194

SPATIAL DISTRIBUTION OF BUILDING USE

RECOGNITION AND PREDICTION OF USE WITH MACHINE LEARNING

TAITAWIP THIRAPONGPHAIBOON [1] ; SEAN HANNA [2]

ABSTRACT

Spatial measures that can be applied at different scales, such as angular choice at specific radii, have been shown to effectively identify distinct and important spatial patterns at each scale. Foreground movement routes, local commercial centres, and identifiable urban regions have each be associated with different scales of such measures. This paper investigates and quantifies the degree to which the likelihood of use types of individual buildings can be determined based on spatial measures of the street segment graph alone, using supervised machine learning on a detailed dataset of buildings in London.

A vector of 66 dimensions representing the spatial profile of each street segment is taken from analysis in DepthmapX and includes graph centrality measures such as choice at various radii, along with immediate features such as segment length. The proportion of building use is taken from OpenStreetMap and presented as a proportion of buildings in each category of Residential, Commercial or Business. A multilayer perceptron with between one and three hidden layers is trained to output the expected proportions of building use category given the various-dimensional spatial input vector for each segment. Various configurations of the training set and the network configuration are tested and discussed.

Results of training indicate a best accuracy of approximately 85% with correlation coefficient of 0.696 for training set and 0.517 (moderately strong) for test set with 37-dimensional spatial input vector, suggesting that the spatial predictability of building use is comparable to that of other human factors such as movement. The spatial factors that appear most significant to the distribution of use types are variations in metric total length and metric mean depth SLW at medium metric radii. An analysis of the most significant dimensions of the input is given, to suggest a spectral spatial profile of each building use type, and the effectiveness of these in prediction or planning is discussed.

KEYWORDS

Neural Networks, Land Use, Typology

1. Taitawip Thirapongphaiboon The Bartlett School of Architecture, UCL Gower St London WC1E zhengthwp@gmail.com
2.Sean Hanna The Bartlett School of Architecture, UCL Gower St London WC1E s.hanna@ucl.ac.uk

1. INTRODUCTION

Economics and social processes are known as main contributors in the creation of cities. ‘It would be strange indeed if they came into existence despite the pattern of life that go on them’ (Hillier, 2014). These attributes also lead to the spatial configuration or the physical space of the city. The being of a spatial configuration propels characteristics of its own area, which could mathematically be evaluated. For example, the value of choice or a spatial measure can account for the human movement as a route or a spatial attribute. The value of the Metric Mean Depth, or a spatial measure on a street can explain the shape and size of a building block or a spatial attribute. This mathematic evaluation can be called a spatial measure. However, only a spatial measure cannot adequately account for spatial pattern It must follow a more complex relationship than only one spatial measure. For instance, with only the Metric Mean Depth or only choice, it cannot indicate living centres or land-use patterns (a spatial pattern). Through only a spatial measure, it is unlikely to represent for a spatial pattern. With a certain set of spatial measures, one can possibly predict and explain a spatial pattern. This set of spatial measures can be referred to as a "syntactic signature", (Silva, 2017).

By using a syntactic signature, a complicated relationship can emerge. Thus, it might be hard to explicitly explain such relation. However, in a recent decade, the Machine Learning technique has been introduced to handle such problem. This technique learns from a large set of corrected data. One of the main difficulty of this method is known to be an insufficient data. Presently, since data resources have been growing and becoming accessible such as Open Street Map, the study of spatial pattern are now plausible.

This research addresses the following question: what is a certain degree to which the likelihood of use types of individual buildings can be determined based on spatial measures of the street segment graph alone, using supervised machine learning on a detailed dataset of buildings in London?

The paper will be presented in four parts. Firstly, establishment and development of a detailed review of studies handling with a relationship between spatial measures and spatial pattern, an evaluation of the feasibility of this project using Machine Learning techniques, and a study of preceding works dealing with Machine Learning to predict the spatial pattern will be put forwarded. Secondly, the methodology section will be illustrated. This part will explain how to deal with dataset and the creation of machine learning model. In the third part, the results of the model will be shown and discussed. The net result from the methodology will be explicated. Then, several experiments will be processed to examine in detail of a spatial feature. The significance of each feature will be introduced. Conclusions in responding to the research question will be drawn and sets of the findings of which spatial features might be correlated to particular uses will be shown in the final part.

2. RELATIONSHIP BETWEEN SPATIAL MEASURES AND LANDUSE, SCALE, AND MACHINE LEARNING TECNIQUES IN SPACE SYNTAX

Space Syntax has been effectively shown to identify the correlation between spatial configuration and various phenomena such as social, economic, and environmental phenomena. Based on Hiller and Hanson (1984), these phenomena include patterns of movement, land use density, land use mix, urban growth, safety and crime, and so on. By using the space syntax approach, many researchers have shown how movement pattern and flows in cities significantly formed by the spatial of street networks. This event forms the progression and evolution of the centres and sub-centres affecting the way of life in a city. It could affect the level of security, social segregation, and organisational cultures. One of the main theory called the natural movement (B Hillier, 1993) suggesting that the inequality of street networks produces different levels of accessibility affecting business on the selection of location. A certain business may pick a characteristic of a location that suitable itself the most. For instance, a retail owner may choose the street segment having a high human movement in order to gain customers as much as possible. According to such a hypothesis, a spatial pattern, including land use, in the urban street network can be predicted by spatial configuration to a certain degree.

While the relationship between spatial configuration and street network pattern becomes an accepted practice, it is crucial to review in which spatial measures associate which particular street network pattern, in this case, the pattern of the land use. To understand spatial measurements, two subjects must be introduced. The first subject is the type of spatial measures allowing to gain a different aspect of implication. The second subject is scale allowing to understand the level of hierarchy we are dealing.

For spatial measurements, two types of the measure have been mentioned the most related to the aspect of correlation between land use and spatial configuration. The first type of measures is Integration, which measures how close each segment is to all other under different types of distance and at a different scale, known as "to-movement" (Bill Hillier S .I., 2005). This measure implies how likely pedestrians would turn out to arrive in the space more than other spaces without making an attempt. The second type of measure is Choice, which measures how much movement is likely to pass through each segment on trips between all other segments (Bill Hillier S .I., 2005). This measure implies how people likely to choose the segment as a way to pass through on trips. According to such implications, both spatial attributes can identify the type of land use that would fit best in a certain space (Nadia Charalambous, 2012)

Integration can be used to study, within an urban system, human movement based on a substantial number of studies over the past two decades. In 1997, Desyllas presented a street pattern correlated with integration. He mentioned that the street pattern forms the movement, which attracts the shops rather than shops generating movement. In 2000, Bin Jiangl explained that, by using space syntax techniques, urban planners can envision movement flows before the actual construction within urban systems. Such techniques could be applied to many aspects such as business, interaction, and security. For example, if urban planners have to design a location of residential buildings, one of the main concerned aspect would be security. In recent theory, Shu (1999) has presented that people are better protected in more integrated areas with more visual field and flow potentials. Accordingly, integration can be used as a part for the prevention of crime in the residential area. Thus, by using integration as a component, both research suggests that it is possible to predict spatial pattern through street segment analysis.

Choice can also be used to study human movement. However, choice measures between rather than closeness. It could be implied of a street segment that has a potential as a route (Nadia Charalambous, 2012). It also indicates how often a segment has been chosen as part of the shortest path between every pair of a segment on the specific metric radius. Compare to integration, the choice seems to have more research about movement such as activities like walking and driving (Haq, 2017) rather than land use. This spatial feature can determine the type that would suit best in such street, possibly certain land uses would expect areas with a high integration rate (Nadia Charalambous, 2012).

Both spatial measures have been mentioned as the main ways to reveal the human movement pattern. However, there are more than two features that could be used in space syntax as spatial measures such as Metric Mean Depth and Metric Total Length. These spatial measures could be applied as components for analysis in order to gain a better correlation between land use and spatial configuration.

With scale, the analysis can be a focus on a certain selected area from the whole system. Conventionally, terms of radius have been differentiated as "local" and "global". However, these terms of the radius can be seen as problematic for two reasons (Krenz, 2017). First, the boundary selection has a strong influence on the observed structure. Second, the radius n is not a distance free measure. These problems have been rooted in the area of space syntax that lack of theorising in scale.

One of the theories that could be challenged such method is central place theory of the economic distribution of urban space (Christaller,1993). The idea of this theory is that cities are the centre of economic exchange generating hierarchical order. The system has been divided into seven hierarchical level of urban form such as small town to large cities. Based on this theory, Krenz (2017) believe that Walter Chirstaller's Central Place Theory (CPT) (1993) of hierarchical can bring a significant insight into the emergence of scales because a regional distribution of urban areas must follow a more complex relationship than just "local" and "global". In his research, he tries to discover latent scale structures among a huge set of different metric radii. As a result, he gained insight into the intrinsic scales embedded in the graph and merely four of scales in central place theory appears. He also suggests that the remaining centres might not be adequately displayed to form its own spatial scales and worth considering intersections between these four spatial scales as emergent spatial scales. According to Region Morphology (Krenz, 2017), we could define eight levels of hierarchical scale including neighbourhood, city, between city and metropolitan, metropolitan, between metropolitan and inter-regional, inter-regional, between inter-regional and intra-regional, and intra-regional. Presumably, such spatial scales have been rooted in the human-shaped spatial networks as the product of repetitive human activities and informed by economic processes. This interesting finding indicates that there are levels of hierarchy inside the complex system of the city. Spatial pattern hence can be distinct in each scale and could be unveiled by explored in a certain spatial scale.

In the last decade, one of the techniques for prediction is machine learning. Machine learning is the systems automatically learn program from a large set of data, which has been spread widely throughout computer science and other fields. It could identify a subtle pattern in the data, which could hence predict and recognize such a pattern.

The Machine Learning algorithms are generally classified into three categories, which are supervised learning, unsupervised learning, and reinforcement learning. These algorithms are comparable to data mining, which explores from end to end data to assume for patterns (M Praveena, 2017). However, these algorithms are used for a particular purpose. For example, while the purpose of supervised learning is to estimate the best mapping function that when you have a new input data that the algorithm can predict the outcomes, the purpose of unsupervised learning is to model the highlighted structure in the data for studying more about the data system (Brownlee, 2018). In this research, the prediction of the spatial distribution of building use based on spatial measures will be the main focus. The algorithm which will be used as a method and discussion is hence supervised learning.

The most common form of machine learning scheme used in solving the engineer problems is supervised learning (Xu, 2013). The system learns to suggest a program from a large set of labelled training data. In other words, the algorithm is learned from what we advise as a supervisor. The training set will contain a set of input features and corrected outcomes. In image recognition, for instance, we can use each pixel in the picture of a human face as input features and label a human face as an outcome (Abdelfattah, 2017). Based on this method, a set of spatial measures can be used as input features and the proportion of building used as outcomes. With a certain number of data, the model should predict and recognise the pattern of any spatial street network in the research regional area.

Based on my knowledge, application of machine learning between a set of spatial measures and real-world data such as spatial distribution of building use has not been widely studied. Most of the research has been studies on certain spatial measures rather than as a set; they also study other real word data such as spatial distribution urban pollution (Ben Croxford, 1996) and spatial distribution of nature change (Jonathan Cheung-Wai Chan, 2001) rather than spatial distribution of building use. This occurrence might be caused by the lack of building use data in street networks in the past. However, recently, the supply of data grows rich. Many open sources data become available for common people such as Google API and Open Street Map. The feasibility of prediction between the spatial distribution of building use and spatial measures hence becomes possible.
One of the recent related work worth to mention is the syntactic signature of Starbuck' location towards a machine-learning approach to location decision making (Silva, 2017). Silva used the machine learning methods to distinguish the pattern of ten business types by using spatial measures, or syntactic properties, of such street segment location in London, UK. In this research, the algorithm will predict

whether a certain business is likely to located on the certain street or not. He hypothesizes that if the natural movement is true, and building use caused by syntactic properties, then it should be conceivable to determine in which degree of each spatial measures affected a particular building use located on such street. According to his research, the model has a prediction performance above 70% on the testing set, which could indicate that there is a relationship between land use and the syntactic properties of spatial configuration. Thus, he suggested that it seems possible to find a syntactic signature for certain business type.

Even though there is not much research dealing with a large set of the spatial distribution of building use and a set of spatial measures, perhaps because of the lack of data in the past, the data becomes rich and possible for this research today. By using machine learning techniques, it seems feasible to use a set of spatial measures as input and the proportion of building use as output in supervised learning with a purpose to predict the spatial pattern

3. METHODS AND DATASETS

The methodology has been structured into two main phrases (Figure 1). The first phrase is pre-processing. This phrase aims to collect spatial measures as input feature and spatial distribution of building use as output features. The second phrase is supervised learning by using the data from the first step as an input dataset. If the results gather data that has spatial pattern enough, the model of supervised learning then could be trained and predict the building proportion by using merely spatial measures.

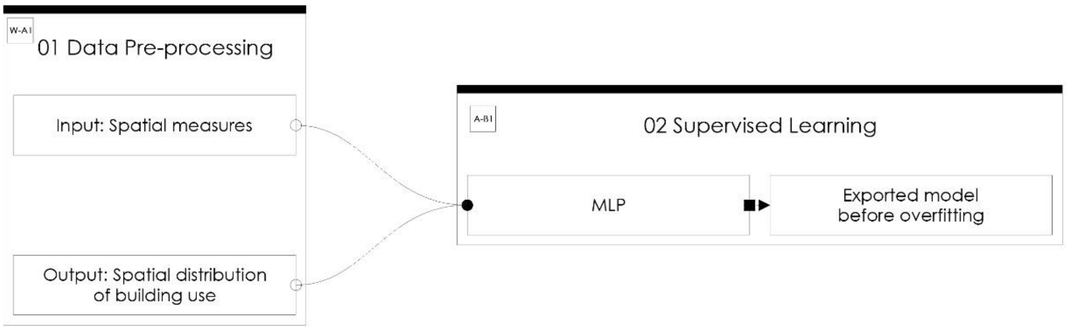

Figure 1– Main workflow

For the purpose of prediction and recognition of spatial distribution of building use by using spatial measure, building use proportion on segments and spatial measures will be used to train in supervised learning. However, there is no complete data of building use proportion on segments in any source based on my knowledge. Thus, the way to collect building use data and spatial measures need to be done before the machine learning steps. In this step, it is called data pre-processing (phrase 1).

The purpose, in this phrase, is to transform the data from geometry into a format, which can be easily interpreted by machine learning. In order to clarify, this phrase aims to project the building data (building use and so on) into street segments and export both building use proportion and spatial measures as a set of value. There are three stages and fourteen steps in pre-processing data workflow (Figure 1, 01-Data pre-processing). Three stages consist of street segments data collection, buildings data collection, and spatial measures as input & building use proportion as output as shown in Figure 3. For the details in each stages and steps of technical coding knowledge, I would suggest to in my dissertation called the spatial distribution of building uses (Thirapongphaiboon, 2018) in the Bartlett library since this paper will be discussed more about the experiment of machine learning and datasets.

Figure 2 – Example of Applying projection buildings uses proportion on street segments algorithm in Camden.

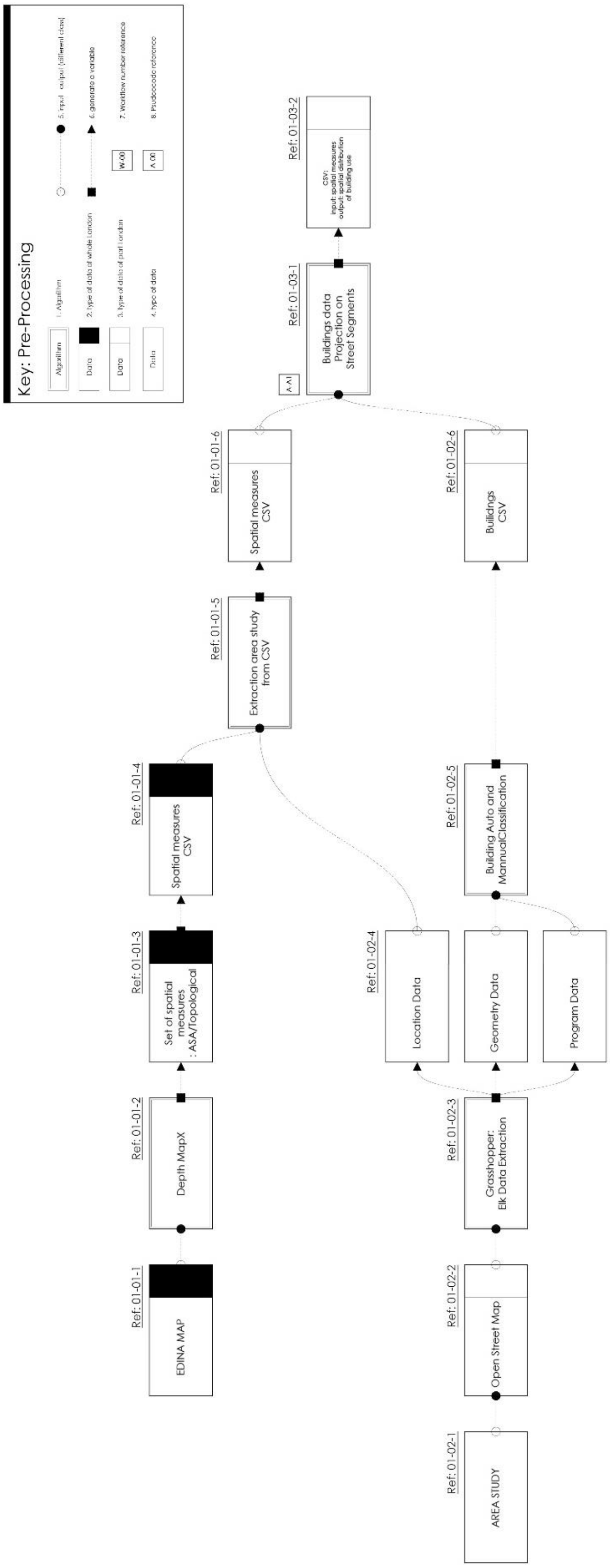

Figure 3 – Pre-processing workflow

8

Afterwards, the data will be used to train by using one of the Machine Learning techniques called Multi-Layer Neural Network (MLP). This network aims to find the model that gain the best correlation coefficient with high prediction performance. In this step, it is called supervised learning (phrase 2).

The only main algorithm in this phrase is Multi-Layer Perceptron MLP for supervised learning (Figure 1,02-Supervised learning). The artificial neural network will be created to train and learn from pre-processing data. According to the theory, after training, the model should be able to predict building proportion by using spatial measures only.

The process can differentiate into two parts in the workflow (Figure 4, workflow). The first part is set up, which will process only one time. The second part will process continuously as a loop. In the first part, the data will be read from the pre-processing data file. The data will be loaded into the algorithm. Then the data will be separated into two datasets, which are training set (80% of the original data) and test set (20 % of the original data). The purpose of file separation is to measure the correlation coefficient in both sets to indicate the over-fitting event. Lastly, in this part, the multi-layer neural network will be initiated by the number of layer and number of neural in each layer. In this case, there are three layers in the network and fifty neurons in each layer.

In the second part, the algorithm will run as a loop. The primary function of this part is training. The network in will be trained by using the data from the training set and guess outcomes, find an error, and adjust the weight accordingly. This algorithm uses both feedforward and backpropagation techniques. Afterwards, the network will evaluate the correlation by using both the training set and test set correlation.

In most cases, before the over-fitting events occur in the test set will reach around the point of lowest error, which conforms with correlation most of the time. However, in this case, we want the result that could have the best prediction performance. The model, thus, uses the best correlation as a threshold to export the network model when it is reaching the highest correlation, which locates around the error landscape before an over-fitting event occurs.

Beyond the typical method of differentiating data into two datasets for training and testing, there is one more issue concerned in the data, the noise in the data. In most streets there is not much chance that the building uses proportion can be two hundred percent (since it has buildings in two sides of a street) because a street can contain other elements rather than buildings such as a gap between building. This kind of noises have been concerned that it would reduce the quality of machine learning prediction and recognition. For the sake of such hypothesis, the normalized datasets have been introduced.

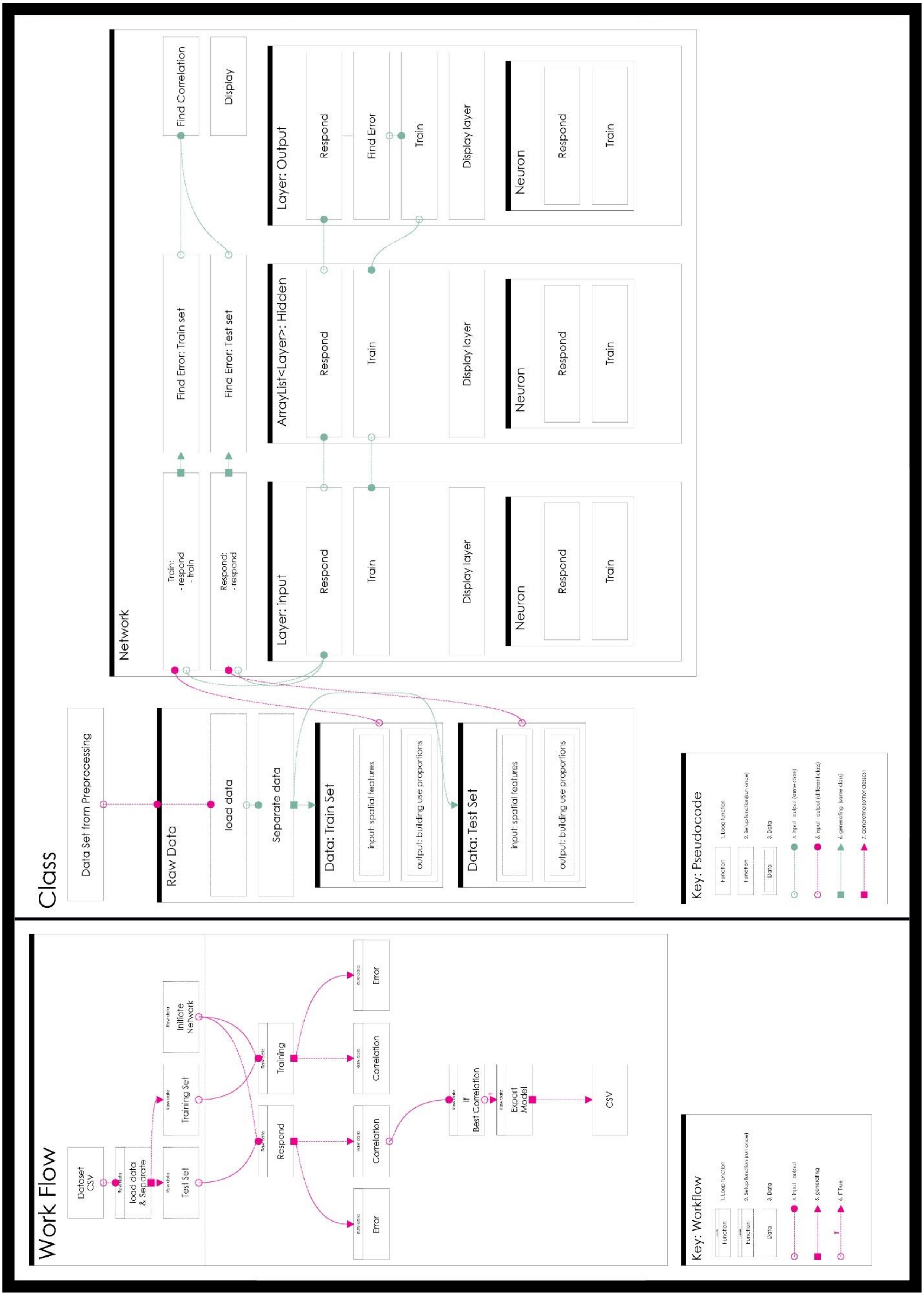

Figure 4 – Multi-Layer Perceptron algorithm (MLP)

To normalize the building use proportion in each street segment, all of the building uses distance will be summed up. Then, such value will be used to get new of each building use percentage. In this case, when the new normalized building use proportions have been summed up, it will be two hundred percent as shown in the formula below. In addition of this technique, some of street segments will not be included in the datasets such as a street that has building use proportion less than hundred percent since most of the time it indicate a lack of information.

$$np_i = d_i \div \sum_{i=0}^{n} d_i$$

$d_i = distance\ of\ building\ use\ type\ i$

$np_i = normalized\ percentage\ of\ building\ use\ type\ i$

For such reasons, in each experiment, there will be four datasets of normalized and unnormalized data with testing and training set in the next chapter.

4. RESULTS AND DISCUSSION

Following the methodology, the multilayer perceptron has been training and select the network that gains the best fit before the over-fitting events in the network occurs. In the algorithm, the error has been used to train the neural network; then, it exports the model gaining the highest correlation in the test set. Two sets of data have been introduced to test in the network, which is unnormalized building use proportion set and normalised building use proportion set. If the building use proportions have been normalised, it should decrease noise data in street networks such as a gap between buildings and presumably increase the value of correlation. The following explanation will show the detail in each data set and compare to all set.

In training set with unnormalized building use proportion as output, the model has a prediction performance of above 88% at 0.613 correlation coefficient. Typically, the training set correlation coefficient can be more than 0.8, a strong correlation. If the correlation in the test set is low, however, this result could indicate over-fitting. For such reason, the algorithm in this research chooses the model that has the highest correlation in the test set, which indicates the point before over-fitting. According to these results, the correlation in the training set with unnormalized building use proportion indicates a high positive correlation based on the Rule of Thumb for Interpreting the Size of Correlation Coefficient.

In training set with normalised building use proportion as output, the result of training indicates an accuracy of approximately 88% at 0.692 correlation coefficient. While the training set with

11

unnormalized building use proportion gains correlation coefficient of 0.613, this set gains 0.692. Based on these result, it indicates a high positive correlation.

In the test set with unnormalized building use proportion as output, the performance of the prediction model is above 87% at 0.489 correlation coefficient. As we mentioned before, in standard practice, the test set correlation coefficient will be lower. According to these results, it indicates a moderate positive correlation.

In the test set with normalised building use proportion as output, the model gains an accuracy of estimate above 83% at 0.501 correlation coefficient. Compare to the set that has not been normalised the data, this set has a higher correlation coefficient value. Based on these result, this set has a moderate positive correlation

From these values, we could investigate more in the detail of each use in every set. As you can see (Figure 4.1), both commercial and residential use gain a higher correlation than business use in most set. For example, in training set with normalized data, the correlation coefficient between actual data and responded data in commercial and residential is approximately 0.7, while the correlation coefficient for business use is 0.61. This event happens in most cases.

Regarding the issue of normalised the building use proportion, normalised the building use proportion does effect on the correlation coefficient. While both training and test set that has not been normalised gain the correlation coefficient of 0.613 and 0.489, both training and test set that has been normalised gain the correlation coefficient of 0.692 and 0.501. According to these value, the normalized proportion can be indicated that it has an effect.

From two sets of data, the set gaining highest correlation coefficient is the set with normalised building use proportion as output. The performance of the prediction model is approximately 87% at 0.692 correlation coefficient in the training set and above 83% at 0.501 correlation coefficient in the test set. In the training set, the correlation coefficient in each use are 0.742 for commercial buildings proportion, 0.616 for business buildings proportion and 0.717 for residential buildings proportions. In the test set, the correlation coefficient in each use are 0.555 for commercial buildings proportion, 0.408 for business buildings proportion and 0.539 for residential buildings proportions. For the test set, these results indicate a moderate positive correlation according to the Rule of Thumb for Interpreting the Size of Correlation Coefficient.

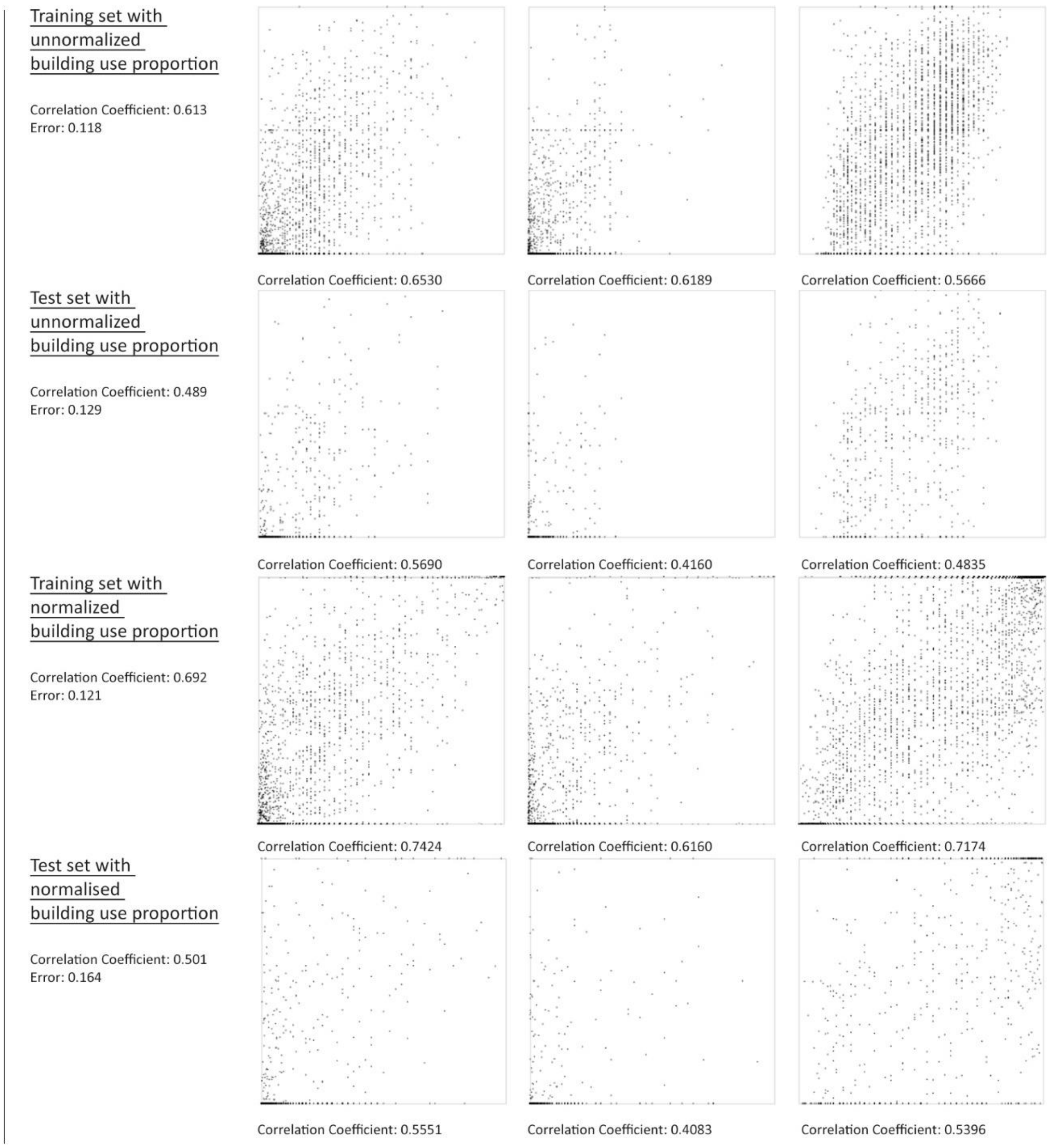

Figure 5 – The Correlation Coefficient between actual data (x value) and responded data (y value) in each use (Commercial, Business, and Residential). The diagram portrays the correlation of 4 sets, which are training set with unnormalized building use proportion, training set with normalised building use proportion, test set with unnormalized building use proportion, and test set with normalised building use proportion.

13

Since the research question is whether spatial distribution of building use is relevant to spatial measures; and, which set of these syntactic features affect which building use proportion, the investigation of the relationship between spatial features and certain use is conducted. In this part, we will begin with the correlation coefficient between every spatial feature and building proportion in each use. The first experiment will be tested with both sets of actual data and the training data. Then, in the second experiment, training with a different set of features will be tested to find a set of spatial features that gain a strong correlation. The following part will be discussed in the first and second experiment with implications, visualise and present the responded data on London map.

In the statistics field, there is a measure of how strong a relationship between two variables called Correlation coefficients. In the first part of the discussion, an investigation of each spatial features will be conducted to pinpoint a particular set of features gaining strong positive or negative correlation coefficient. The experiment will measure the correlation between building use proportion in each category and all spatial features. Four sets of data will be showing in this investigation, consisting of unnormalized actual building use proportion data set, normalised actual building use proportion data set, unnormalized responded building use proportion data set, and normalised responded building use proportion data set.

Most of the detail show strikingly similar results (Figure 6 and Figure 7). In these four data sets have shown the same best positive and negative correlation coefficient for each use. For commercial use, the best positive correlation feature is Metric Total Length at 600 m, and the best negative correlation features are Metric Mean Depth SLW at 1800 m. For business use, the best positive correlation feature is Node Count at 300 m, and the best negative correlation feature is Metric Mean Depth at 7200 m. For residential use, the best positive correlation feature is Metric Mean Depth at 7200m, and the best negative correlation feature is Metric Total Length at 7200m.

Based on the literature review, Angular Segment Analysis such as integration and choice are expected to be an excel feature correlated with the spatial pattern, the pattern of land use. However, according to the results, the topological analysis seems to play a significant role. In residential and commercial, for example, the best correlation is Metric Mean Depth and Metric Total Length. These findings hence suggest that topological analysis could be a crucial element for prediction distribution of building use.

One of noticeable relationship from the results is the reverse effect of residential and commercial. While the positive correlation is Metric Total Length and negative correlation is Metric Mean Depth SLW in commercial, in residential, the positive correlation is Metric Mean Depth, and the negative correlation is Metric Total Length. Possibly, this reverse relationship could be compared to intuitive observation. For example, the living centre tends to locate in more mean depth to avoid noise, and commercial tend to locate in lower mean depth to gain better accessibility. However, it is still not a certain fact from this result.

14

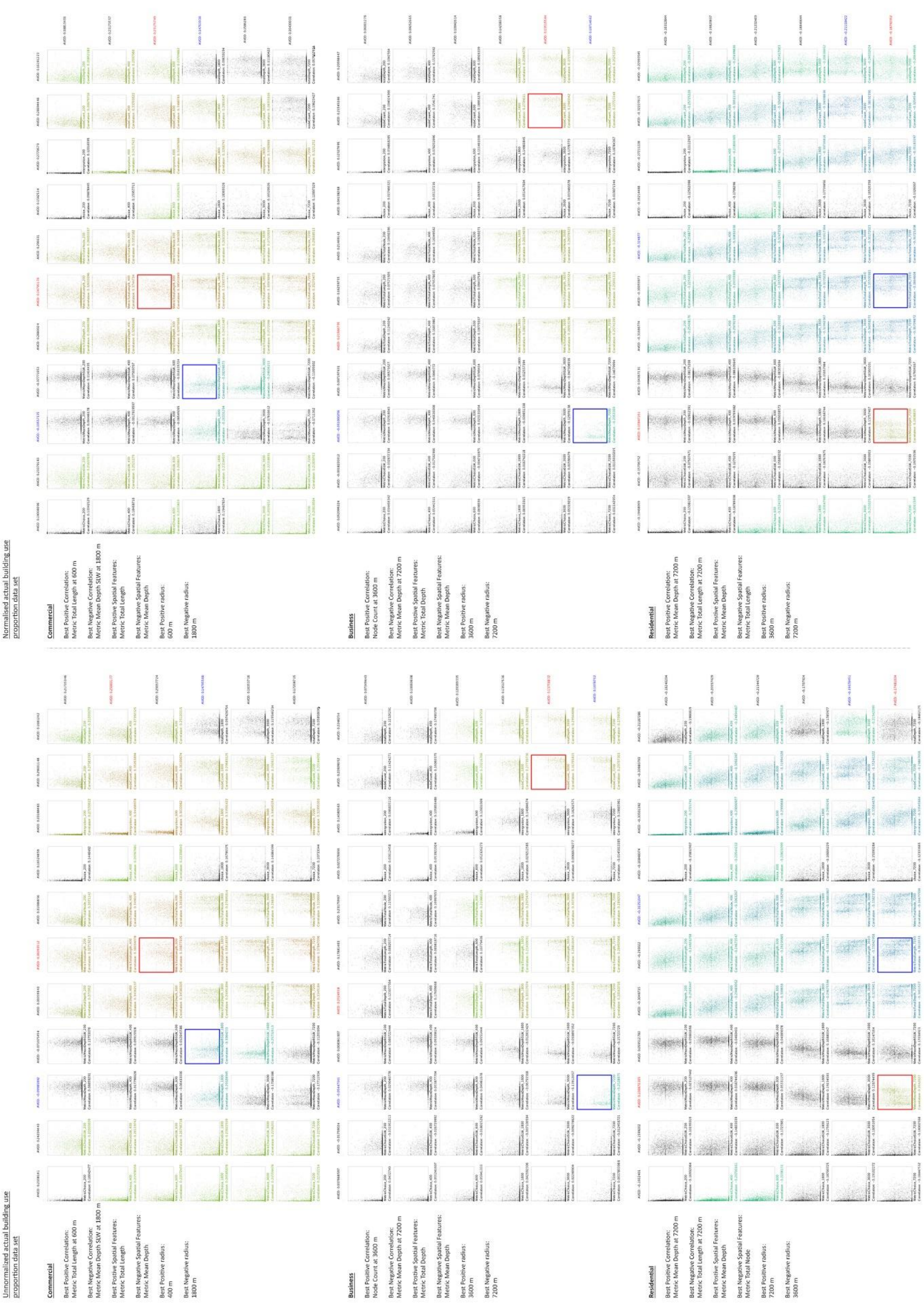

Figure 6 – Correlation between building use proportion (x value) in each category and all spatial features (y value) with unnormalized actual building use proportion data set (left) and normalised actual building use proportion data set (right).

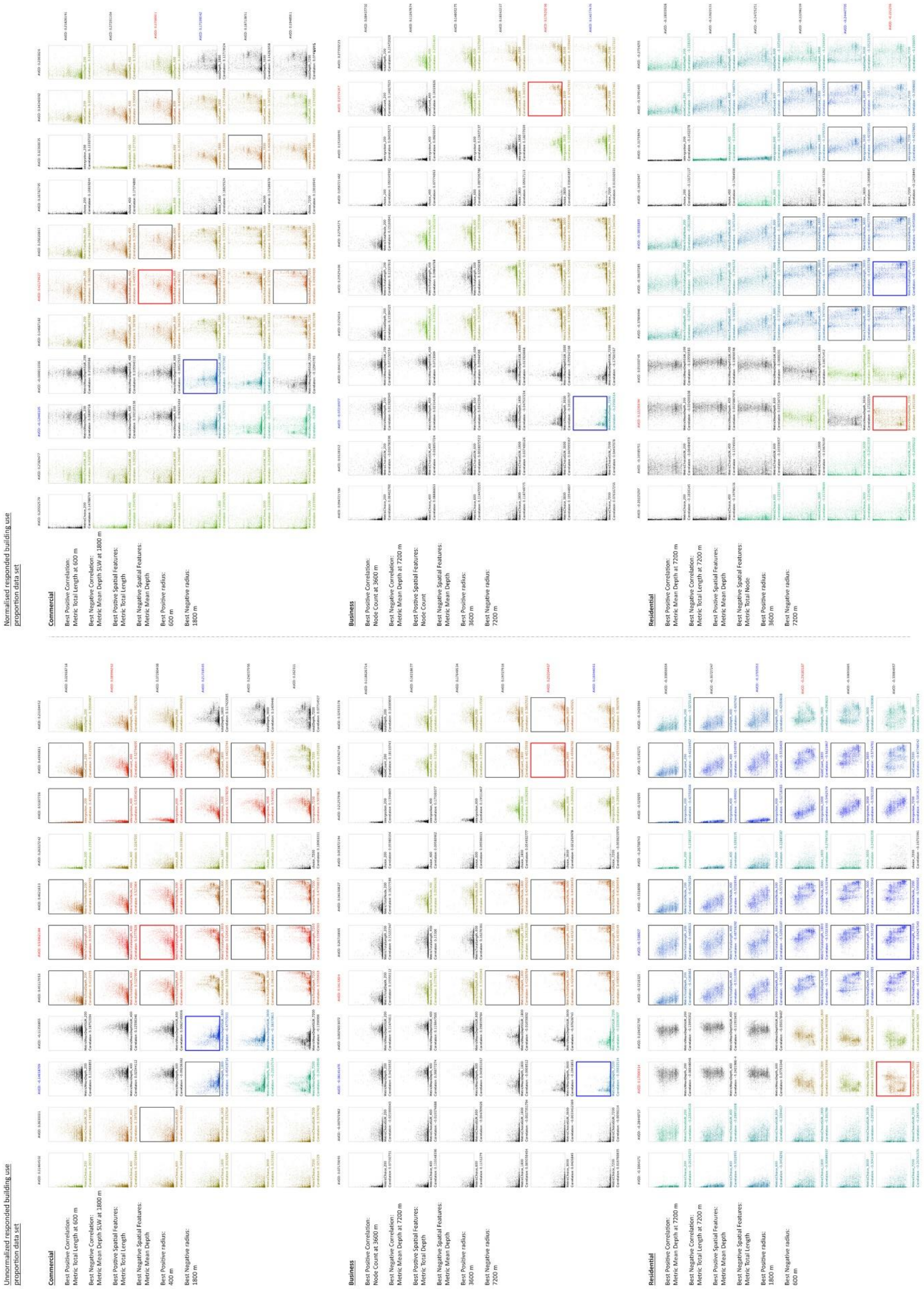

Figure 7 – Correlation between building use proportion (x value) in each category and all spatial features (y value) with unnormalized responded building use proportion data set (left) and normalised responded building use proportion data set (right).

16

In Metric And Topo-Geometric Properties Of Urban Street Networks (Bill Hillier A .T-.P., 2012) indicated that Metric Mean Depth does reflect block size and shape, but this feature alone could not account for living centres. Metric Mean Depth hence could be one of the element to find living centres. From the result, Metric Mean Depth plays a significant role in all use. While the best negative correlation in commercial and business is Metric Mean Depth and Metric Mean Depth SLW, the best positive correlation in residential is Metric Mean Depth. These results support that Metric Mean Depth perhaps should be used as a feature element to indicate land use.

Even though most of the detail shown similar results, some details are different in each set. In these four sets, there are differences in the best correlation in the type of spatial measures and radii. The differences in the type of spatial measures occur in business use and residential use. For business use, in most sets, the best positive correlation type of spatial measure is Metric Total Depth and the best negative correlation measurement is Metric Mean Depth. However, in normalised responded building use proportion data, the best positive correlation type of spatial measure is Node Count. For residential use, in most sets, the best positive correlation type of spatial measure is Metric Mean Depth, and the best negative correlation type of spatial measure is Metric Total Node. However, in unnormalized responded building use proportion data set, the best negative correlation is Metric Total Length. For commercial, on the other hand, the best correlations are similar, which are Metric Total length for the positive type of spatial measure, and Metric Mean Depth for the negative type of spatial measure.

By the complexity of the system in an urban street network, a single or several spatial measures might not be enough to approximate the spatial pattern. According to the result of uncertainty in each set, it indicates that such a hypothesis is presumably correct. Even though there is a similarity in each set, it still varies in value. Due to such fact, a set of best type spatial features in different data set can be seen as a useful set of features to predict the proportion of land use, but not as significant as the best positive and negative spatial features.

The differences in the radii appear in commercial use and residential use. For commercial use, while the strongest negative radius in all set is 1800m, the strongest negative radius is different between the actual building use proportion sets and the responded building use proportion sets. In the actual building use proportion sets, the best negative radius is 400 m. In the responded building use proportion sets, the best negative radius is 600 m. For residential use, in most sets, the best positive radius is 7200 m, and the best negative radius is 3600m. However, in unnormalized responded building use proportion data set, the best positive radius is 1800 m and the best negative radius is 600 m. For business, on the other hand, have the same radius correlation with a positive correlation radius at 3600 m and a negative correlation radius at 7200 m.

In the sense of community hierarchy, commercial and resident is expected to correlate with small to medium radii since living centre would have a relationship between community rather than city or metropolitan. In the same way, the correlated commercial radii correspond with small to medium radii from 400 to 1,800 meters. According to Region Morphology (Krenz, 2017),these radii can be referred to as the scale of a neighbourhood to a city. However, the correlated residential radii seem to correspond with medium to high radii from 600 to 7,200 meters. Based on Region Morphology (Krenz, 2017), these radii can be referred as a scale of betweenness of neighbourhood and city to betweenness of city and metropolitan. Perhaps, in a commercial use, the network could indicate a spatial pattern in a neighbourhood such as a bar in the corner and local shop as a neighbourhood centre, while the living centre is corresponding with a larger radius rather than neighbourhood radius.

With business use, however, it is expected to correlate with high radii since the business location is likely to locate in the centre of a city. Accordingly, the investigation supports such a hypothesis. The business correlated radii have corresponded to high radii from 3,600 – 7,200 meters. These radii could be called as Betweenness city and Metropolitan in Region Morphology (Krenz, 2017), which I would refer these radii as Central Business District (CBD) radii.

To investigate more in detail, the type of the spatial measures must be understood. There are two main types of spatial measure in these data. The first set is the Topological Analysis including Metric Choice, Metric Choice SLW, Metric Mean Depth, Metric Mean Depth SLW, Metric Total Length, and Metric Total Node. The second set is Angular Segment analysis (ASA) including Choice, Integration, Node Count, and Total Depth. These two analyses allow measuring different definitions of distance in the urban street network. While topological analysis refers to every change of direction or turn between a street segment and its neighbouring street segments, the angular analysis assigns a value of the degree of angular change of the direction between a street segment and its neighbour)Laura Narvaez, 2015(. With these measurements, we could differentiate a set of strong correlation (bold border in Figure 6 and Figure 7) into two groups.

In Topological Analysis, the results have shown both positive and negative correlation. For commercial use, a set of moderately strong positive correlation consists of Metric Total Depth, Metric Total Length, and Metric Total Node. A set of moderately strong negative correlation consists of Metric Mean Depth and Metric Mean Depth SLW. Similar to commercial use, business use has the same direction with commercial use with lower correlation. In contrast to commercial use, the residential use has a reverse relationship with the commercial.

In Angular Segment Analysis (ASA), the result has shown either positive or negative correlation. For commercial use, a set of moderately strong positive correlation consists of Integration and Node Count. For business use, a set of moderately strong positive correlation consists of Node Count and Total Depth in higher radii from 1800 m to 7200 m. Contrary to commercial use, the residential use has a reverse relationship with commercial correlation with a set of Total Depth in lower radii.

Spatial measures gaining best correlation	Commercial		Business		Residential	
	Positive	Negative	Positive	Negative	Positive	Negative
All Data Set	Metric Total Length At 600 m	Metric Mean Depth SLW At 1800 m	Node Count At 3600m	Metric Mean Depth At 7200 m	Metric Mean Depth At 7200 m	Metric Total Length At 7200 m

Table 1 – The table of spatial measures gaining best correlation. The results in all set are similar.

Spatial measure gaining strong correlation compared to other features in both normalised and unnormalized training set		Commercial		Business		Residential	
		Positive	Negative	Positive	Negative	Positive	Negative
Set of Spatial Measures	**Topological Analysis**	Metric Total Length, Metric Total Node	Metric Mean Depth SLW	-	Metric Mean Depth	Metric Mean Depth	Metric Total Depth, Metric Total Length, Metric Total Node
	Angular Analysis	Integration, Node Count	-	Node Count	-	-	Integration, Node Count
Length of radii		400 – 600 meters	1800 – 3600 meters	1800-7200 meters	7200 meters	1800 – 7200 meters	400 – 7200 meters

Table 2 - The table of spatial measure set gaining strong correlation compared to other features in both normalised and unnormalized training set.

From the first experiment, the results have shown that a single or several set of features might not be adequate to predict spatial pattern. Also, Topological Analysis seems to play a significant role to predict spatial distribution of building use and worth to be investigated more. For example, Metric Mean Depth can be use as one of an element feature to predict living centres. Lastly, the radii in each use is varies, yet acceptable. While the commercial radii tend to correlate with small-medium radius with could be referred as neighbourhood to city scale, the residential radii tend to correlate with medium-high radius with could be referred as betweeness neighbourhood and city to betweeness city and metropolitan. The business radii, however, is distinct in high radius with could be referred as betweeness city and metropolitan or could be referred as Central Business District radius.

19

Still this discussion is based on the best correlation in each scale, it is important to recognise that these experiments have been structure in a certain set of scale and spatial measures and can be improved if the experiment is tested in a full range of scale and spatial measures.

In the second experiment, a different set of features will be tested to find a set of spatial features that gain a strong correlation. In statistics and machine learning theory, this technique is known as dimensionality reduction (Rowies ST, 2000). To be more specific, this experiment will conduct Feature selection with the embedded strategy by selecting to add or remove features while building the model based on the prediction errors. Hypothetically, if the number of features is reduced, the precision performance should be improved with a set of strong correlation features.

In this experiment, four sets of grouping features will be training and export the correlation efficient value before over-fitting event appears. The first set is the set of all dimensions we have been shown in the result. In this set, all spatial features have been used to train the model. In the second set, the feature will be based on the first experiment. This set of features included best correlation features in all use, strong correlation features in training set, excluded weak correlation features in all use. The third set is a set of spatial features with different radii such as integration and choice in all radii we have. Finally, the last set based on the literature review. The main features in this set are integration and choice with a set of radii.

Correlation based on selection set of features		**Number of features**	**Best Correlation Coefficient**	**Error**
Set 01	**All features (normalised)**	66	0.501	0.164
Set 02	**Best correlation in all use**	5	0.496	0.174
	Strong correlation in training set (unnormalized)	37	0.517	0.166
	Strong correlation in training set (normalised)	18	0.507	0.160
	Excluded weak correlation (unnormalized)	59	0.495	0.160
	Excluded weak correlation (normalised)	53	0.502	0.165
Set 03	**Metric Choice**	6	0.240	0.190
	Metric Choice SLW	6	0.219	0.200
	Metric Mean Depth	6	0.371	0.180
	Metric Mean Depth SLW	6	0.378	0.192
	Metric Total Depth	6	0.436	0.188
	Metric Total Length	6	0.456	0.172
	Metric Total Node	6	0.481	0.177
	Choice	6	0.259	0.200
	Integration	6	0.374	0.193
	Node Count	6	0.476	0.173
	Total Depth	6	0.368	0.194
Set 04	**Integration and Choice**	12	0.393	0.184

Table 3 - The table shown feature selection set and the best correlation value of each set.

Since the set of all dimensions, the first set of features, has already been tested, we will begin with the result of the second set of features (Table 3, set 02). This set used sets of features based on the first experiment (See 6.1 FEATURE SELECTION SET 02 for features elements). These sets included a group of best correlation features in all use (5 dimensions), a group of strong correlation features in training set (37 and 18 dimensions), a group of excluded weak correlation features in the normalised set (53 dimensions). Among each group in this set, three groups gain a better correlation than using all dimension, 66 dimensions. In this set, a group of features that gain the strongest correlation coefficient is a set of strong correlation features in the unnormalized training set, which is 0.5168 at 0.16 error with 37 dimensions.

Previously, the hypothesis of reduction dimension has been suggested that if we use such a technique, the correlation coefficient might be improved. The results of the second set, accordingly, supported that the correlation coefficient can be enhanced. From these group of features, three groups correlation coefficient exceed the original correlation coefficient; in one of these group, it has only 18 dimensions. These results indicate that the correlation coefficient can be improved, and more data probably need in order to gain a better result.

In set 03 (Table 3), each type of spatial features has been grouped together. The correlation coefficient value of each group is around 0.2 – 0.48. These groups can be separated into three main divisions based on the correlation coefficient value. The first division with a correlation of 0.40 – 0.50 consists of Metric Total Node, Node Count, Metric Total Length, and Metric Total Depth. The second division with a correlation of 0.30 – 0.40 consists of Metric Mean Depth SLW, Integration, Metric Mean Depth, and Total Depth. The last division with a correlation of 0.20 -0.30 consists of Choice, Metric Choice, and Metric Choice SLW.

In the literature review, Choice and Integration have been mentioned that perhaps they are a good candidate to predict land use of an urban street network (Jiangl et al., 2000; Shu, 1999; Desyllas, 1997; Charlambous and Mavridou, 2012). According to the results, a set of choice alone might not be a proper set to predict building uses proportion on the certain street segment. While the correlation coefficient in integration set gains medium-high level compared to other spatial measures, the correlation coefficient in choice set gains a low level, suggesting that integration set probably be a proper set to predict land use, but not choice set alone. Also, unexpectedly, sets of features from Topological analysis such as Metric Total Node gain a high level of correlation compared to other spatial measures. These sets are hence worth to consider as featured correlated with land use.

In the last set, Integration and Choice have been grouped. According to Charalambous and Mavridou (2012), they mentioned that Choice could determine the land use type that would suit best on the location, and specific land uses would possibly expect area with a high integration value. Integration and choice together are hence a set that worth experimenting. Accordingly, choice and integration might be an interesting combination for improving prediction performance

The result, in the same way, has shown that, by using integration and choice as the same group, the correlation coefficient value is better than using each group alone. The value has been improved from 0.259 in choice and 0.374 in integration to 0.393. This finding indicates that integration and choice are perhaps one of a proper set to predict land use.

From the second experiment, this evidence suggested that the correlation coefficient can be improved, and more data probably need for a better result. Also, a set of choice alone might not be a good set of features. By using integration and choice, however, the better result can be gained. Additionally, sets of features worth to be studied more about their implications are a set from Topological analysis such as Metric Total Length and Metric Total node, since they gain a high level of correlation compared to other sets. Finally, strong correlation in training set (unnormalized) is worth to be considered as a syntactic signature features for commercial, residential, and business since this set's prediction performance had exceed other sets. However, because there is a wide range of possibility for combination of features, more investigation is needed.

To portray the prediction of the MLP model, this study generates the map of London using responded data from the MLP model (Figure 8 -10). The MLP used weights and thresholds of the set that gains the best correlation coefficient at 0.517, which is the set of strong correlation in training set (unnormalized) at 37 dimensions.

Both result from training set area and outside training set area have been observed. The result in both cases seems to predict properly, but the prediction of business use outside training area tends to predict something different. According to the results in training set area, most of the prediction result appears to be the same as real-world data, such as the spatial pattern of building use in Camden, and Tottenham. Results outside the training set area are also investigated. They appear to predict the spatial distribution of building use adequately such as the commercial street in Kilburn High Road and commercial street around Maryland station. In business use, however, the prediction is quite varied, yet interesting. Some of the prediction indicates other uses as business such as harbour and factory. One of the interesting observation is that the emergence of false intensity in business use seems to appear in isolated grouping street.

Even though the model has not predicted the proportion of use on street segments all precisely, many are fairly accurate and similar. These results suggest that the spatial predictability of building use is comparable to other human factor such as movement.

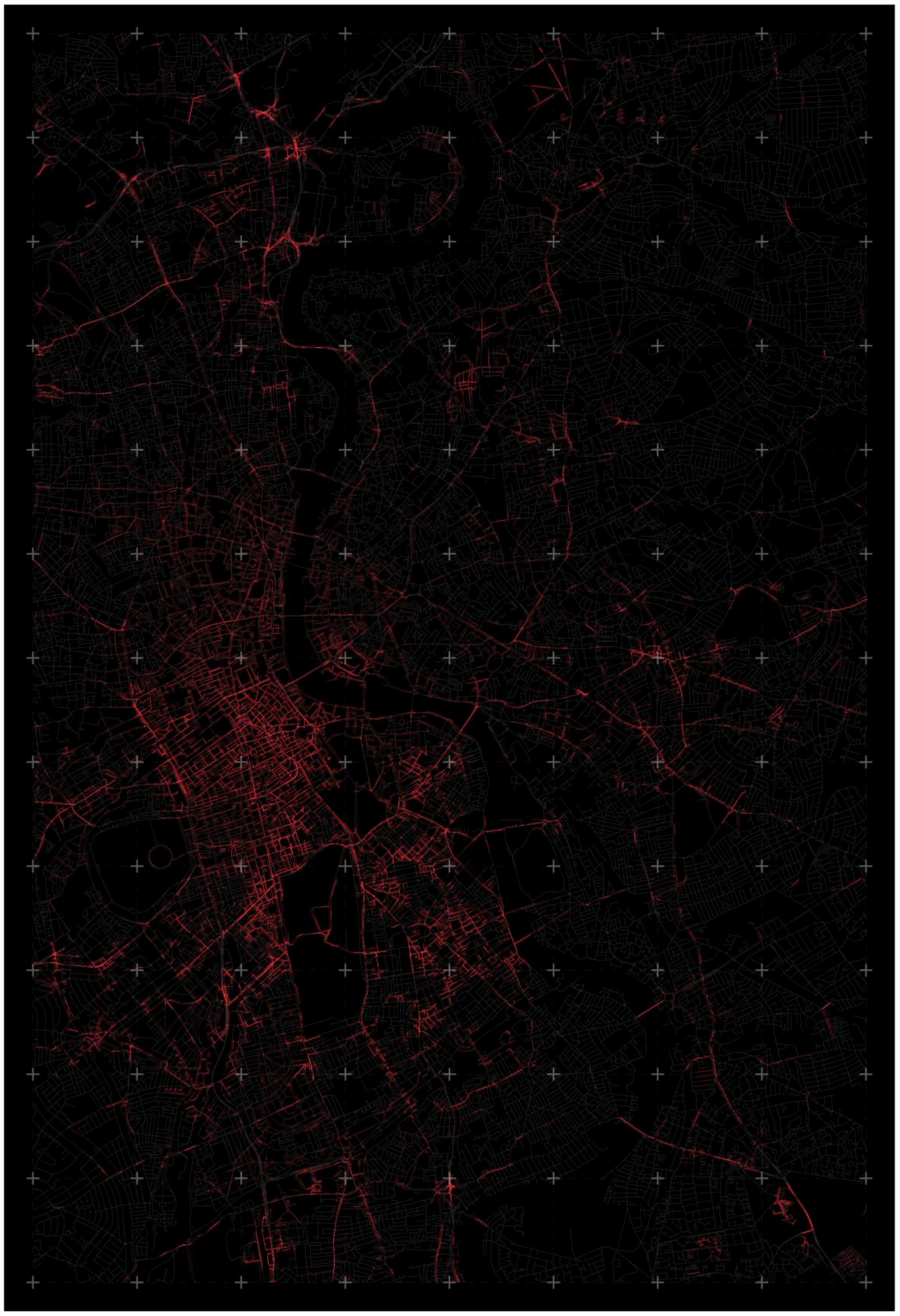

Figure 8 – Recognition and Prediction of use with Machine Learning: Commercial

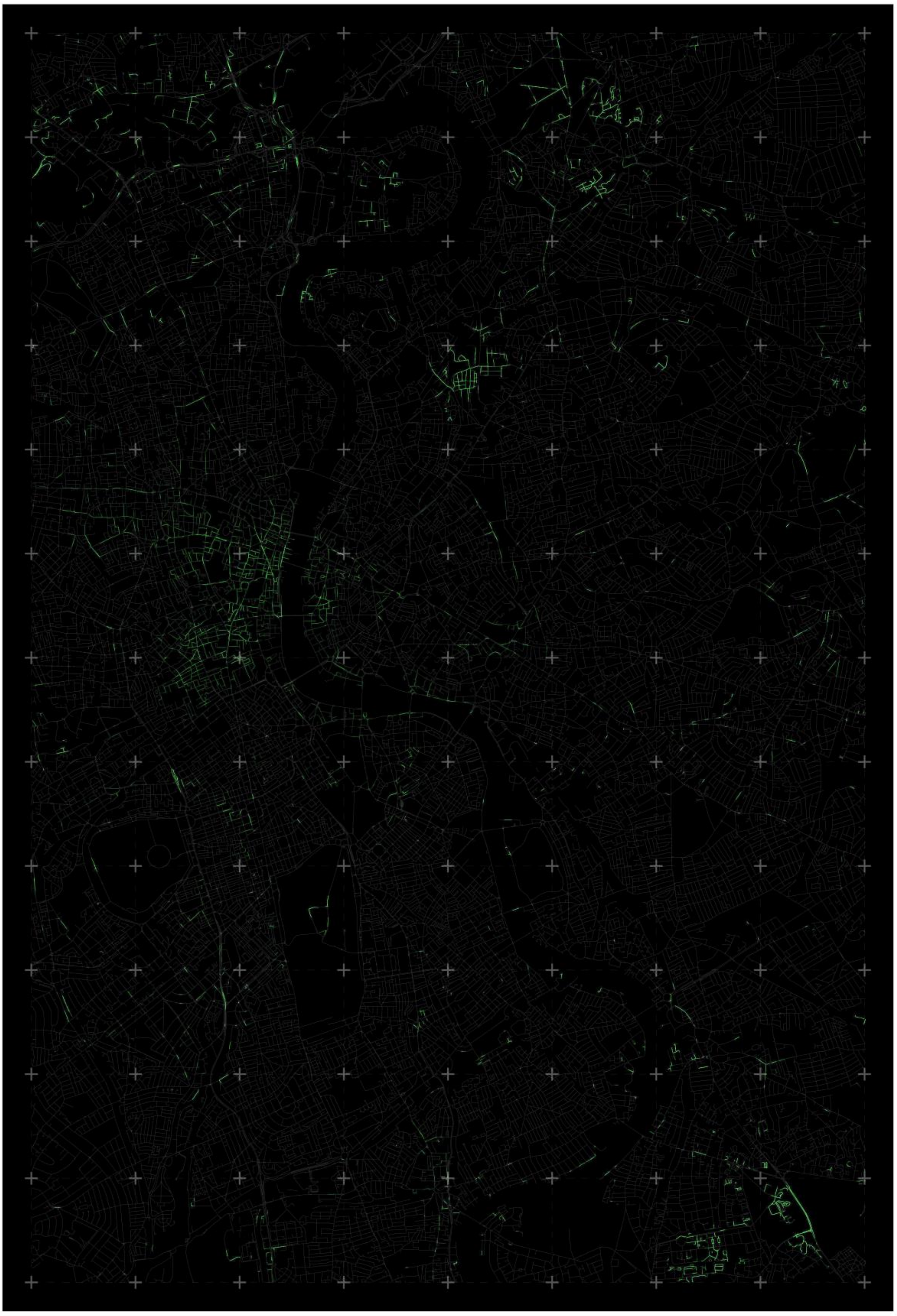

Figure 9 - Recognition and Prediction of use with Machine Learning: Business

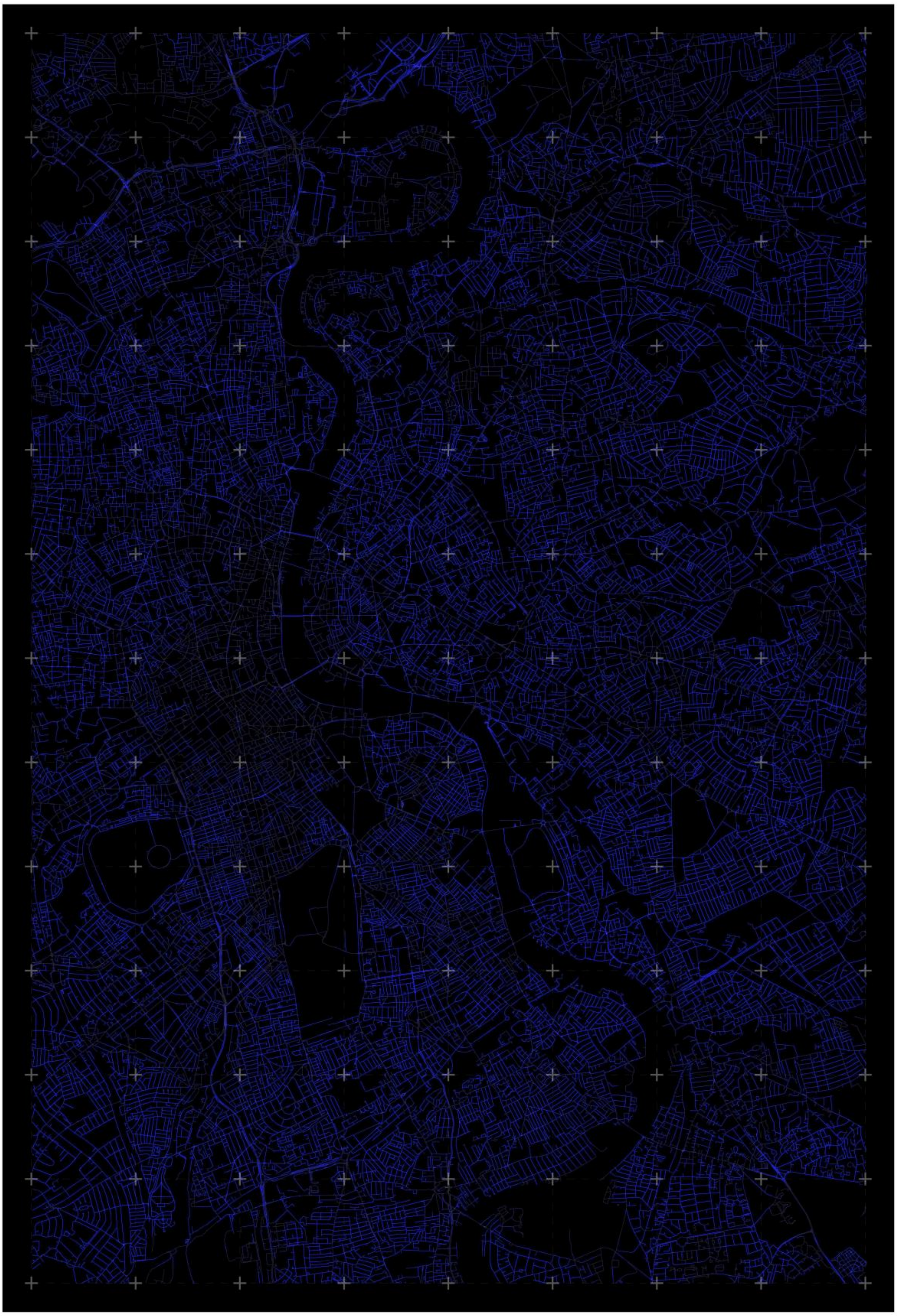

Figure 10 - Recognition and Prediction of use with Machine Learning: Residential

5. CONCLUSIONS

This paper investigates whether spatial distribution of building use has correlations with spatial measurement or not. It also examines which set of spatial measures affects which building use type. More than 10,000 street segments in London have been trained in a multilayer perceptron network .

The result of training indicates an accuracy of approximately 85% and the value of correlation coefficient of 0.696 for the training set and 0.517 for the test set, suggesting that the spatial predictability of building use is comparable to that of other human factors such as movement.

The approximate syntactic signature in each use, correlated radii and best spatial features, are different. For commercial buildings, the length of correlated radii is low to medium radii (200 - 1,800 meters). These radii can be called as a neighbourhood to city radius in the UK. The strongest positive correlation feature in this use us Metric Total Length and the best negative correlation feature is Metric Mean Depth. For residential buildings, the length of correlated radii medium-low to high (400 – 7,200 meters). These radii can be called as betweenness neighbourhood – city to betweenness City-metropolitan. The best positive and negative correlation feature, however, have a reverse relationship with commercial buildings. While the best positive correlation feature is Metric Mean Depth, the best negative correlation feature is Metric Total Length. For business building, the length of correlated radii is distinct, which is medium to high (1800 – 7,200 meters). These radii can be called as a city to betweenness city – metropolitan. Because of the use distinct radii, such radii possibly can be called as Central Business District radius (CBD). Most of the feature in business use has the same direction as commercial use. However, this use has a distinct strongest positive feature, which is Node Count.

From the results and discussion, until now, a set of strong correlation features in training set (unnormalized) is worth to be considered as a syntactic signature for all use (see Appendix 6.1 FEATURE SELECTION SET 02); however, the odds are that other combination of features can exceed such set. More investigation might be necessary.

Since the derivation of better correlation and less error came from experimentation with a different set of spatial features, it indicates that correlation can be improved, and more data is probably essential. However, even though more data is collected, to what extent that spatial measures of street segments alone can indicate and predict the spatial distribution of building use. This spatial pattern might follow a more complex, complicated, and variable relationship than a set of spatial measures. Further research may need to examine the degree of signification between spatial features and spatial distribution of building use. Supplementary factors might be introduced. Thus, a better result should be generated and can be used to inform design and planning.

26

REFERENCES

A Penn, J Hanson, T Grajewski, J Xu B Hillier. (1993). *Natural Movement.* London: a Pion publication.

A. Turner. (2008). *Getting Serious with DepthMap: Segment Analysis and Scripting.* London: UCL.

Abdellatif Abdelfattah. (2017). *Image Classification using Deep Neural Networks.* Medium.

Alan Penn, Bill Hillier Ben Croxford. (1996). *Spatial distribution of urban pollution.* London: ELSEVIER.

Alan Penn, Sam Griffiths, Laura Narvaez. (2015). *The architectural adaptation of urban economic life: Location, use and form of the commercial-residential building in Cardiff.* London: ResearchGate.

Alasdair Turner Joao Pinelo. (2010). *Introduction to UCL Depthmap 10.* London: UCL.

Alasdair Turner, Tao Yang, and Hoon Tae-Park Bill Hillier. (2012). METRIC AND TOPO-GEOMETRIC PROPERTIES OF URBAN STREET NETWORKS: some convergences, divergences, and new results. *The Journal of Space Syntax*.

Bill Hillier. (2007). *Space is the machine* . London: Space Syntax .

Bill Hillier. (2008). *The Social Logic of Space* . London: Cambridge University Press; Reprint edition.

Bill Hillier. (2014). *The Generic city and its origins.* John Wiley & Sons Ltd.

Bill Hillier, Yolandes Barnes, and Christian Schwander Alain Chiaradia. (2009). *Residential Property Value Patterns in London.* Stockholm: Proceedings of the 7th International Space Syntax Symposium .

Björn Klarqvist. (1993). *A Space Syntax Glossary.* NORDISK ARKITEKTURFORSKNING .

Carlo Ratti. (2004). Urban texture and space syntax: some inconsistencies. *Environment and Planning B: Planning and Design*.

Christophe Claramuntz and Bjijrn Klarqvist Bin Jiangl. (2000). *Integrataion of space syntax into GIS for modeling urban spaces.* Elsevier B.V.

Dr Laura Vaughan . (2008). Mapping the East End ‘Labyrinth’. in *Jack the Ripper and the East End Labyrinth’, Museum of London and Random House.* London.

Girmay Kifle Berhie and Saif Haq. (2017). *Land Use and Transport Mode choices: Space Syntax Analysis of American Cities.* Enquiry, The ARCC Jounrnal .

Hong-Kyu Kim and Dong Wook Sohn. (2002). *An analysis of the relationship between land use density of office buildings and urban street configuration.* Elsevier Science Ltd.

Jason Brownlee. (2018). *Machine Learning Mastery with Python.* Mastery Pty Ltd.

Jingyi Lin and Yifang Ban. (2017). Comparative Analysis on Topological Structures of Urban Street Networks. *International Journal of Geo-Information*.

Joao Pinelo Silva. (2017). *The syntactic signature of Starbuck's locations.* Kingdom of Bahrain: University of Bahrain.

Kimon Krenz. (2017). *Region Morphology.* London: Preceeding of the 11th Space Syntax Symposium.

Kwok-Ping Chan, and Anthony Gar-On Yeh Jonathan Cheung-Wai Chan. (2001). *Detecting the Nature of Change in an Urban Environment.* American Society for Photogrammetry and Remote Sensing.

land value and urban morphology Berlin in transition: analysis the relationship between land use. (1997). *Desyllas J.* London: International Symposium on SpaceSyntax.

Magda Mavridou Nadia Charalambous. (2012). *Space Syntax: Spatial Integration Accesibility and Angular Segment Analysis by Metric Distance (ASAMeD).* Cyprus: Accessibility Instruments for Planning Practice.

Roger Grosse, Rajesh Ranganath, and Andrew Y. N Honglak Lee. (2011). *Unsupervised Learning of Hierarchical Representations with Convolutional Deep Belief Network.* communications of the acm .

S Shu. (1999). *Housing layout and crime vunerability.* Proceeding, Second International Symposium on SPace Syntax.

Saul LK Rowies ST. (2000). *Nonliner dimensionality reduction by locally linear embedding.* Science.

Shinichi Iida Bill Hillier. (2005). *Network effects and psychological effects: a theory of urban movement.* London: University College London.

V Jaiganesh M Praveena. (2017). *A literature review on supervised machine learning algorithms and boosting process.* International Journal of Computer Applications.

X., Yang G Xu. (2013). *Robust manifold classification based on semi supervised learning.* International Journal of Advancements in Computing Technology.

28

APPENDICES

6.1 FEATURE SELECTION SET 02

Best correlation in all use	Best correlation in all use
Dimensions	5
Correlation Coefficient	0.496
Error	0.174
Features	1. Metric Total Length at 600 m. 2. Metric Total Length at 7200 m. 3. Metric Mean Depth at 7200 m. 4. Metric Mean Depth SLW at 1800 m. 5. Node Count at 3600 m.

Table 1 - 1a – Feature selection set 02 – Best correlation in all uses set features.

Strong correlation in training set	unnormalized	normalised
Dimensions	37	18
Correlation Coefficient	0.517	0.507
Error	0.166	0.160
Features	1. Metric Choice SLW at 600 m. 2. Metric Mean Depth at 1800 m. 3. Metric Mean Depth at 7200 m. 4. Metric Mean Depth SLW at 1800 m. 5. Metric Total Depth at 200 m. 6. Metric Total Depth at 400 m. 7. Metric Total Depth at 600 m. 8. Metric Total Depth at 1800 m. 9. Metric Total Depth at 3600 m. 10. Metric Total Depth at 7200 m. 11. Metric Total Length at 200 m. 12. Metric Total Length at 400 m. 13. Metric Total Length at 600 m. 14. Metric Total Length at 1800 m. 15. Metric Total Length at 3600 m. 16. Metric Total Length at 7200 m. 17. Metric Total Node at 200 m. 18. Metric Total Node at 400 m. 19. Metric Total Node at 600 m. 20. Metric Total Node at 1800 m. 21. Metric Total Node at 3600 m. 22. Metric Total Node at 7200 m. 23. Integration at 200 m. 24. Integration at 400 m. 25. Integration at 600 m. 26. Integration at 1800 m. 27. Integration at 3600 m. 28. Integration at 7200 m. 29. Node Count at 200 m. 30. Node Count at 400 m. 31. Node Count at 600 m. 32. Node Count at 1800 m. 33. Node Count at 3600 m. 34. Node Count at 7200 m. 35. Total Depth at 400 m. 36. Total Depth at 600 m. 37. Total Depth at 3600 m.	1. Metric Mean Depth at 7200 m. 2. Metric Mean Depth SLW at 1800 m. 3. Metric Total Depth at 3600 m. 4. Metric Total Depth at 7200 m. 5. Metric Total Length at 400 m. 6. Metric Total Length at 600 m. 7. Metric Total Length at 1800 m. 8. Metric Total Length at 3600 m. 9. Metric Total Length at 7200 m. 10. Metric Total Node at 600 m. 11. Metric Total Node at 1800 m. 12. Metric Total Node at 3600 m. 13. Metric Total Node at 7200 m. 14. Integration at 3600 m. 15. Integration at 7200 m. 16. Node Count at 600 m. 17. Node Count at 1800 m. 18. Node Count at 3600 m.

Table 1 - 2a – Feature selection set 02 – Strong correlation in training set features.

Excluded weak correlation	unnormalized	normalised
Dimensions	59	53
Correlation Coefficient	0.495	0.502
Error	0.160	0.165
Features	1. Metric Choice at 200 m. 2. Metric Choice at 400 m. 3. Metric Choice at 600 m. 4. Metric Choice at 1800 m. 5. Metric Choice at 3600 m. 6. Metric Choice at 7200 m. 7. Metric Choice SLW at 200 m. 8. Metric Choice SLW at 400 m. 9. Metric Choice SLW at 600 m. 10. Metric Choice SLW at 1800 m. 11. Metric Choice SLW at 3600 m. 12. Metric Choice SLW at 7200 m. 13. Metric Mean Depth at 1800 m. 14. Metric Mean Depth at 3600 m. 15. Metric Mean Depth at 7200 m. 16. Metric Mean Depth SLW at 1800 m. 17. Metric Mean Depth SLW at 3600 m. 18. Metric Mean Depth SLW at 7200 m. 19. Metric Total Depth at 200 m. 20. Metric Total Depth at 400 m. 21. Metric Total Depth at 600 m. 22. Metric Total Depth at 1800 m. 23. Metric Total Depth at 3600 m. 24. Metric Total Depth at 7200 m. 25. Metric Total Length at 200 m. 26. Metric Total Length at 400 m. 27. Metric Total Length at 600 m. 28. Metric Total Length at 1800 m. 29. Metric Total Length at 3600 m. 30. Metric Total Length at 7200 m. 31. Metric Total Node at 200 m. 32. Metric Total Node at 400 m. 33. Metric Total Node at 600 m. 34. Metric Total Node at 1800 m. 35. Metric Total Node at 3600 m. 36. Metric Total Node at 7200 m. 37. Choice at 200 m. 38. Choice at 400 m. 39. Choice at 600 m. 40. Choice at 1800 m. 41. Choice at 3600 m. 42. Integration at 200 m. 43. Integration at 400 m. 44. Integration at 600 m. 45. Integration at 1800 m. 46. Integration at 3600 m. 47. Integration at 7200 m. 48. Node Count at 200 m. 49. Node Count at 400 m. 50. Node Count at 600 m. 51. Node Count at 1800 m. 52. Node Count at 3600 m. 53. Node Count at 7200 m. 54. Total Depth at 200 m. 55. Total Depth at 400 m. 56. Total Depth at 600 m. 57. Total Depth at 1800 m. 58. Total Depth at 3600 m 59. Total Depth at 7200 m.	1. Metric Choice at 400 m. 2. Metric Choice at 600 m. 3. Metric Choice at 1800 m. 4. Metric Choice at 3600 m. 5. Metric Choice at 7200 m. 6. Metric Choice SLW at 200 m. 7. Metric Choice SLW at 400 m. 8. Metric Choice SLW at 600 m. 9. Metric Choice SLW at 1800 m. 10. Metric Choice SLW at 3600 m. 11. Metric Choice SLW at 7200 m. 12. Metric Mean Depth at 1800 m. 13. Metric Mean Depth at 3600 m. 14. Metric Mean Depth at 7200 m. 15. Metric Mean Depth SLW at 1800 m. 16. Metric Mean Depth SLW at 3600 m. 17. Metric Mean Depth SLW at 7200 m. 18. Metric Total Depth at 200 m. 19. Metric Total Depth at 400 m. 20. Metric Total Depth at 600 m. 21. Metric Total Depth at 1800 m. 22. Metric Total Depth at 3600 m. 23. Metric Total Depth at 7200 m. 24. Metric Total Length at 200 m. 25. Metric Total Length at 400 m. 26. Metric Total Length at 600 m. 27. Metric Total Length at 1800 m. 28. Metric Total Length at 3600 m. 29. Metric Total Length at 7200 m. 30. Metric Total Node at 200 m. 31. Metric Total Node at 400 m. 32. Metric Total Node at 600 m. 33. Metric Total Node at 1800 m. 34. Metric Total Node at 3600 m. 35. Metric Total Node at 7200 m. 36. Choice at 600 m. 37. Integration at 400 m. 38. Integration at 600 m. 39. Integration at 1800 m. 40. Integration at 3600 m. 41. Integration at 7200 m. 42. Node Count at 200 m. 43. Node Count at 400 m. 44. Node Count at 600 m. 45. Node Count at 1800 m. 46. Node Count at 3600 m. 47. Node Count at 7200 m. 48. Total Depth at 200 m. 49. Total Depth at 400 m. 50. Total Depth at 600 m. 51. Total Depth at 1800 m. 52. Total Depth at 3600 m 53. Total Depth at 7200 m.

Table 1 - 3a – Feature selection set 02 – Exclude weakness correlation set features

30