions either based only on tabular data, or on images. Each model's parameters (shown in Table 1) were tuned via exhaustive Grid search, using the GridSearchCV instance of the scikit-Learn [28] open source library, implemented in Python language. XGBoost [29] was used to process the tabular data, while a CNN [24] was employed for the images (*Tr_to_R_area*). The same CNN was also used for the pseudoimages of the proposed model.

The size of all the (pseudo)images was set equal to 64×64 pixels. A bigger image size has been shown able to produce better results [8] and could therefore be tried; however, this work focuses on the improvement brought by combining images and tabular data for a given image size and network configuration.

In order to use the same CNN architecture for both the monochrome images and the colored pseudoimages, we have transformed the monochrome images to colored ones, by simply making three copies of the same channel.

Figure 6 presents the process of fine-tuning the CNN: it can be seen that for the first 100 epochs the training error keeps
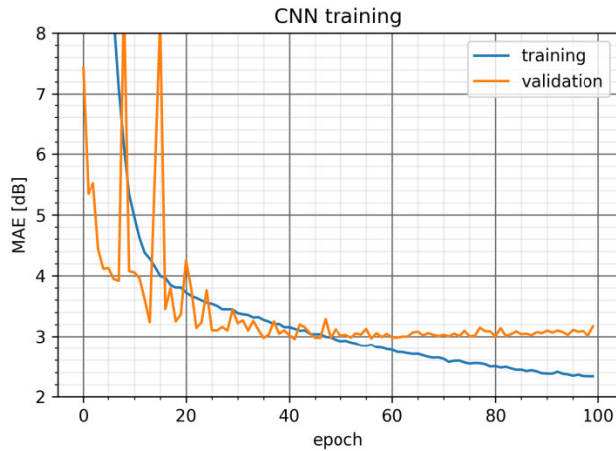
FIGURE 6. Training and validation error for 100 epochs.

on becoming smaller (due to overfitting), while the validation error remains practically stable after a number of epochs. The validation error reached its minimum value (2.96 dB) at the 42th training epoch. The CNN's weights at this epoch, were used in order to produce its estimations for the testing set.

C. COMPARATIVE RESULTS

Figure 7 presents the MAPE, MAE and RMSE values for the single mode approaches (based either only on images or on tabular data), as well as for the proposed bimodal approach which the pseudoimages facilitate. It is clear that the proposed approach superceeds the single mode inputs, profiting from the combination of the available data types.

The series of diagrams included within Figure 8, depict the comparison between the actual and the predicted path loss values, as well as the statistical distribution of the error values and the absolute error values.

A comparison among the models' predictions for a randomly chosen urban area, is presented through Figure 9. It can be seen that the contour plot that was produced according to the pseudoimage method, is the most similar to the one that relies on Ray Tracing's calculations.

D. OTHER BIMODAL APPROACHES

Besides testing against its unimodal counterparts, we have also compared our method with the state-of-the-art bimodal approaches of stacked generalization and feature concatenation, that were presented in Section II C.

Figure 10 highlights the architectural differences among the three bimodal approaches, with regard to the implementation of modality fusion. These differences are further explained in Table 2. The pseudoimage approach facilitates an early fusion at the input level, while feature concatenation fuses the different modalities after feature extraction. The ensembling technique of stacked generalization provides an indirect fusion approach, through the combination of the unimodal models' results.

We have implemented the aforementioned approaches in order to compare their results with those obtained from
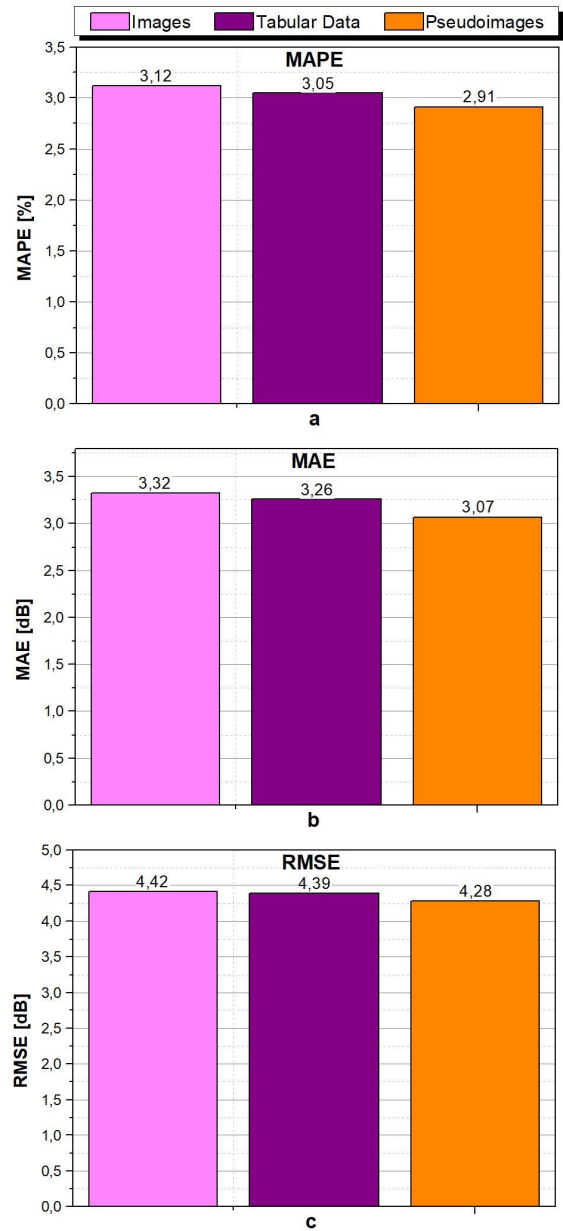

FIGURE 7. The combination of images and tabular data, via the compound pseudoimages, leads to predictions with enhanced precision. (a) Comparing MAPE values (b) Comparing MAE values (c) Comparing RMSE values.

the pseudoimage-based method. We have therefore used the StackingRegressor instance of the Mlxtend [30] package in order to deploy the stacked generalization ensemble. Its base learners were the unimodal models of Table 1, while a linear regressor was chosen as a meta-model, according to [31].

Feature concatenation was implemented by adding a second feature extraction branch [7], dedicated to the processing of the tabular data input within the CNN. After parameter tuning by exhaustive Grid search, the branch was defined to be consisting of two fully connected layers, having 23 and 16 nodes respectively.
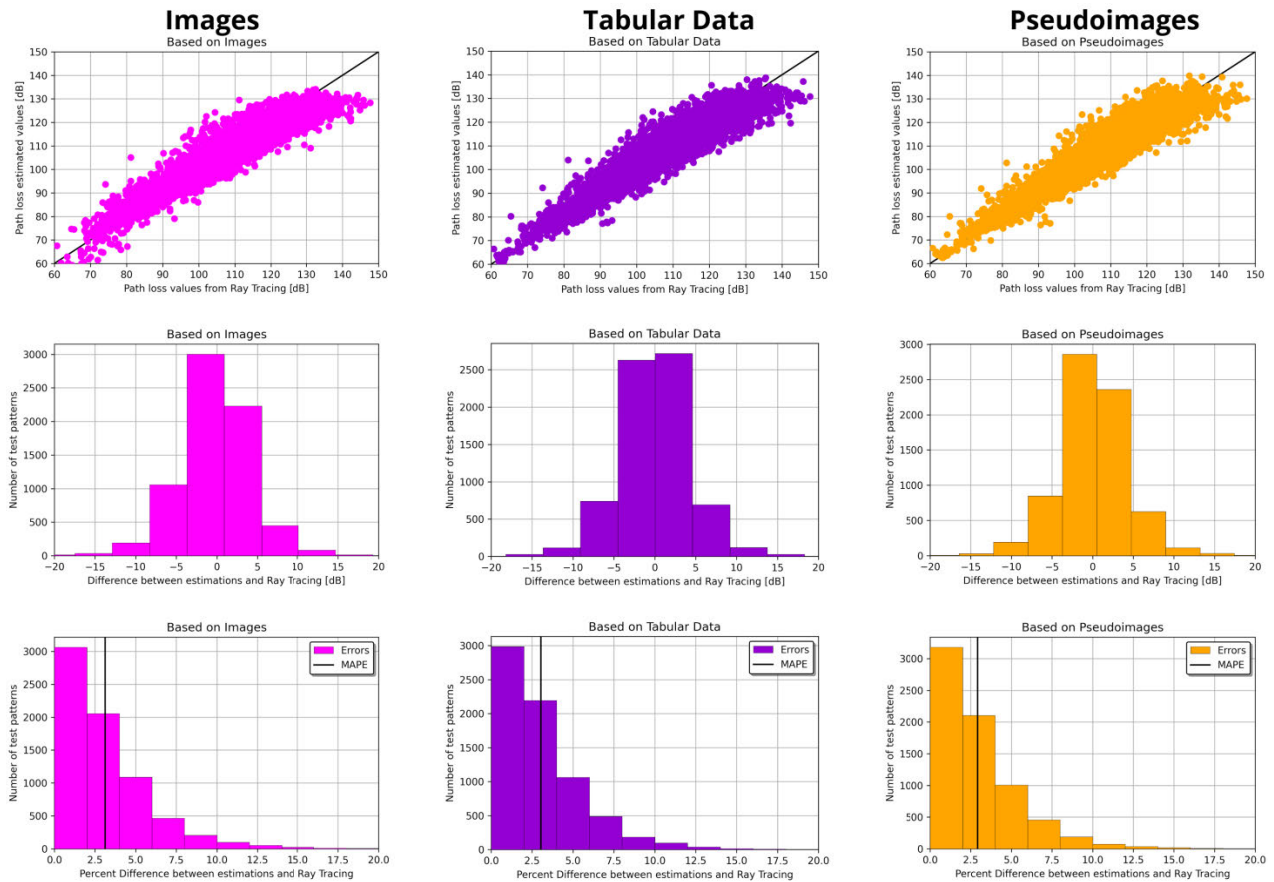
FIGURE 8. Scatter plots, error histograms and absolute error histograms for the three different types of inputs.

Table 3 presents the error metrics for the three different bimodal approaches (stacked generalization, feature concatenation and pseudoimages), along with the unimodal ones. It can be seen that the proposed approach provides better results according to all error metrics. Moreover, all bimodal approaches outperform those that rely on a single input modality. A final remark is that the earlier the fusion happens, the better the results are. That is, an early integration of the available sources of information leads to the most efficient prediction model [19].

V. DISCUSSION

A. EVALUATING CHANNEL CONTRIBUTIONS

In order to evaluate the contribution of the individual channels to the prediction performance, we have examined the cases where each image was fed to the CNN in the absence of the other two, as well as the cases where two images formed the CNN's input: more specifically, we have converted each of the three grayscale images into colored ones, so as to either fill all three channels from three copies of the same image, or to use one of them for two of the channels and fill the third channel with a second image.

The three first cases of Table 4 correspond to using each monochrome image as a colored one. Case_1 has already been presented in the results' section. It is clear

that *Tr_to_R_area* is important for the path loss prediction. The same holds true for case_3. It is clear that tabular data, though having been converted to an image, enables trustworthy predictions. Comparing case's 3 MAE value with that obtained when processing the tabular data in its original form with the XGBoost method, we notice a small performance drop (3.34 dB against 3.26 dB). This means that tabular data information remains almost intact despite its transformation to an image. This small performance drop can be tolerated because it facilitates the creation of pseudoimages and the benefits associated with them.

Case_2 leads to a total performance degradation. This is due to the fact that the image of the area around the receiver, on its own, cannot provide sufficient information for reliable predictions. When combined with another image, the *Area_around_R* offers a small improvement, as seen when comparing case_4 to case_1 or case_5 to case_3.

Case_4 shows that map information (through *Tr_to_R_area* and *Area_around_R*) cannot surpass the 3.27 dB barrier. However, the combination of *Tab_to_Im* and *Tr_to_R_area* (case_6 and case_7) offers a significant performance boost, proving that these two images are the most useful ones. The addition of the *Area_around_R* leads to case_8 which exhibits the best performance, stemming from the concurrent employment of all three images. It can therefore be clearly
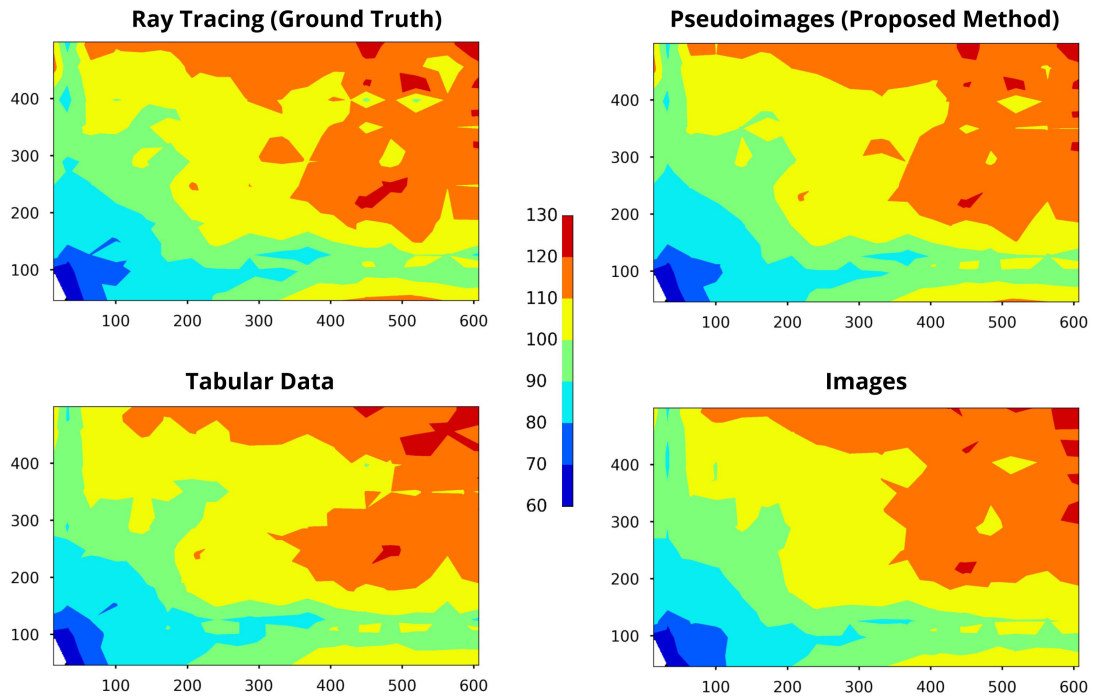
FIGURE 9. Path loss predictions for a randomly chosen urban area: the proposed bimodal method superceeds the unimodal ones.
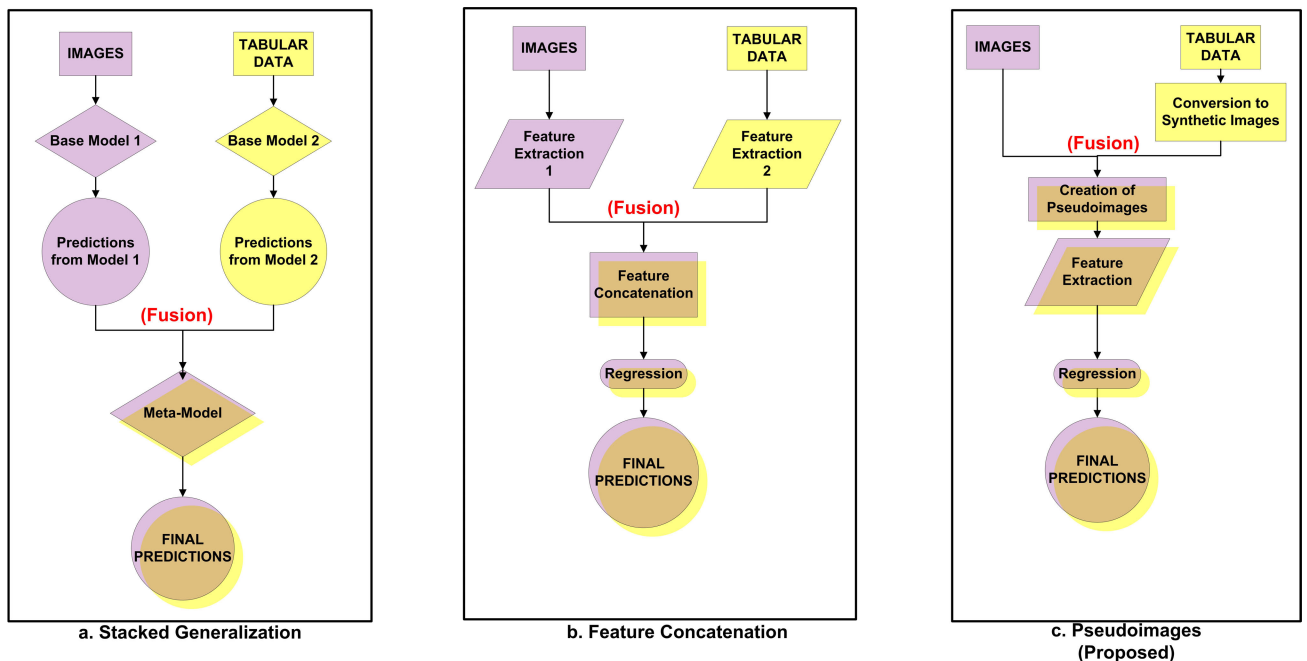

FIGURE 10. Architecture of the three modality fusion approaches.

concluded that the coexistence of the three different channels offers the best results.

B. TABULAR DATA AND PERFORMANCE GAIN

The current paragraph highlights the performance improvement brought by the injection of tabular data, by comparing the CNN's performance with and without its presence.

As Figure 11 implies, the incorporation of either a single image (*Tr_to_R_area*), or even the addition of a second one (*Area_around_R*), lead respectively to error values of 3.32 and 3.27 dB.

However, the injection of transformed tabular data (*Tab_to_Iim*) at the third channel of the pseudoimage, provides an error drop of 0.2 dB, thus offering a MAE value equal to 3.07 dB.

TABLE 2. Comparative description of the modality fusion approaches.

Method	Fusion Implementation	Description
Stacked Generalization	At the combination of the base models' predictions.	A meta-model combines the predictions of the base models, through the ensembling technique of stacked generalization. Each base model has its own unimodal input.
Feature Concatenation	At the features' level.	Two distinct input branches extract independently features from the two input modalities. The features are then concatenated and used from the model's regression unit.
Pseudoimages (Proposed)	At the inputs' level.	Pseudoimages, containing distinct channels occupied either from images or from transformed tabular data, are the inputs of the model. Features are jointly extracted and fed to the model's regression unit.

TABLE 3. Results of all the unimodal and bimodal approaches.

Input type	Fusion Method	MAE [dB]	MAPE (%)	RMSE [dB]
Unimodal (images)	No	3.32	3.12	4.42
Unimodal (tabular data)	No	3.26	3.05	4.39
Bimodal (tabular data + images)	Stacked Generalization	3.23	3.02	4.37
	Feature Concatenation	3.15	2.97	4.32
	Pseudoimages (proposed)	**3.07**	**2.91**	**4.28**

TABLE 4. Pseudoimage configurations and their results.

Case	Ch_1	Ch_2	Ch_3	MAE [dB]
1	Tr_to_R_area	Tr_to_R_area	Tr_to_R_area	3.32
2	Area_around_R	Area_around_R	Area_around_R	8.05
3	Tab_to_Im	Tab_to_Im	Tab_to_Im	3.34
4	Tr_to_R_area	Tr_to_R_area	Area_around_R	3.27
5	Area_around_R	Area_around_R	Tab_to_Im	3.28
6	Tab_to_Im	Tab_to_Im	Tr_to_R_area	3.13
7	Tab_to_Im	Tr_to_R_area	Tr_to_R_area	3.12
8	Tr_to_R_area	Area_around_R	Tab_to_Im	3.07

FIGURE 12. Two ways to transform tabular data to images: (a) Treating all features as equally important (b) Giving more space to the most important features.
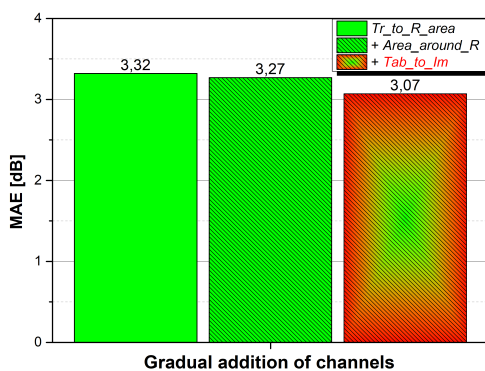

FIGURE 11. The injection of tabular data offers a performance enhancement of 0.2 dB.

C. THE ROLE OF FEATURE IMPORTANCES

Feature importances have been used in order to define the percentage of pixels occupied from each feature. However, the transformation of tabular data to images could take place without considering feature importances: that is, all pixels could be equally distributed among the available features.

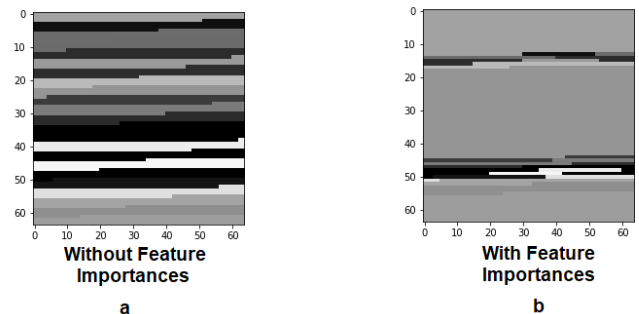This approach would lead to images having 178 pixels for 21 of the features, while two of them would have 179 pixels

(for images with 64 × 64 pixels). A comparison among the images produced with, and without, the consideration of feature importances is shown in Figure 12.

In order to examine the usefulness of the feature importance consideration, we have replaced the (feature importance dependent) corresponding channel of the pseudoimages, with its counterpart which treats all features as equally important. It turned out that the prediction performance became a little worse, resulting to a MAE value of 3.12 dB (as opposed to the 3.07 dB of the original setting).

This means that the CNN was actually helped from seeing more pixels of the most important features. However, even the simplistic approach of considering all features as equally important and presenting them as such in the CNN, is still offering a considerable enhancement as compared to the case where only map information was presented to the CNN.

Moreover, we have randomly changed the order of features, thus changing the relative positions of the corresponding patches inside the Tab_to_Im channel. A negligible fluctuation (of 0.005 dB) was observed within a set of ten experiments.

VI. CONCLUSION AND FUTURE WORK

The integration of images and tabular data in the form of compound pseudoimages has facilitated the estimation of path loss with a higher degree of precision, in comparison with the single-mode respective models. Moreover, the proposed approach led to better results in comparison with other methods of modality fusion, resulting in a MAE value of 3.07 dB against the 3.15 dB of the established bimodal approaches.

The presented framework is general and can therefore incorporate changes and/or additions within the tabular data or the images that constitute its individual channels. Furthermore, transfer learning and image augmentation techniques can be applied in order to further profit from the introduction of the pseudoimages and the deep-learning-based approach that was introduced in the work at hand.

REFERENCES

[1] M. Uccellari, F. Facchini, M. Sola, E. Sirignano, G. M. Vitetta, A. Barbieri, and S. Tondelli, "On the use of support vector machines for the prediction of propagation losses in smart metering systems," in *Proc. IEEE 26th Int. Workshop Mach. Learn. Signal Process. (MLSP)*, Sep. 2016, pp. 1–6, doi: 10.1109/MLSP.2016.7738887.

[2] S. P. Sotiroudis and K. Siakavara, "Mobile radio propagation path loss prediction using artificial neural networks with optimal input information for urban environments," *AEU-Int. J. Electron. Commun.*, vol. 69, no. 10, pp. 1453–1463, Oct. 2015, doi: 10.1016/j.aeue.2015.06.014.

[3] Y. Zhang, J. Wen, G. Yang, Z. He, and X. Luo, "Air-to-air path loss prediction based on machine learning methods in urban environments," *Wireless Commun. Mobile Comput.*, vol. 2018, Jun. 2018, Art. no. 8489326. Accessed: Jul. 27, 2020. [Online]. Available: https://www.hindawi.com/journals/wcmc/2018/8489326/

[4] S. P. Sotiroudis, S. K. Goudos, and K. Siakavara, "Neural networks and random forests: A comparison regarding prediction of propagation path loss for NB-IoT networks," in *Proc. 8th Int. Conf. Mod. Circuits Syst. Technol. (MOCAST)*, May 2019, pp. 1–4, doi: 10.1109/MOCAST.2019.8741751.

[5] N. Faruk, N. T. Surajudeen-Bakinde, A. Abdulkarim, S. I. Popoola, A. Abdulkarim, L. A. Olawoyin, and A. A. Atayero, "ANFIS model for path loss prediction in the GSM and WCDMA bands in urban area," *ELEKTRIKA-J. Electr. Eng.*, vol. 18, no. 1, pp. 1–10, Apr. 2019, doi: 10.11113/elektrika.v18n1.140.

[6] H. F. Ates, S. M. Hashir, T. Baykas, and B. K. Gunturk, "Path loss exponent and shadowing factor prediction from satellite images using deep learning," *IEEE Access*, vol. 7, pp. 101366–101375, 2019, doi: 10.1109/ACCESS.2019.2931072.

[7] J. Thrane, D. Zibar, and H. L. Christiansen, "Model-aided deep learning method for path loss prediction in mobile communication systems at 2.6 GHz," *IEEE Access*, vol. 8, pp. 7925–7936, 2020, doi: 10.1109/ACCESS.2020.2964103.

[8] S. P. Sotiroudis, S. K. Goudos, and K. Siakavara, "Deep learning for radio propagation: Using image-driven regression to estimate path loss in urban areas," *ICT Exp.*, vol. 6, no. 3, pp. 160–165, Sep. 2020, doi: 10.1016/j.icte.2020.04.008.

[9] X. Zhang, X. Shu, B. Zhang, J. Ren, L. Zhou, and X. Chen, "Cellular network radio propagation modeling with deep convolutional neural networks," in *Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, New York, NY, USA, Aug. 2020, pp. 2378–2386, doi: 10.1145/3394486.3403287.

[10] S. P. Sotiroudis, S. K. Goudos, K. A. Gotsis, K. Siakavara, and J. N. Sahalos, "Optimal artificial neural network design for propagation path-loss prediction using adaptive evolutionary algorithms," in *Proc. 7th Eur. Conf. Antennas Propag. (EuCAP)*, Apr. 2013, pp. 3795–3799.

[11] S. P. Sotiroudis, S. K. Goudos, K. A. Gotsis, K. Siakavara, and J. N. Sahalos, "Application of a composite differential evolution algorithm in optimal neural network design for propagation path-loss prediction in mobile communication systems," *IEEE Antennas Wireless Propag. Lett.*, vol. 12, pp. 364–367, 2013, doi: 10.1109/LAWP.2013.2251994.

[12] S. I. Popoola, E. Adetiba, A. A. Atayero, N. Faruk, C. T. Calafate, and C. Yuen, "Optimal model for path loss predictions using feed-forward neural networks," *Cogent Eng.*, vol. 5, no. 1, Jan. 2018, Art. no. 1444345, doi: 10.1080/23311916.2018.1444345.

[13] S. I. Popoola, A. Jefia, A. A. Atayero, O. Kingsley, N. Faruk, O. F. Oseni, and R. O. Abolade, "Determination of neural network parameters for path loss prediction in very high frequency wireless channel," *IEEE Access*, vol. 7, pp. 150462–150483, 2019, doi: 10.1109/ACCESS.2019.2947009.

[14] M. Piacentini and F. Rinaldi, "Path loss prediction in urban environment using learning machines and dimensionality reduction techniques," *Comput. Manage. Sci.*, vol. 8, no. 4, pp. 371–385, Nov. 2011, doi: 10.1007/s10287-010-0121-8.

[15] S. P. Sotiroudis, S. K. Goudos, K. A. Gotsis, K. Siakavara, and J. N. Sahalos, "Modeling by optimal artificial neural networks the prediction of propagation path loss in urban environments," in *Proc. IEEE-APS Top. Conf. Antennas Propag. Wireless Commun. (APWC)*, Sep. 2013, pp. 599–602, doi: 10.1109/APWC.2013.6624896.

[16] S. P. Sotiroudis, K. Siakavara, and J. N. Sahalos, "A neural network approach to the prediction of the propagation path-loss for mobile communications systems in urban environments," *PIERS Online*, vol. 3, no. 8, pp. 1175–1179, 2007, doi: 10.2529/PIERS070220023434.

[17] P. Domingos, "A few useful things to know about machine learning," *Commun. ACM*, vol. 55, no. 10, pp. 78–87, Oct. 2012, doi: 10.1145/2347736.2347755.

[18] J. Thrane, B. Sliwa, C. Wietfeld, and H. Christiansen, "Deep learning-based signal strength prediction using geographical images and expert knowledge," Aug. 2020, *arXiv:2008.07747*. Accessed: Jan. 6, 2021. [Online]. Available: http://arxiv.org/abs/2008.07747

[19] K. Gadzicki, R. Khamsehashari, and C. Zetzsche, "Early vs late fusion in multimodal convolutional neural networks," in *Proc. IEEE 23rd Int. Conf. Inf. Fusion (FUSION)*, Jul. 2020, pp. 1–6, doi: 10.23919/FUSION45008.2020.9190246.

[20] D. H. Wolpert, "Stacked generalization," *Neural Netw.*, vol. 5, no. 2, pp. 241–259, Jan. 1992, doi: 10.1016/S0893-6080(05)80023-1.

[21] (2017). *OpenStreetMap Contributors S. Planet Dump*. [Online]. Available: https://planet.osm.org

[22] Z.-W. Xu, X.-M. Liu, and K. Zhang, "Mechanical properties prediction for hot rolled alloy steel using convolutional neural network," *IEEE Access*, vol. 7, pp. 47068–47078, 2019, doi: 10.1109/ACCESS.2019.2909586.

[23] G. Louppe, L. Wehenkel, A. Sutera, and P. Geurts, "Understanding variable importances in forests of randomized trees," in *Proc. 26th Int. Conf. Neural Inf. Process. Syst.*, Red Hook, NY, USA, vol. 1, Dec. 2013, pp. 431–439. Accessed: Sep. 30, 2020.

[24] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[25] *EDX Wireless Microcell/Indoor Module Reference Manual, Version 7*, EDX Wireless, Eugene, OR, USA, 2011.

[26] Z. Yun and M. F. Iskander, "Ray tracing for radio propagation modeling: Principles and applications," *IEEE Access*, vol. 3, pp. 1089–1100, 2015, doi: 10.1109/ACCESS.2015.2453991.

[27] S. P. Sotiroudis, S. K. Goudos, and K. Siakavara, "Feature importances: A tool to explain radio propagation and reduce model complexity," *Telecom*, vol. 1, no. 2, pp. 114–125, Aug. 2020, doi: 10.3390/telecom1020009.

[28] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, and D. Cournapeau, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, no. 85, pp. 2825–2830, 2011. Accessed: Jan. 6, 2021. [Online]. Available: http://jmlr.org/papers/v12/pedregosa11a.html

[29] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[30] S. Raschka, "MLxtend: Providing machine learning and data science utilities and extensions to Python's scientific computing stack," *J. Open Source Softw.*, vol. 3, no. 24, p. 638, Apr. 2018, doi: 10.21105/joss.00638.

[31] M. J. van der Laan, E. C. Polley, and A. E. Hubbard, "Super learner," *Stat. Appl. Genet. Mol. Biol.*, vol. 6, no. 1, Jan. 2007, Art. no. 25, doi: 10.2202/1544-6115.1309.

SOTIRIOS P. SOTIROUDIS was born in Thessaloniki, Greece, in 1976. He received the B.Sc. degree in physics, the M.Sc. degree in electronics, and the Ph.D. degree in physics from the Aristotle University of Thessaloniki, in 1999, 2002, and 2018, respectively, and the B.Sc. degree in informatics from the Hellenic Open University in 2011.

From 2004 to 2010, he worked at the Telecommunications Center of the Aristotle University of Thessaloniki. Since 2010, he works at the Greek Ministry of Education as a Teacher of physics and informatics. His research interests include wireless communications, radio propagation, and machine learning algorithms. He is currently a Postdoctoral Researcher with the RadioCommunications Laboratory, School of Physics, Aristotle University of Thessaloniki.

PANAGIOTIS SARIGIANNIDIS (Member, IEEE) received the B.Sc. and Ph.D. degrees in computer science from the Aristotle University of Thessaloniki, Thessaloniki, Greece, in 2001 and 2007, respectively. He has been an Associate Professor with the Department of Electrical and Computer Engineering, University of Western Macedonia, Kozani, Greece, since 2016. He has published over 180 papers in international journals, conferences, and book chapters, including the IEEE COMMUNICATIONS SURVEYS AND TUTORIALS, IEEE INTERNET OF THINGS, IEEE TRANSACTIONS ON BROADCASTING, IEEE SYSTEMS JOURNAL, *IEEE Wireless Communications Magazine, IEEE/OSA Journal of Lightwave Technology*, IEEE ACCESS, and *Computer Networks*. He has been involved in several national, European, and international projects. He is currently the project coordinator of three H2020 projects, namely a) H2020-DS-SC7-2017 (DS-07-2017), SPEAR: Secure and PrivatE smArt gRid, H2020-LC-SC3-EE-2020-1 (LC-SC3-EC-4-2020), b) EVIDENT: bEhaVioral Insgihts anD Effective eNergy policy acTions, and H2020-ICT-2020-1 (ICT-56-2020), and c) TERMINET: nexT gEneRation sMart INterconnectEd ioT, while he coordinates the Operational Program MARS: sMart fArming with dRoneS (Competitiveness, Entrepreneurship, and Innovation). He also serves as a principal investigator in the H2020-SU-DS-2018 (SU-DS04-2018-2020), SDN-microSENSE: SDN-microgrid reSilient Electrical eNergy SystEm and in the Erasmus+ KA2 ARRANGE-ICT: pArtneRship foR AddressiNG mEgatrends in ICT (Cooperation for Innovation and the Exchange of Good Practices). His research interests include telecommunication networks, the Internet of Things, and network security. He participates in the Editorial Boards of various journals, including *International Journal of Communication Systems* and *EURASIP Journal on Wireless Communications and Networking*.

SOTIRIOS K. GOUDOS (Senior Member, IEEE) received the B.Sc. degree in physics, the M.Sc. degree in electronics, and the Ph.D. degree in physics from the Aristotle University of Thessaloniki, in 1991, 1994, and 2001, respectively, the master's degree in information systems from the University of Macedonia, Greece, in 2005, and the Diploma degree in electrical and computer engineering from the Aristotle University of Thessaloniki, in 2011. He joined the Department of Physics, Aristotle University of Thessaloniki, in 2013, where he is currently an Associate Professor. His research interests include antenna and microwave structures design, evolutionary algorithms, wireless communications, and semantic web technologies. He is the Director of the ELEDIA@AUTH lab member of the ELEDIA Research Center Network.

Dr. Goudos is the founding Editor-in-Chief of the *Telecom* open access journal (MDPI publishing). He is currently serving as an Associate Editor for IEEE ACCESS and IEEE OPEN JOURNAL of the COMMUNICATION SOCIETY. He was honored as an IEEE Access Outstanding Associate Editor for 2019. He is currently serving as the Chapter/AG coordinator for IEEE Greece Section. He is also member of the Editorial Board of the *International Journal of Antennas and Propagation (IJAP)*, the *EURASIP Journal on Wireless Communications and Networking*, and the *International Journal on Advances on Intelligent Systems*. He is also member of the topic board of the *Electronics* open access journal. He was the Sub-Committee Chair of the Asian-Pacific Microwave Conference (APMC 2017) in the track of smart and reconfigurable antennas. He has also served as a member of the Technical Program Committees in several IEEE and non-IEEE conferences. He is a member of the IEICE, the Greek Physics Society, the Technical Chamber of Greece, and the Greek Computer Society. He currently serves as the Chapter/AG Coordinator for IEEE Section Greece. He has served as the Technical Program Chair of the International Conference on Modern Circuits and Systems Technologies (MOCAST). He was the Lead Guest Editor with the Special Issues of the *IJAP* with topic Evolutionary Algorithms Applied to Antennas and Propagation: Emerging Trends and Applications, in 2016 and 2017. He was also the Lead Guest Editor for the Special Issue of the *EURASIP Journal on Wireless Communications and Networking*, in 2018, with topic Optimization methods for Key Enabling Technologies: 5G, the IoT, and Big Data. He was the Editor of the book *Microwave Systems and Applications* (In Tech publishers, 2017). He is the author of the book *Emerging Evolutionary Algorithms for Antennas and Wireless Communications* (Institution of Engineering and Technology, 2021).

KATHERINE SIAKAVARA (Member, IEEE) received the B.Sc. degree in physics, the M.Sc. degree in electronics, and the Ph.D. degree in electrical engineering from the Aristotle University of Thessaloniki, Greece, in 1977, 1979, and 1982, respectively. She is currently an Associate Professor with the Department of Physics of the University of Thessaloniki, Greece. She is the author or coauthor of 42 international journal articles, of 43 papers in proceedings of international conferences, of two book chapters in English and two book chapters in Greek. Her research interests are in the areas of applied electromagnetism, analysis and design of antenna systems, microwaves, and radio-wave propagation and radio-communications.

• • •