A Decision Support System for Home BP Measurements

Akshay Jain*, Mihail Popescu, James Keller
ECE Department
University of Missouri (UM)
*aj4g2@mail.missouri.edu

JL Belden, RJ Koopman, SJ Patil, SM Canfield; LM Steege**
UM School of Medicine
University of Wisconsin**

VA Shaffer, P Wegier, KD Valentine, A Hathaway
UM Health Sciences, Psychology

ABSTRACT
Wearable and non-wearable sensors are pervasive. However, the health implications of the data they provide is not always clear for the user. In this paper we present a Decision Support System (DSS) that assists a user of a Home Blood Pressure (HBP) monitor to decide timely consultation with a doctor. While HBP is more reliable than office readings, it is more variable due to factors such as food, exercise or error in recording measurements. Our DSS is based on fuzzy rules composed of linguistic summaries of the data. The rules are designed from the current US clinical guidelines and are tuned using an evolutionary algorithm. On a dataset of 40 patients monitored over 3 months, we obtained an interrater agreement of 0.97 between the physicians and DSS trained with their data, while the average agreement between these same physicians was 0.95.

Author Keywords
Home BP, DSS, linguistic summaries, fuzzy rules

ACM Classification Keywords
Information systems ~ Expert systems, Computing methodologies ~ Vagueness and fuzzy logic

INTRODUCTION
The amount and variety of wearable and non-wearable sensors continue to increase in everyday life which, we believe, enables the discovery of many early warnings of medical conditions. While these sensors provide a wealth of data from heart rate to blood pressure, they do not provide any insight into the medical knowledge associated with their patterns. Home Blood Pressure (HBP) measurements are shown to be better indicators of health risks as compared to measurements taken at clinics [1]. Therefore, the health risk associated with hypertension has led to an increased usage of electronic BP monitors in homes, generating large amounts of HBP data. Since the HBP is not always measured in ideal conditions, the data generated might be more variable and not easily interpretable by the user. We propose a Decision Support System (DSS) for HBP measurements that assesses the degree to which the blood pressure is uncontrolled, thus, allowing the user to identify when they might require a clinical intervention. The proposed DSS may also help save time during clinical encounters by providing text summaries of large amounts of data in BP display or by providing support in the clinical BP control workflow [2].

The Eighth Joint National Committee (JNC 8) [3] guidelines regarding hypertension suggests starting a pharmacologic treatment to decrease BP once it reaches beyond a clinically significant threshold. For example, the guidelines state that for people older than 60 years of age, treatment should be started when BP approaches or surpasses 150/90 mm Hg, while a threshold of 140/90 is used for those younger than 60. These guidelines however, do not address the situation when BP values are both inside and outside the target range. In [4] authors conducted a study to assess BP control. They monitored ambulatory BP around the clock and related it to sporadic HBP values. They concluded that, if 3 or more of the last 10 systolic HBP readings are greater than or equal to 135 mm Hg, then BP is not in control and an intervention may be required. In both studies, a crisp threshold is used to assess BP. However, in our pilot study from three family medicine physicians assessing patients' HBP data, we observed that there is some variation in the way these experts assessed control of HBP. This prompted questions regarding what data features experts utilize when assessing HBP measurements. Is it just the proportion of readings above a certain threshold? Is it the trend of the data? Do all experts consider the same control threshold?

In this work, we propose a DSS which helps in assessing HBP control. We make use of a fuzzy rule system (FRS) to mimic human process of rating a series of HBP data. The input to FRS consists of trend and linguistic summaries of the data which help in addressing the temporal nature of the data and provide clear explanations (i.e. "The BP is not controlled due to many days out of range in the last two weeks"). The fuzzy rules are designed by taking into account the guidelines in [3, 4]. They are tuned using an evolutionary approach to match the physicians' assessment of BP control on a dataset obtained from [5].

BACKGROUND
The input to the FRS system is a series of systolic HBP measurements obtained in a given period of time, two weeks in our case. Although diastolic BP has clinical significance,

PervasiveHealth '17, May 23–26, 2017, Barcelona, Spain

ACM ISBN 978-1-4503-6363-1/17/05…$15.00
https://doi.org/10.1145/3154862.3154891

most clinical guidelines for BP control [4] are only based on the systolic values, as in this study.

Linguistic Feature Extraction

Since a person might be irregular at recording their BP over a time period, the first step is to summarize the series of BP measurements into a set of features. Though numeric statistical features are possible, we chose linguistic features. Linguistic features are informative for explanation of DSS recommendations and provide a good data summary for the clinician who does not wish to analyze large numbers of BP data points during a patient's short visit [6]. The conversion of the numerical time series to a linguistic representation was performed using Linguistic Protoform Summaries (LPS). A simple LPS can be of the form: "*Q y's are P*" where *Q* and *P* are quantifiers and summarizers, respectively, and the *y's* are the objects to be summarized. In this work we used one summarizer, {*out of range*}, and five quantifiers, {*Almost none*, *Few*, *Some*, *Many*, *Almost all*}. For example, an LPS summarizing BP data for two weeks can be of the form: "*The BP was out of range for a few days in the last two weeks*". Summarizers and quantifiers are modelled by fuzzy sets defined over suitable domains, as shown in Figure 1 and 2. Each LPS is accompanied by a truth value, ranging from 0 to 1, which is the measure of validity of the summary. More details about LPS generation can be found in [7]. We also explore the trend of a series of BP measurements.

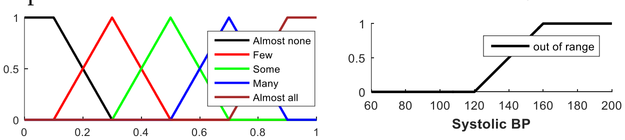


Figure 1 & 2: Quantifiers (left) and the *Out of Range* Summarizer (right) used to generate (LPS)

Fuzzy Rule Base System

A fuzzy rule system (FRS) has a set of rules based on expert input. Each fuzzy rule has a set of inputs (antecedents) and a set of outputs (consequents). A fuzzy rule which uses trend and LPS features as inputs to assess BP control may have rules of the form: *If BP is out of range for some days and trend is increasing, then BP is not in control.* Various rules will have different LPS quantifiers (instead of *some*) and different trend values (instead of *increasing*).

The presence of LPS in the fuzzy rules listed above makes our system different from a 'standard' FRS [8]. There, the degree to which an antecedent is satisfied is the membership value of the crisp input. However, this is not the case when using LPS as one of the antecedents of the fuzzy rules. In order to incorporate them, we treat the truth value of LPS as the membership degree which is directly fed into the fuzzy rule system. There are several types of fuzzy rule systems available in literature. We use the Mamdani – Assilion (MA) FRS framework [8] to implement our approach. A more detailed description of our LPS FRS will be presented in a subsequent paper.

DATASET

The dataset used in this work resulted from the study conducted in [5] where 43 intervention patients recorded their BP at home one or more times per day for about 3 months. For each patient, information was extracted from their clinic visit notes at different points in time and was annotated with pre-defined tags such as {*BP out of range*, *BP in range}*. These are cases where a clinician inspected the BP data and tagged it as *out of/in range*. In all there were 30 such events.

We extracted the systolic BP data for the previous 14 days from the dataset for all 30 events. For cases with multiple BP readings per day, we take the average for that day. The BP *in range* and *out of range* tags for a lot of these cases were considered unreliable by physicians in our team. Therefore, to obtain a new set of BP assessments, these cases were then presented to the 3 physicians in the form of line graphs displaying the systolic BP of each case over 2 weeks. The experts are part of our research team and have relevant experience of 21, 38 and 10 years, respectively.

For each case, the physicians were asked to rate the BP as *not in control* with a degree of confidence ranging from 0 to 10. These ratings were then linearly scaled in the range 0 to 1. We observed that even though the ratings for each expert were similar to one another for most of the examples, there was significant difference among them. This suggests that the experts do not use a set of well–defined rules while assessing BP control. Agreement between the ratings can be quantified by computing Interclass Correlation Coefficient (ICC) [9]. When comparing two or more raters, an ICC of 1 denotes a 100% agreement between the raters. ICC for the ratings provided by the experts are shown in Table 1. We see that the average ICC for our panel of 3 clinicians was 0.95.

Table 1: Inter Class Coefficient (ICC) for the ratings provided by 3

	ICC
Expert 1 – Expert 2	0.959
Expert 2 – Expert 1	0.936
Expert 3 – Expert 2	0.956
Average Expert ICC	**0.95**

EXPERIMENTS

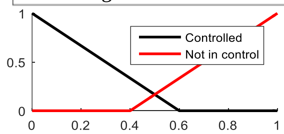
The diversity in the expert ratings presented above suggests a 'fuzzy' system to match the process of the experts in assessing a set of BP measurements. The experiments in this section are designed to further explore this notion.

Fuzzy Rules with just LPS

We start with LPS of the form: *The BP was out of range for Q days*, where, *out of range* is the summarizer and *Q* a quantifier. The Summarizer, *out of range*, assigns a degree at which a certain blood pressure value is not in range. A typical fuzzy set (based on the guidelines in [3]) that might represent *out of range* is shown in Figure 2 (i.e. a BP of 140 is *out of range* with a degree of 0.5). Quantifiers, as the name suggests, specify the applicability of Summarizers. For the LPS shown above, they are used to define how many of the measurements in a series of BP readings are out of range. Figure 1 shows the quantifiers {*Almost None*, *Few*, *Some*, *Many*, *Almost all*}, used in our system. Tables 2 presents the Fuzzy Rules composed of LPS. Both BP *controlled* and *not in control* are also fuzzy sets defined over suitable domains, as shown in Figure 3. The fuzzy rules used in this experiment

Table 2: FRS using only LPS for BP control assessment. The rules are of the form: If LPS is A then BP is B

Antecedent (A)	Consequent(B)
Out of range for *almost all* or *many* or *some* days	not in control
Out of range for few or almost none of the days	controlled



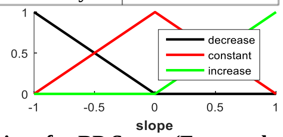


Figure 3 (left): Membership Functions for BP Status (Fuzzy rule consequent). Figure 4 (right): Membership Functions for Trend

are quite simplistic in nature. Basically, the FRS is used to combine the truth values of the five LPS and transform them to a degree in *not in control* BP. Nevertheless, we will see next, they perform well. The MA fuzzy inference engine described in Section 2 is used, with one change. The defuzzified output of the BP status variables shown in Figure 3, using centroid defuzzification is scaled between 0 and 1. The fuzzy set describing the summarizer, *out of range*, is a very important and medically relevant parameter of the system. Looking at the ratings provided by the experts and speaking with them, we concluded that the definition of *out of range* is subjective. Therefore, we used a data centric approach to define *out of range* BP and we employed a simple evolutionary algorithm (EA), which finds the definition that would best match the output of the FRS to the expert ratings. Note that the EA is carried out separately for each expert. Table 4 (Column 1) shows the ICC value of the fuzzy rule base system and the expert for which it was trained. We see that the agreement between the fuzzy rule base system and corresponding experts is more than that among the experts (Table 1). Also, the ICC for Expert 3 is very close to 1, signifying that this expert weights the quantity of measurements out of range very highly. The ICC value of not equal to 1, suggests presence of cases in which the LPS based system could not match the ratings of the expert. Some of these cases are analyzed in detail in the Section 5.

Fuzzy Rules with LPS + Trend

The fuzzy rules described above only make use of LPS when assessing a series of BP measurements. Even though this system is able to explain most of the ratings provided by the experts, there are some where this is not so. These cases might be explained by including some more features to our fuzzy rule base. Trend seems to be a potential feature, which appears in a lot of instances. To this end, we introduce trend to our fuzzy rule base, as shown in Table 3.

Trend is computed by linear regression of order 1. The output slope of the best fit line is passed through the membership functions shown in Figure 4 to get the degree of satisfaction in the fuzzy sets {*decrease, constant, increase*}.

We follow the procedure using an EA described in Experiment 1, to learn the fuzzy set representing the summarizer, *out of range*. However, given the small number of examples (only 30), we fix the definitions of fuzzy sets representing {*decrease, constant, increase*}. The rest of the details of the FRS are same as Experiment 1. Table 4 shows

Table 3: FRS e using LPS and trend for BP control assessment. The rules are of the form: If LPS is $A_1$and trend is A_2 then BP is B

Antecedents		**Consequent**
A_1	**A_2**	**B**
Out of range for almost all or many days		not in control
Out of range for some days	decrease	controlled
Out of range for some days	constant	not in control
Out of range for some days	increase	not in control
Out of range for few days	decrease	controlled
Out of range for few days	constant	controlled
Out of range for few days	increase	not in control
Out of range for almost none of the days		controlled

Table 4: ICC values for both systems with Experts 1, 2 and 3.

	LPS	LPS + Trend
Expert 1 - System	0.968	0.969
Expert 2 - System	0.966	0.972
Expert 3 - System	0.987	0.976
Average System ICC	**0.974**	**0.972**

the ICC values for the system with both the LPS and trend used in the rule base. We observe that, according to this metric, the introduction of trend did not change the system performance by much for any of the 3 experts. Nevertheless, the performance for Expert 2 increased a little, while that for Expert 3 deteriorated. In both cases, the average ICC between the system and the 3 experts was about 0.97.

CASE STUDIES

In this section we analyze the results obtained in the two experiments with the help of illustrative case studies. The expert ratings and the *not in control* degree by the two fuzzy rule base systems are presented in Table 5.

Case A (Figure 5a) is an interesting case where the trend is evident. There is a significant variation between ratings of the experts. Experts rated it as 0.1, 0.4 and 0 *not in control*, respectively. The impact of including trend in the fuzzy rule base is easy to observe for Expert 1 and 2. For the systems trained for both these experts, the upward trend increases the confidence in the class *not in control*. For Expert 1, the system seems to 'overshoot' from 0.02 to 0.15 while for Expert 2, the degree in *not in control* doesn't rise enough to match the rating. However, it matches much better compared to just using LPS. For Expert 3, the inclusion of trend doesn't have any impact here. Combining the results for all 3 experts, the inclusion of trend clearly helps to match ratings of the experts with the system for this case.

In Case B (Figure 5b), the FRS with trend has a negative impact on the system performance. We can see for Expert 1, the ratings matched better when using just LPS. The presence of downward trend decreases the system confidence in *not in control* class. For Expert 2, the trend

Table 5: For each expert, the *not in control* rating computed by both the fuzzy rules bases and the rating of the expert is shown

	Expert 1			Expert 2			Expert 3		
	Rules 1	Rules 2	User	Rules 1	Rules 2	User	Rules 1	Rules 2	User
A	0.02	0.15	0.1	0	0.28	0.4	0.02	0	0
B	0.35	0.05	0.3	0.05	0.05	0	0.35	0.05	0.2
C	0.53	0.5	0.2	0.68	0.5	0.3	0.77	0.5	0.8
D	0.63	0.68	0.2	0.72	0.68	0.5	0.65	0.7	0.8

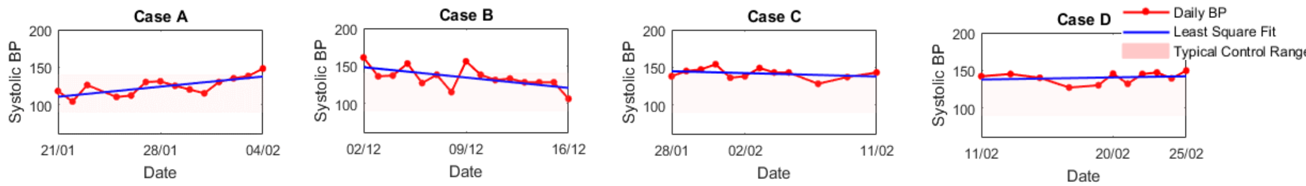


Figure 5(a,b,c,d): Systolic BP data used in case studies. The light colored patch shows the typical Systolic BP control range (90 to 140)

does not have any impact while for Expert 3, the system rating significantly undershoots the rating provided.

Case C (Figure 5c) is ambiguous in the sense that measurements bounce up and down from the typical control range and there is no evident trend in the data. This is also shown in the variation in the ratings provided by the experts as 0.2, 0.3 and 0.8. We see from Table 5 that neither of the two rule base systems provide a good match for any of the experts except the LPS based rules for Expert 3. This is consistent with results obtained until now that just LPS based system works well for Expert 3.

In Case D (Figure 5d) also, expert ratings have a significant variation among them (0.2, 0.5, 0.8), which again might be due to the fact that the measurements bounce around the control range boundary. This again leads to a poor performance of the system in terms of matching to the expert ratings. Neither of the two rule bases were able to match well with the ratings provided by the three experts.

DISCUSSION

With the help of the ICC metrics and the case studies we see that the Fuzzy Rule Base system has a potential to be a decision support system for HBP data. This system allowed us to gain important insights about the factors that the physicians might consider while assessing BP data presented as a time series. Also, we observed that there is no set of well–defined crisp rules to assess BP that are universally accepted. This is evident in Cases C and D, where the data points bounce above and below the typical BP control range.

The case studies show that not all physicians give equal weight to all the features. We see that our three experts tend to give a lot of importance to the amount of BP measurements which are out of range. However, Expert 2 gives more weight to trend in the data as compared to Experts 1 and 3 (See Case A). Also, the ratings provided by Expert 3, mirrored very well with just LPS based fuzzy rules, which shows that they give little or no weight to the trend. This is in line with our conversation with the experts about factors they consider while assessing BP control.

The inclusion of trend to our Fuzzy Rule Base was helpful in explaining some of the cases/experts, but not always. This is in accordance with what the experts had to say about which features they give importance to while rating Blood Pressure. This can also be observed in Case A and B. While in Case A, the inclusion of trend helped to better match the system rating to that of Expert 1 and 2, in Case B, the trend lead to a poorer matching degree for Expert 1.

CONCLUSION

We presented a DSS for HBP data based on fuzzy rules with linguistic summaries and trend as input. Our system achieved a 0.97 average agreement with 3 primary care physicians, comparable to the average agreement between physicians themselves, 0.95. We also pointed out the subjective nature of BP assessment task. Future work will investigate adding more features and learning the parameters for all input features using EA. A rigorous evaluation of our DSS requires more data in order to separate the training and testing set, which we intend to do next. Nevertheless, this pilot study allowed us to gain important insights about assessment of HBP data.

ACKNOWLEDGMENTS

This work was supported by AHRQ grant 1R01HS023328 and by NIH-NLM grant 1R01LM012221-01A1.

REFERENCES

1. T. G. Pickering, *et al*, "Call to action on use and reimbursement for home blood pressure monitoring," *Hypertension*, vol. 52, pp. 10-29, 2008.
2. A. Cuenca, "Improving Blood Pressure Control With Strategic Workflows," *Family practice management*, vol. 23, pp. 23-28, 2015.
3. P. A. James, *et al.*, "2014 evidence-based guideline for the management of high blood pressure in adults: report from the panel members appointed to the Eighth Joint National Committee (JNC 8)," *Jama*, vol. 311, pp. 507-520, 2014.
4. J. E. Sharman *et al.*, "Pragmatic method using blood pressure diaries to assess blood pressure control," *The Annals of Family Medicine*, vol. 14, pp. 63-69, 2016.
5. R. J. Koopman, *et al.*, "Implementing home blood glucose and blood pressure telemonitoring in primary care practices for patients with diabetes: lessons learned," *Telemedicine and e-Health*, vol. 20, pp. 253-260, 2014.
6. R. J. Koopman. (2015). *Optimizing Display of Blood Pressure Data To Support Clinical Decision*. Available: https://healthit.ahrq.gov/ahrq-funded-projects/optimizing-display-blood-pressure-data-support-clinical-decision
7. A. Jain and J. M. Keller, "On the computation of semantically ordered truth values of linguistic protoform summaries," in *Fuzzy Systems (FUZZ-IEEE), 2015 IEEE International Conference on*, 2015, pp. 1-8.
8. J. M. Keller, D. Liu, and D. Fogel, *Fundamentals of Computational Intelligence: Neural Networks, Fuzzy Systems, and Evolutionary Computation,*: Wiley/IEEE Press, 2016.
9. K. A. Hallgren, "Computing inter-rater reliability for observational data: an overview and tutorial," *Tutorials in quantitative methods for psychology*, vol. 8, p. 23, 2012.