Enquiries:
If you have questions about this document, contact openresearch@mmu.ac.uk. Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines)

Modelling Road Congestion using a Fuzzy System and Real-World Data for Connected and Autonomous Vehicles

Luke Abberley Member, IEEE, Keeley Crockett SMIEEE, Jianquan Cheng
Science and Engineering, Manchester Metropolitan University, UK
{Luke.Abberley@mmu.ac.uk, K.Crockett@mmu.ac.uk, J.Cheng@mmu.ac.uk}

Abstract - **Road congestion is estimated to cost the United Kingdom £307 billion by 2030. Furthermore, congestion contributes enormously to damaging the environment and people's health. In an attempt to combat the damage congestion is causing, new technologies are being developed, such as intelligent infrastructures and smart vehicles. The aim of this study is to develop a fuzzy system that can classify congestion using a real-world dataset referred to as Manchester Urban Congestion Dataset, which contains data similar to that collected by connected and autonomous vehicles. A set of fuzzy membership functions and rules were developed using a road congestion ontology and in conjunction with domain experts. Experiments are conducted to evaluate the fuzzy system in terms of its precision and recall in classifying congestion. Comparisons are made in terms of performance with traditional classification algorithms decision trees and Naïve Bayes using the Red, Amber, and Green classification methods currently implemented by Transport for Greater Manchester to label the dataset. The results have shown the fuzzy system has the ability to predict road congestion using volume and journey time, outperforming both decision trees and Naïve Bayes.**

Keywords - Intelligent Transport Systems; Big Data; Fuzzy System; Urban Road Network; Congestion

I. INTRODUCTION

For centuries, people have naturally been migrating from rural to urban areas causing the natural occurrence of urbanization, which has contributed to one of the biggest challenges' society faces each day, which is road congestion. Road congestion in urban areas is estimated to cost the UK economy a total of £307 billion by 2030 [1]. Furthermore, road congestion contributes enormously to damaging the environment, due to air pollution which has an impact on peoples well-being [2], [3].

In an attempt to reduce the impact of road congestion, many large corporations, such as Google, Tesla, and Uber are developing *'smart vehicles'*, such as connected and autonomous vehicles (CAVs) that will be implemented as part of an Intelligent Transport System (ITS) of the future. Smart vehicles are expected to reduce congestion levels and the number of fatal accidents on the roads, with an estimated 37,000 lives a year predicted as being saved in the United States (U.S.) alone [4]. This is due to smart vehicles being able to communicate faster than a human and make better decisions based on information collected by sensors embedded within vehicles with other vehicles and infrastructure [5]. However, due to the limited access to these smart vehicles and their associated infrastructure, this study will use alternative data sources, which comprise of data similar to what is collected by CAVs and Road Side Units (RSUs) that will be used within ITS of the future, such as a VANETs. Furthermore, these types of ITS will provide data from vehicle to vehicle (V2V) and vehicle to infrastructure (V2I) providing a constant stream of big data [5]–[7], which can be used to provide different information, such as volume, journey time, speed, and weather conditions, which are also known as dimensions.

Little work has been conducted using fuzzy systems to model road congestion [8]–[10]. However, this limited work has indicated that fuzzy models of road congestion are better for a stakeholder, such as a domain expert to understand that the conventional quantitative models previously implemented, such as the probability model [11] and the spatial-temporal model [12]. Fuzzy sets are the ideal choice for modelling road congestion because of their ability to handle the ambiguity, multifaceted nature, and uncertainty within traffic data, such as journey time, speed, traffic flow, accidents, road works etc. They have the ability to capture such variables through the use of linguistic variables and hedges which are easier for a domain expert to understand [13].

The contribution of this paper is taking an unbalanced real-world big dataset with the support of an ontology and domain experts to construct a fuzzy model of road congestion. This is achieved through the construction of fuzzy systems comprises of a set of fuzzy membership functions and fuzzy rules that can be used to identify road congestion. An experiment is conducted to determine whether the fuzzy model can be used to analyse traffic data to classify congestion. Comparisons are made with an existing system used by Greater Manchester Transport authority in the UK and other known classification algorithms.

This paper is organised as follows: Section II provides a brief overview of recent studies of fuzzy systems within the field of transportation. Section III provides an in-depth explanation for developing a fuzzy system capable of capturing the levels of congestion on an urban road network. Section IV presents the experimental methodology and Section V provides the results and a comparative discussion. Section VI concludes this study and provides insight into further work.

II. Fuzzy Systems in Transportation

A. What is a fuzzy system?

A fuzzy system is typically a control system based on fuzzy logic. The term "fuzzy" refers to the system's ability to deal with terms that are not binary or predefined- often referred to as linguistic variables [13]. For instance, a humans' understanding of the phrase, near or far, could imply: very near, near, not near, far, and very far. Hence, fuzzy terms are subjective and mean different things to different people. The main advantage of a fuzzy system is that the model itself is made up of a number of fuzzy rules, which can model a problem, such as congestion that can be expressed in terms a human operator can understand.

B. Transport applications

The approach to use fuzzy systems within the discipline of transportation to classify road traffic congestion is a relatively new field. For instance, a study was presented in [14] into a cooperative V2V road traffic detection congestion on freeways. The study uses a level of service metric created by a third party that collected aerial surveys to define the levels of congestion: slight, moderate, and severe. The author then created a new metric that uses four membership function: VerySlow, Slow, Medium, and Fast, two inputs: Speed and Density, and sixteen rules to define an output for one of three levels of congestion. However, this study does not consider non-congestion as an output and has described only testing the model in a simulation with simulated data, furthermore, the focus of the study is on highways and does not reflect an urban road network, which has very different characteristics. Another study [9], examined road traffic anomalies that contribute to congestion at a single junction using a one-way traffic video sequence. This study uses two data inputs: Traffic flow and traffic density. Traffic flow has three membership functions called low, medium, and high. Traffic density also has three membership functions (sparse, normal, and dense) which are calculated using a statistical analysis of the pixels. This study uses nine rules, which were obtained through experts and empirical experiments and has an output of either: normal traffic, slight congestion, and heavy congestion. However, one of the limitations of this experiment is it was only tested on 3 different scenes and in total had 142 observations. Although the use of fuzzy systems is very new, despite limitations in current work, fuzzy set representation of the variables that model congestion encapsulate a greater human understanding of a multifaceted and dynamic environment.

III. Developing a Fuzzy System for Congestion

This section describes the methodology that was used to develop a fuzzy system for road congestion on an urban city network. The model utilises real-world data from Bluetooth sensors and inductive loop counters provided by Transport for Greater Manchester (TfGM) for Manchester, UK. These data sources will provide data that is equivalent to what will be provided by CAVs and RSUs. Moreover, experts in road congestion management (TfGM) and a road congestion ontology [3], [15] was used to help define the fuzzy sets to ensure thorough domain coverage. The road congestion ontology which was used to support the development of a fuzzy system capable of classifying road congestion was presented in [15]. The road congestion ontology states that congestion can be measured using multiple dimensions, such as journey time and volume. Furthermore, congestion is often the consequence of an event, such as rush hour, a road accident, a concert, a football match and roadworks. Finally, depending on the severity of congestion the magnitude can vary from very low to very high. Therefore, in this study, the magnitude ranges defined in the urban road congestion ontology [15] will be used to determine the membership functions: Very low, *VL*, low, *L*, medium, *M*, high, *H*, and very high, *VH* which will ensure coverage of the domain.

A. Data Sources and Variables

For this study, a real-world spatial-temporal dataset, known as the Manchester Urban Congestion dataset (MUCD) was used. The dataset consists of journey time, volume, weather, bank holidays, and event information. However, due to the nature of this study, only journey time and volume data collected from Bluetooth sensors and IDC will be used to simulate data collected by CAVs and RSUs. The MUCD has 17376 records and each record consists of 126 attributes. Furthermore, the MUCD dataset is labelled using the Red, *R*, Amber, *A*, and Green, *G*, (RAG) method implemented by TfGM, UK. Where (G)reen is non-congestion (1), (A)mber is slight congestion (2), and (R)ed is major congestion (3).

$$G = JT \leq median * 1.25 \qquad (1)$$
$$A = Median * 1.25 < JT \leq Median * 1.5 \qquad (2)$$
$$R = JT > Median * 1.5 \qquad (3)$$

Where *JT* is the average journey time for all Bluetooth enabled vehicles travelling between two sensors on each link. The *Median* is the 50th percentile of journey time for a single link within the MUCD. 1.25 and 1.5 are the congestion factors that TfGM experts use to measure network performance. The problems associated with the MUCD can be summarised as:

- Due to the limited number of inductive loops counters, the ability to calculate the volume of traffic for each link in the network is limited.
- The data quality of the Bluetooth sensors has many issues. For example, capture rates; during the night periods or a period where no vehicle with a Bluetooth device passes the sensors cause the sensors to provide an incorrect average journey time when being observed.
- In bad weather, the sensors which use a mobile network to transmit the data to a central location, can fail and cause the dataset to have missing data.
- The non-congestion class significantly outweighs the other, causing the MUCD dataset to be imbalanced, which imposes challenges for machine learning classification algorithms that is a problem because classification algorithms are often biased towards the majority class, which in this study is non-congestion.

Dimension (Variables)	Data sources	Linguistic values (Membership functions)
Journey time	Bluetooth remote sensors	Very Low (*VL*) Low (*L*) Medium (*M*) High (*H*) Very High (*VH*)
Volume	Inductive loop counters	Very Low (*VL*) Low (*L*) Medium (*M*) High (*H*) Very High (*VH*)

B. Membership Function Determination

Table I shows the dimensions, data sources, and linguistic values determined from the urban road congestion ontology [15]. The linguistic values of the membership functions representing journey time and volume are also shown.

Using the linguistic values identified in Table I, the creation of the fuzzy membership functions can be performed using three steps:

- **Step 1**: Perform K-means clustering [16] on both journey time and volume data.
- **Step 2**: Identify the final boundary values for a set of clusters (referred to in this work as groups) where they connect and define this value as *dt (boundary threshold)*.
- **Step 3**: Using the *dt* value, determine membership function domain coverage using one of three membership functions: linear increasing, linear decreasing, and trapezoidal.

The primary objective of data mining is to discover patterns within large datasets, such as the MUCD dataset used within this study. K-means clustering is an unsupervised algorithm used within data mining to find a cluster of patterns in data. K-means uses the inherent structures in the data to best organise the data into groups of maximum commonalities [17]. This is achieved by partitioning *n* observations into *k* (in this study *k=5*) clusters. Five clusters were used based upon early empirical experiments which found that five clusters provided sufficient resolution [15]. Figure 1 shows, as an example, 17376 journey time records plotted on a 24-hour scale. Each observation within Figure 1 belongs to the cluster with the nearest mean value. Once K-means has been performed, it becomes possible to identify the boundary values between each cluster, which will be used to create the membership functions in the fuzzy system.
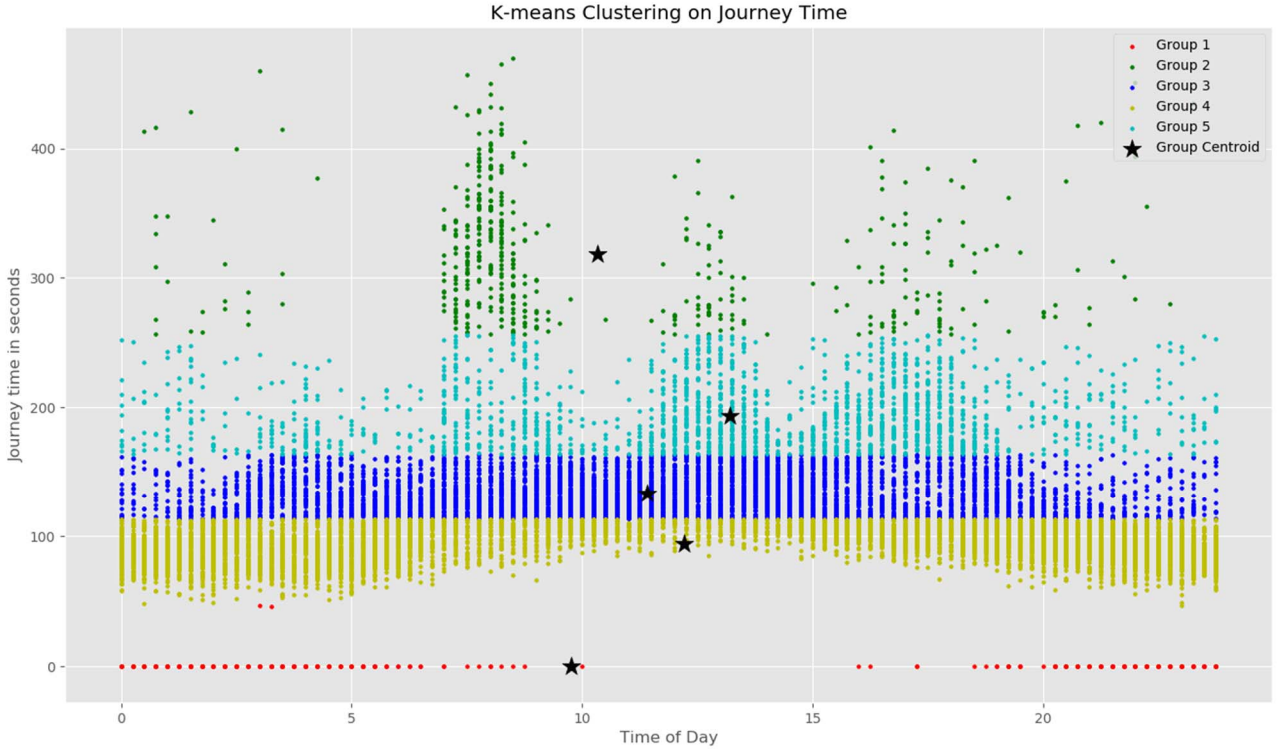

Figure 1: Example of k-means clusters on 6 months' worth of journey time data where k=5.

Figure 2 shows an example pair of linear opposing membership functions, which will be used for the *VL* and *VH* memberships. The two pairs (4) and (5) are both linear increasing and decreasing membership functions *L*, can be defined as [18]:

$$L \uparrow (x, dm, dn) = \begin{cases} 0, & x \leq dm \\ \dfrac{x - dm}{dn - dm}, & dm \leq x \leq dn \\ 1, & x \geq dn \end{cases} \quad (4)$$

$$L \downarrow (x, dm, dn) = \begin{cases} 1, & x \leq dm \\ 1 - \dfrac{x - dm}{dn - dm}, & dm \leq x \leq dn \\ 0, & x \geq dn \end{cases} \quad (5)$$

Where *dm* is defined as $dm = dt - n\sigma$ and *dt* is the value generated by K-means clustering on all variable *i* records. *n* is a real number $n \rightarrow [0.0, \infty]$, σ is the standard deviation, and *x* is the value of the variable *i*. *n* is empirically determined. Additionally, *dn* is defined as $dn = dt + n\sigma$.
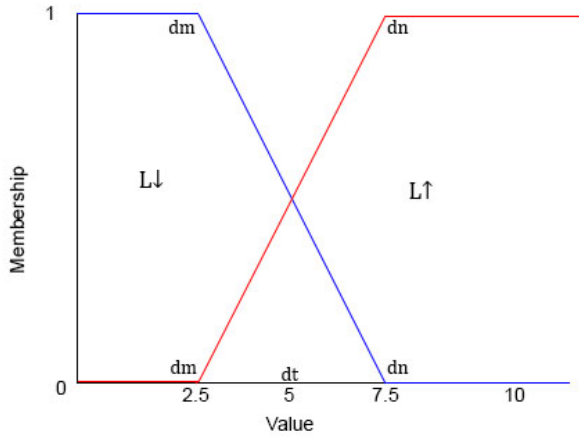

Figure 2: Example of a linear pair opposing fuzzy memberships functions.

Figure 3 shows an example of a trapezoidal-shaped membership function, which will be used for the *L*, *M*, and *H* memberships. The trapezoidal-shaped membership function *T* (6), may be defined as:

$$T(x, dm^1, dn^1, dm^2, dn^2) = \begin{cases} 0, & x \leq dm^1 \\ \dfrac{x - dm^1}{dn^1 - dm^1}, & dm^1 \leq x \leq dn^1 \\ 1, & dn^1 \leq x \leq dm^2 \\ 1 - \dfrac{x - dm^2}{dn^2 - dm^2}, & dm^2 \leq x \leq dn^2 \\ 0, & x \geq dn^2 \end{cases} \quad (6)$$

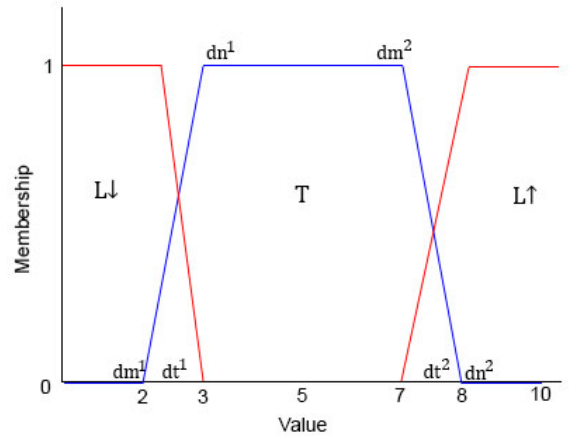Where *dm1*, *dn1*, *dm2*, and *dm2* are defined using the same method as *dm* and *dn*.

Figure 3: Example of a trapezoidal-shaped membership function.

C. Fuzzy Rules Determination (manual and expert)

The fuzzy rules were initially created with every possible variation of each five membership functions, such as *VL*, *L*, *M*, *H*, and *VH* for journey time and volume, which gave a total of 25 rules. However, with the support of the urban road congestion ontology [15] and domain experts, TfGM [19], the rules were humanly optimised down to just six. This manual optimisation revealed that several rules were not firing so therefore, they were not relevant. For example, if journey time was *VH* then the output is congested regardless of the volume.

Algorithm 1 uses both antecedents and consequents membership functions to fire six unique rules to acquire each rule strength ready for fuzzy inference.

Algorithm 1
Rules for congestion

Antecedents: Journey time, *JT*. Volume, *V*.
Antecedents memberships: Very low, *VL*. Low, *L*. Medium, *M*, High, *H*. Very high, *VH*.
Consequents: Congestion, *C*.
Consequents memberships: Congested, Con. Non-congested, *Non*.

IF JT is VH **THEN** Con
IF JT is H **THEN** Con
IF JT is M **AND** V is VH **THEN** Con
IF JT is M **AND** V is **NOT** VH **THEN** Non
IF JT is L **THEN** Non
IF JT is VL **THEN** Non

D. Fuzzy inference

One of the first control systems and most commonly implemented methods for computing fuzzy inference is Mamdani [20]. Furthermore, Mamdani was first implemented within the transport domain, where it was used in an attempt to control a steam engine and boiler combination [20].
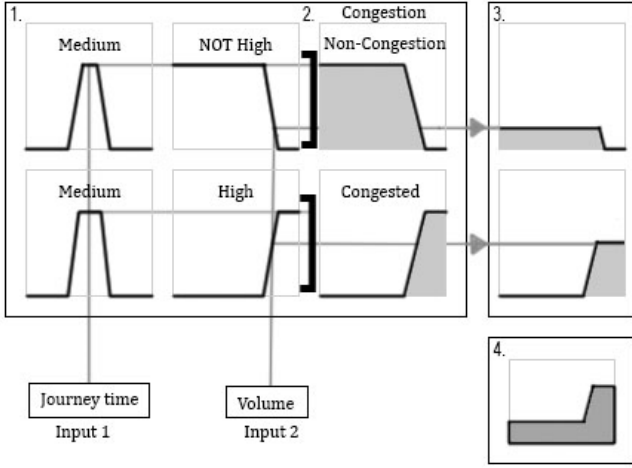
Figure 4: A example of how Mamdani fuzzy inferences works.

Figure 4 shows the composition of fuzzy inference, the four stages are:

Stage 1. Fuzzification of the non-fuzzy inputs (the average journey time and volume over a 15-minute slot for one link in the network), which are crisp, numerical, and specific to the attribute domain. The inputs are fuzzified according to membership functions.

Stage 2. If the antecedent of a given rule has more than one part, the application of a fuzzy operator is required to obtain a single value that represents the individual rule. For instance, the top rule within Figure 4 has two parts in the antecedent, so a AND operator is used to identify the minimum value as the result.

Stage 3. Using the single value acquired in stage 2, the consequent is reshaped to provide the result of implication which is weighted depending on the linguistic characteristics that are attributed to it.

Stage 4. Aggregation is the combination of the fuzzy sets that represent the outputs of each rule into a single fuzzy set (fuzzy output distribution).

Stage 5. The input for defuzzification is the single aggregated fuzzy set and the output is a single value. This is discussed in more detail in section E.

E. Defuzzification

The method centroid of area (COA), also known as the centre of gravity (COG) (7) is used to defuzzify the final output fuzzy set (Figure 4) and output a crisp numeric value, which in this study is the probability of congestion. To achieve this, the total area of the output distribution membership is divided into a number of sub-areas and then the COA is calculated for each sub-area. Finally, all sub-areas COA are summed together to find the defuzzied value (probability of congestion).

$$Z^* = \frac{\int \mu_{\bar{A}}(z) \cdot z \, dz}{\int \mu_{\bar{A}}(z) \, dz} \qquad (7)$$

Where μ is defined as the degree of membership (y-axis), z is defined as the value on the x-axis, $\bar{A}$ is the fuzzy set, and dz is the derivative of z.

IV. EXPERIMENTAL METHODOLOGY

The aim of the experiment is to determine whether a fuzzy system can be used to analyse traffic data to classify congestion. The hypothesis for this study is H_1: Using journey time and volume data, it is possible to classify congestion using a fuzzy system. To evaluate the performance of the fuzzy system, it was compared against two alternative machine-learning algorithms: The decision tree C4.5 (using the Weka implementation J48) [21] and Naïve Bayes, on the MUCD dataset. The training and testing strategy is described as follows: the MUCD dataset was split into two parts: Training Set containing 8688 records of which 6665 were classified as non-congestion and 2023 were classified as congestion, which accounts for only 23% of records. The test set containing the remainder of the dataset. Datasets were mutually exclusive.

In order to evaluate the three methods using an unbalanced dataset, five statistical measurements were chosen, which are: True Positive Rate, *TPR*, also known as recall and sensitivity. TPR measures the proportion of actual positives that are correctly identified. TPR is defined in equation (8) where *TP* is true positive, and *FN* is false negative.

$$TPR = \frac{TP}{TP + FN} \qquad (8)$$

False Positive Rate, *FPR,* measures the negative instance that is wrongly classified as positive. FPR is defined in equation (9) where *FP* is false positive, and *TN* is true negative.

$$FPR = \frac{FP}{FP + TN} \qquad (9)$$

Precision, also known as positive predictive value, *PPV,* measures the number of positive predictions divided by the total number of positive class values predicted. Precision is defined in equation (10).

$$PPV = \frac{TP}{TP + FP} \qquad (10)$$

F-measure, also known as F1 Score, *F1,* measures the balance between the precision and TPR. F-measure is defined in equation (11).

$$F1 = \frac{2 * TP}{2 * TP + FP + FN} \qquad (11)$$

Overall efficiency, also known as accuracy measures the amount of correctly classified instances. Overall efficiency is defined in equation (12)

$$Overall\ Efficiency = \frac{TP + TN}{TP + TN + FP + FN} \qquad (12)$$

However, due to the class imbalance as mentioned above, it is important to provide a single value that represents the performance of both classes for TPR, False Positive Rate, precision, and F-measure. To achieve this a weighted average will be used and is defined in equation (13). Where C_{non} represents the statistical measurement being weighted for the class non-congestion. C_{con} represents the statistical measurement being weighted for the class congested.

$$W = \frac{C_{non} * (TP + FN) + C_{con} * (TN + FP)}{TP + FN + TN + FP} \qquad (13)$$

V. RESULTS AND DISCUSSION

The purpose of this study was to determine whether it is possible to classify road congestion using a fuzzy system and real-world traffic data. Table II shows the results for each statistical measurement for the three machine-learning algorithms and their classes: Non-congestion (Figure 5) and Congested (Figure 6) and the weighted average of both classes (Figure 7) defined in equation (13).

Before discussing the results, the authors would like to reiterate the challenges of performing classification on an imbalanced dataset. Global performance measurements, such as overall efficiency, provides an advantage to the majority class and can be misleading. For example, the overall efficiency of the fuzzy system is 88 per cent, which seems good. However, assume the dataset had 100 instances, with a split of 80 for non-congestion and 20 for congestion. Assume the system classifies non-congestion as 92 instances and congested as eight instances. This means the class, congested is only 40 per cent efficient/accurate and not 88 per cent. Therefore, the discussion will focus on TPR, FPR, precision, F-measure.

The results show Naïve Bayes achieved a TPR of 99.8 per cent for non-congestion, which is the higher TPR across all algorithms and both classes. However, it achieved the second highest FPR of 57.8 per cent. This is attributed to the paradox of imbalanced datasets. The FPRs for the minority class across all three algorithms are significantly low, for instance, the fuzzy system is 5.5 per cent, the decision tree is 4.8 per cent, and the Naïve Bayes is 0.2 per cent. The FPRs for the majority class across all three algorithms are noticeably higher, for instance, the fuzzy system is 32.9 per cent, the decision tree is 59 per cent, and the Naïve Bayes is 57.8 per cent. Because of these noticeable differences, it has been decided from this point to only compare the weighted averages of both classes. The TPR weighted average for the fuzzy system is 88 per cent, which is higher than both, the decision tree by ≈6 per cent and Naïve Bayes by ≈2 per cent. The FPR weighted average for the fuzzy system is 26.5 per cent, which is lower than both, the decision tree by ≈20 per cent and Naïve Bayes by ≈18 per cent. The precision weighted average for the fuzzy system is 87.6 per cent, which is higher than the decision tree by ≈6 per cent, however, it was lower than the Naïve Bayes by ≈1 per cent. The F-measure weighted average for the fuzzy system was 87.6 per cent and is higher than both the decision tree by ≈7 per cent and Naïve Bayes by ≈3 per cent. Furthermore, the fuzzy system overall efficiency was the highest of all three machine-learning algorithms.

Although all algorithms perform to a similar level with the fuzzy system performing the best overall, it should be noted that each algorithm has its own level of complexity with some stakeholders possibly struggling to understand how the model produces an explainable decision. For instance, the easiest of the three algorithms for a stakeholder to understand is the fuzzy system. This is because the rules are comprised of linguistic variables, which are easier to understand and interpret by stakeholders. The single defuzzied output of the fuzzy system gives a measure of the probability that congestion occurs in a specific 15-minute slot on road link x, where x is a road link on the urban network being modelled. The second easiest to understand is the decision tree, J48, where a branch of the tree is split based on a value of the variable being used and this is repeated until the leaves are reached and an outcome is decided. It should be noted the bigger the tree and the more leaves (and hence rules) the harder it is to understand the decision transparency and hence, may become harder for stakeholders to follow. The decision tree model in this experiment has a tree size of 17 and a total of 9 leaves. The 9 rules are transparent and could be understood by a transport expert. The most complex algorithm for a stakeholder to understand is Naïve Bayes because it is a probabilistic classifier, which uses a probability distribution over a set of classes, instead of only outputting the most likely class that the observation should belong to.
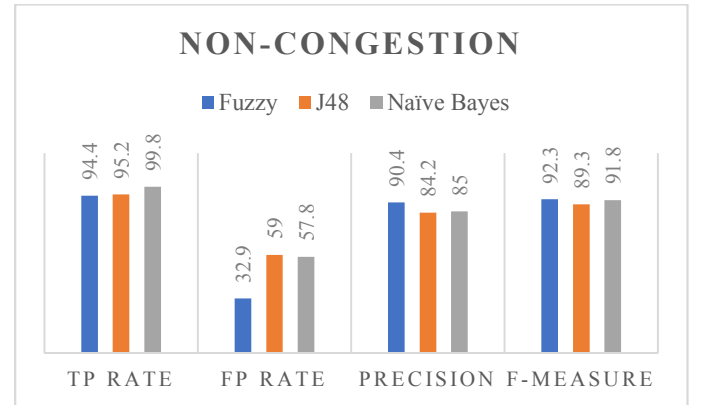

Figure 5: TP rate, FP rate, precision, F-measure, and overall efficiency for non-congestion.

TABLE II.   RESULTS FOR FUZZY SYSTEM, J48, AND NAÏVE BAYES.

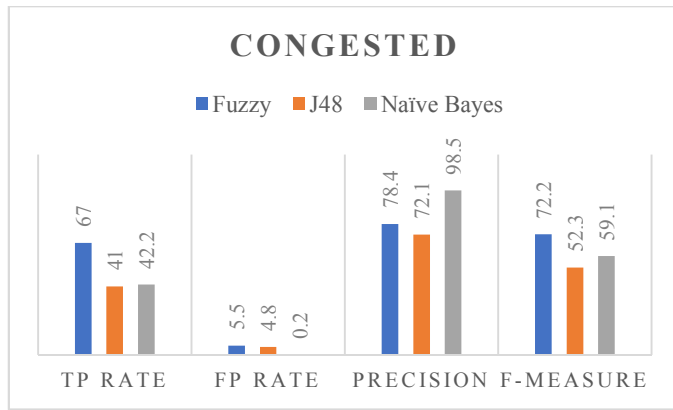| Experiment	Class	TP Rate / Recall (%)	FP Rate (%)	Precision (%)	F-Measure (%)	Overall Efficiency (%)
Fuzzy System	*Non*	94.4	32.9	90.4	92.3	88.0
	Congested	67.0	5.5	78.4	72.2	
	Weighted Avg.	88.0	26.5	87.6	87.6	
Decision tree (J48)	*Non*	95.2	59.0	84.2	89.3	82.5
	Congested	41.0	4.8	72.1	52.3	
	Weighted Avg.	82.6	46.4	81.4	80.7	
Naïve Bayes	*Non*	99.8	57.8	85.0	91.8	86.3
	Congested	42.2	0.2	98.5	59.1	
	Weighted Avg.	86.4	44.4	88.2	84.2	

Figure 6: TP rate, FP rate, precision, F-measure, and overall efficiency for congested.
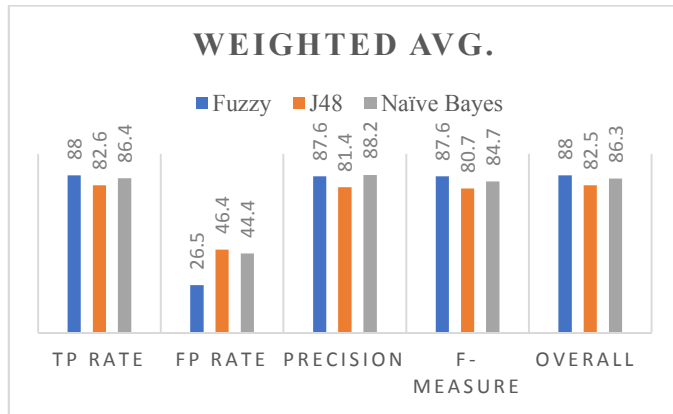

Figure 7: Weighted average of TP rate, FP rate, precision, F-measure, and overall efficiency.

VI. CONCLUSION AND FURTHER WORK

This study has proven the hypothesis, H_I: Using journey time and volume data, it is possible to classify congestion using a fuzzy system and has demonstrated a proof of concept fuzzy model. The initial results have demonstrated the fuzzy systems ability to predict congestion using volume and journey time, outperforming both the decision tree and Naïve Bayes. Moreover, the fuzzy system using, only six rules was able to handle an unbalanced dataset. Additionally, the author believes it would be possible to implement this model on other urban road networks. To further this study, the authors are currently working on expanding the system to classify the three types of congestion [15]: Non-recurrent, Recurrent, and Semi-recurrent. This is an important requirement for TfGM who would benefit from not only being able to identify congestion but the type of congestion, which would allow for different mitigation strategies to be put in place. Additionally, they will be able to measure how much of the network is, at a given time, exhibiting signs of non-congestion, recurrent, non-recurrent, and semi-recurrent congestion. To achieve this goal, the fuzzy system will be expanded to add linguistic variables for different times of day, different days of the week, bank holidays, distance from an attraction, and direction of traffic flow.

REFERENCES

[1] S. Djahel, A. Jones, Y. Hadjadj-Aoul, and A. Khokhar, "CRITIC: A cognitive radio inspired road traffic congestion reduction solution," *IFIP Wirel. Days*, vol. 2018–April, no. February, pp. 151–157, 2018.

[2] L. Rui, Y. Zhang, H. Huang, and X. Qiu, "A new traffic congestion detection and quantification method based on comprehensive fuzzy assessment in VANET," *KSII Trans. Internet Inf. Syst.*, vol. 12, no. 1, pp. 41–60, 2018.

[3] N. Gould and L. Abberley, "The semantics of road congestion," in *UTSG*, 2017.

[4] R. Mudge, D. Montgomery, E. Groshen, J. P. Groshen, S. Helper, and C. Carson, "America's Workforce and the Self-Driving Future Realizing Productivity Gains and Spurring Economic Growth," no. june, 2018.

[5] S. Djahel, R. Doolan, G.-M. Muntean, and J. Murphy, "A Communications-Oriented Perspective on Traffic Management Systems for Smart Cities: Challenges and Innovative Approaches," *IEEE Commun. Surv. Tutorials*, vol. 17, no. 1, pp. 125–151, 2015.

[6] K. Golestan, R. Soua, F. Karray, and M. S. Kamel, "Situation awareness within the context of connected cars: A comprehensive review and recent trends," *Inf. Fusion*, vol. 29, pp. 68–83, May 2015.

[7] N. Isa, M. Yusoff, and A. Mohamed, "A Review on Recent Traffic Congestion Relief Approaches," *2014 4th Int. Conf. Artif. Intell. with Appl. Eng. Technol.*, vol. i, no. November, pp. 121–126, 2014.

[8] P. Pongpaibool, P. Tangamchit, and K. Noodwong, "Evaluation of road traffic congestion using fuzzy techniques," *IEEE Reg. 10 Annu. Int. Conf. Proceedings/TENCON*, pp. 1–4, 2007.

[9] Y. Li, T. Guo, R. Xia, and W. Xie, "Road Traffic Anomaly Detection Based on Fuzzy Theory," *IEEE Access*, vol. 6, pp. 40281–40288, 2018.

[10] Y. Sun, M. Hrušovský, C. Zhang, and M. Lang, "A Time-Dependent Fuzzy Programming Approach for the Green Multimodal Routing Problem with Rail Service Capacity Uncertainty and Road Traffic Congestion," *Complexity*, vol. 2018, pp. 1–22, 2018.

[11] L. Li, "Research on traffic congestion mathematical model in traffic signal control system," *Int. J. Smart Home*, vol. 9, no. 12, pp. 279–288, 2015.

[12] B. Anbaroğlu, T. Cheng, and B. Heydecker, "Non-recurrent traffic congestion detection on heterogeneous urban road networks," *Transp. A Transp. Sci.*, vol. 11, no. 9, pp. 754–771, 2015.

[13] L. A. Zadeh, "Probability Measures of Fuzzy Events," *JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS*, vol. 23. pp. 421–427, 1968.

[14] R. Bauza, J. Gozalvez, and J. Sanchez-Soriano, "Road traffic congestion detection through cooperative Vehicle-to-Vehicle communications," *Proc. - Conf. Local Comput. Networks, LCN*, pp. 606–612, 2010.

[15] L. Abberley, N. Gould, K. Crockett, and J. Cheng, "Modelling road congestion using ontologies for big data analytics in smart cities," in *2017 International Smart Cities Conference (ISC2)*, 2017, pp. 1–6.

[16] S. Khalifa, Y. Elshater, K. Sundaravarathan, A. Bhat, P. Martin, F. Imam, D. Rope, M. Mcroberts, and C. Statchuk, "The Six Pillars for Building Big Data Analytics Ecosystems," *ACM Comput. Surv.*, vol. 49, no. 2, pp. 1–36, 2016.

[17] C. C. Aggarwal, "A Survey of Uncertain Data Clustering Algorithms," *Data Clust. Algorithms Appl.*, vol. 21, no. 5, pp. 455–480, 2013.

[18] K. Crockett, Z. Bandar, D. Mclean, and J. O'Shea, "On constructing a fuzzy inference framework using crisp decision trees," *Fuzzy Sets Syst.*, vol. 157, no. 21, pp. 2809–2832, 2006.

[19] TfGM, "Transport for Greater Manchester." [Online]. Available: https://www.tfgm.com/.

[20] E. H. Mamdani and S. Assilian, "An experiment in linguistic synthesis with a fuzzy logic controller," *Int. J. Man. Mach. Stud.*, vol. 7, no. 1, pp. 1–13, Jan. 1975.

[21] Weka, "Class J48," 2018. [Online]. Available: http://weka.sourceforge.net/doc.dev/weka/classifiers/trees/J48.html. [Accessed: 15-Nov-2018].