


Contrast Set Mining for Actionable Insights into Associations Between Sleep and Glucose in a Normoglycemic Population

Hoang Huyen Nhung¹ and Zilu Liang^{1,2} 

¹Ubiquitous and Personal Computing Lab, Kyoto University of Advanced Science (KUAS), 621-8555 Kyoto, Japan

²Institute of Industrial Science, The University of Tokyo, 113-8654 Tokyo, Japan

Keywords: Data Mining, Personal Informatics, Ubiquitous Computing, Contrast Set Mining.


Abstract: Prior studies have suggested potential associations between poor sleep and glucose dysregulation among diabetic patients. However, little is known about the relationship between sleep and glucose regulation in healthy populations. In this study, we proposed a data mining pipeline based on contrast set mining to identify significant associations between sleep and glucose in a dataset collected from a normoglycemic population in free-living environments. Unlike traditional correlation analysis, our approach does not assume a linear relationship between sleep and glucose and can potentially discover associations when a pair of metrics fall within certain value ranges. The data mining result highlights the total sleep time as an important sleep metric associated with glucose regulation the next day, which is characterised by rules with high lift and confidence. Furthermore, the result suggests that having a higher time ratio in normal glucose range was associated with better sleep continuity at night. These results may provide insights that people can immediately act on for better sleep and better glucose control. Future research may leverage the proposed data mining protocol to develop healthy behaviour recommender systems.

1 INTRODUCTION

Sleep is an essential part of human daily life. There is consensus that adults need 7 or more hours of sleep (Watson et al., 2015). Sleep deprivation was reported to account for a wide spectrum of health problems, including glycaemic disturbances and the development of diabetes (Jiawei et al., 2017; Lou et al., 2015; Xiao et al., 2014; Zuraikat et al., 2020). Modern lifestyle has led to significant changes in human sleep patterns, such as delayed sleep onset and decreased total sleep time. Furthermore, modern abundance has also caused significant shifts in people's eating habits and has consequently led to a surge in chronic metabolic diseases such as diabetes, which is characterised by glucose dysregulation. A few prior studies have demonstrated a strong connection between sleep quality and glucose homeostasis, especially in people with health conditions (Cauter et al., 1991 Sep; Gottlieb et al., 2005; Kothari et al., 2021; Lou et al., 2015; Wang et al., 2017). Nonetheless, those studies were limited by the methods available for data collection and analysis,

and the healthy population is often underrepresented or completely excluded.

With a variety of consumer wearable technologies being developed, researchers are now able to perform longitudinal measurements of sleep and glucose in a more naturalistic environment. Multidimensional sleep structure can be measured using a Fitbit wristband, and 24-hour interstitial glucose can be monitored using a CGM system such as the FreeStyle Libre. Data collected with modern wearable technologies make it possible to examine the reciprocal relationships between sleep and glucose at a finer resolution. Different from prior studies that primarily consider the daily aggregates of sleep and glucose data, our study introduces a circadian perspective. To be more specific, we are interested in how sleep metrics in the previous night associates with glucose during the day and how the glucose level in the daytime associates with the subsequent night-time sleep. To the best of our knowledge, this is the first study investigating the associations between sleep and glucose using ecologically valid and high-resolution data. The contribution of this study is as follow:

 <https://orcid.org/0000-0002-2328-5016>

- We designed a data mining pipeline based on the contrast set mining algorithm STUCCO (Bay & Pazzani, 2001) instead of traditional statistical methods. Different from correlation analysis or linear regression, our method allows the identification of associations that only manifest when the metrics are within certain value ranges.
- The data mining pipeline generates interesting rules and hypotheses that may help inform the design of future studies to deepen our understanding of the relationships between sleep and glucose regulation.

2 RELATED WORKS

Existing studies have found a link between sleep quality and glucose regulation in both diabetic and healthy populations. However, the relationship between sleep and interstitial glucose is complex, which makes it difficult to quantify the connection. Multiple approaches have been proposed to examine this correspondence within and between persons.

Factors such as total sleeping time, sleep structure, time going to bed, age, and eating habits are widely known to have a strong influence on blood glucose levels (Frank et al., 1995; Reutrakul et al., 2013; Tasali et al., 2008). In a study (Cauter et al., 1991), eight normal men were supervised for a total of 53 hours (8h of night sleep, 28h of non-sleep period, and 8h of daytime sleep). Although glucose was infused into the body at a constant rate, plasma glucose levels went through large fluctuations throughout the study. Interestingly, during sleep deprivation, glucose levels rise gradually to reach the maximum at roughly the same time as during regular sleep, then decrease to daytime levels.

Many studies repeatedly report the association of sleep debt, sleep curtailment, and reduced of sleep hygiene contribute to the increase of insulin sensitivity (SI). Even one night of partial sleep deprivation may decrease glucose tolerance and insulin sensitivity significantly (Donga et al., 2010). (Tasali et al., 2008) paid attention to slow-wave sleep (SWS) when his team tried to reduce SWS proportion but sustain total sleep time. His findings suggest that low levels of SWS, as occurs in the elderly and obese patients, can decrease insulin sensitivity by 25%, reaching the reported value for high-risk of diabetes. In the case of sleep debt, the product of insulin sensitivity and acute insulin response to intravenous glucose (AIRg) was decreased by an average of about 40% as compared to the fully rested state, indicating a high risk of diabetes (Xiang et al., 2006).

Despite evidence had been found to prove the connection, they were reported as weak and did not reach statistical significance. In the study focusing on women only, no impact of progressive sleep curtailment over 4 nights was reported on measures of glucose tolerance and SI (Bosy-Westphal et al., 2008). Similarly, (Zielinski et al., 2008) assessed the impact of 8 weeks sleep curtailment on glucose tolerance in self-reported older long sleepers (≥ 8.5 h/night), compared to a control group; through OGTT, the authors observed no effect of sleep restriction on glucose tolerance. Sleep hygiene is more challenging to measure as it is more difficult to define than sleep duration.

Just as sleep affects glucose levels, glucose levels may as well impact sleep quality. For example, one study found that people with prediabetes have a higher rate of suffering from poor sleep than people with normal glucose (Iyegha et al., 2019). Diabetic patients display shorter sleep duration and worse sleep quality, demonstrated by both self-report and objective measures (Yoda et al., 2015).

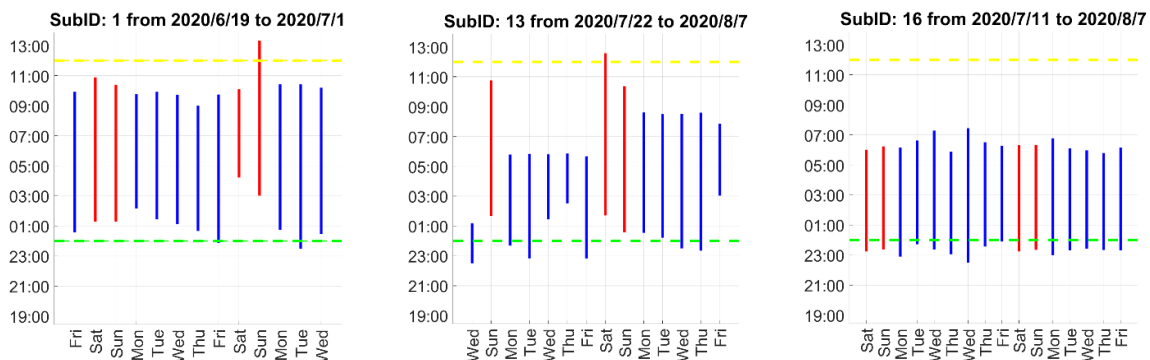


Figure 1: sleep hours of some of our subject indicated by vertical lines (blue lines for weekdays and red lines for weekends).

3 PROPOSED DATA MINING PIPELINE

3.1 Dataset

We used a public dataset which contains sleep and glucose readings recorded between June and August 2020 (Bertrand et al., 2021). In total there are 228 days of data collected from 12 healthy subjects who had no glucose dysregulation. Half of the subjects were female, and the average age was 32.7 years. The sleep patterns varied a lot across participants, as illustrated in Figure 1.

Currently, several technologies are available to quantitatively assess all-day interstitial glucose concentration. In this study, the glucose readings were recorded using the FreeStyle Libre 2 system which is a coin-size continuous glucose monitoring (CGM) device attached at the back of the upper arms. Sleep data were measured with the accelerometer and photoplethysmography (PPG) embedded in the Fitbit Charge 3 worn on the non-dominant wrists. Both the FreeStyle Libre and the Fitbit devices were proven to be able to generate reasonable valid measurements in free-living conditions (Li & Bao, 2018; Liang & Chapa-Martell, 2019).

Table 1: Features constructed from sleep and glucose data.

	Feature	Denotation
Sleep	Total Sleep Time (hours)	TST
	Wake After Sleep Onset	WASO
	Number of wakes ≥ 5 minutes	awake5minCnt
	Sleep Efficiency(%)	SE
	Deep sleep/TST (%)	deepRatio
	Rem sleep/TST (%)	remRatio
Glucose	Mean (mg/dL)	mean
	Maximum (mg/dL)	max
	Minimum (mg/dL)	min
	Mean glucose level outside range (mg/dL)	mge
	Mean glucose level inside range (mg/dL)	mgn
	Standard Deviation (mg/dL)	sd
	Coefficient of variation (%)	cv
	Time spent in range (%)	tir
	Low glucose index	LBGI
	High glucose index	HBGI
	J-index (mg2/dL2)	j_index

3.2 Data Preprocessing

The data preprocessing protocol includes segmenting the CGM data, deriving sleep and CGM features, categorizing numeric features, synchronizing sleep and CGM features, and removing missing data.

The dataset is not in the appropriate format to which the rule induction algorithms may be applied. Therefore, the next step in the pre-processing protocol is to categorise the numeric features into intervals before feeding them into a mining algorithm. Discretization helps improve the performance of contrast set mining algorithms and improve the interpretability of the results. The features were discretized using the methods shown in Figure 2. We mainly focused on features that have well-defined clinical cut-offs. For sleep features, we used medical cut-offs for young adults (18-25 years) reported by (Ohayon et al., 2017). The National Sleep Foundation concluded that sleep quality was a matter of sleep continuity and sleep architecture. The optimal sleep architecture for good sleep quality for adults was agreed to be <5% stage 1 sleep, <81% stage 2 sleep, 16-20% slow wave sleep (SWS), and 21-30% rapid eye movement (REM) sleep. However, stage 1 and stage 2 were combined as light sleep in Fitbit device, therefore the cut-off limits for this parameter remain unknown. Consequently, we did not include light sleep ratio in the sleep feature set.

It is important to properly handle the missing data and ensure the timestamp matches correctly between

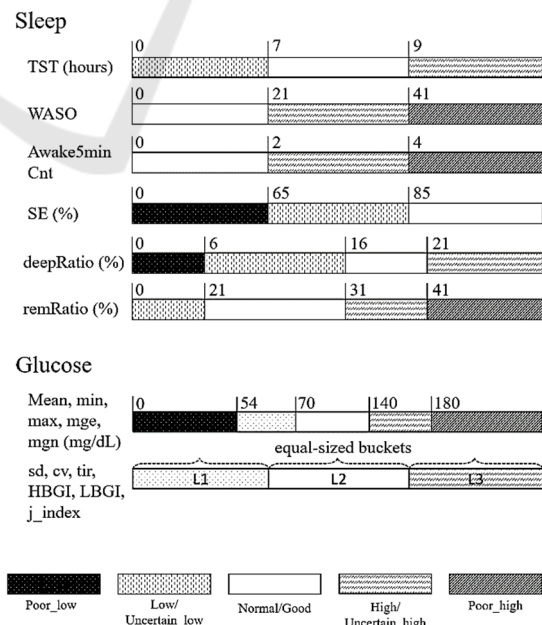


Figure 2: Discretization of sleep and glucose level metrics.

the sleep features and the glucose features. Rows that did not meet the required conditions were removed. Criteria for removing data are nights that lack glucose data, non-consecutive nights (isolated nights without any data recorded the day before or after), features that are dominated by one component (appear >90% number of nights). After pre-processing, the cleaned dataset contains 74 days of data from 8 subjects.

3.3 Contrast Set Mining Algorithm

Contrast set mining is a data mining technique that helps identify contrast patterns between two groups.

The output of contrast set mining are logical rules with 1 or n antecedents on the left-hand-side and the consequent on the right-hand-side.

A typical contrast set mining is the Search and Testing for Understandable Consistent Contrast (STUCCO) algorithm developed in (Bay & Pazzani, 2001). The STUCCO algorithm follows a basic pipeline of identifying a basic idea, controlling the error, filtering the results, evaluating the results, then drawing conclusions based on the problem at hand. For this study, we expect to generate some rules as:

$$TST=low \cap deepRatio=poor \rightarrow Glucose_tir=low$$

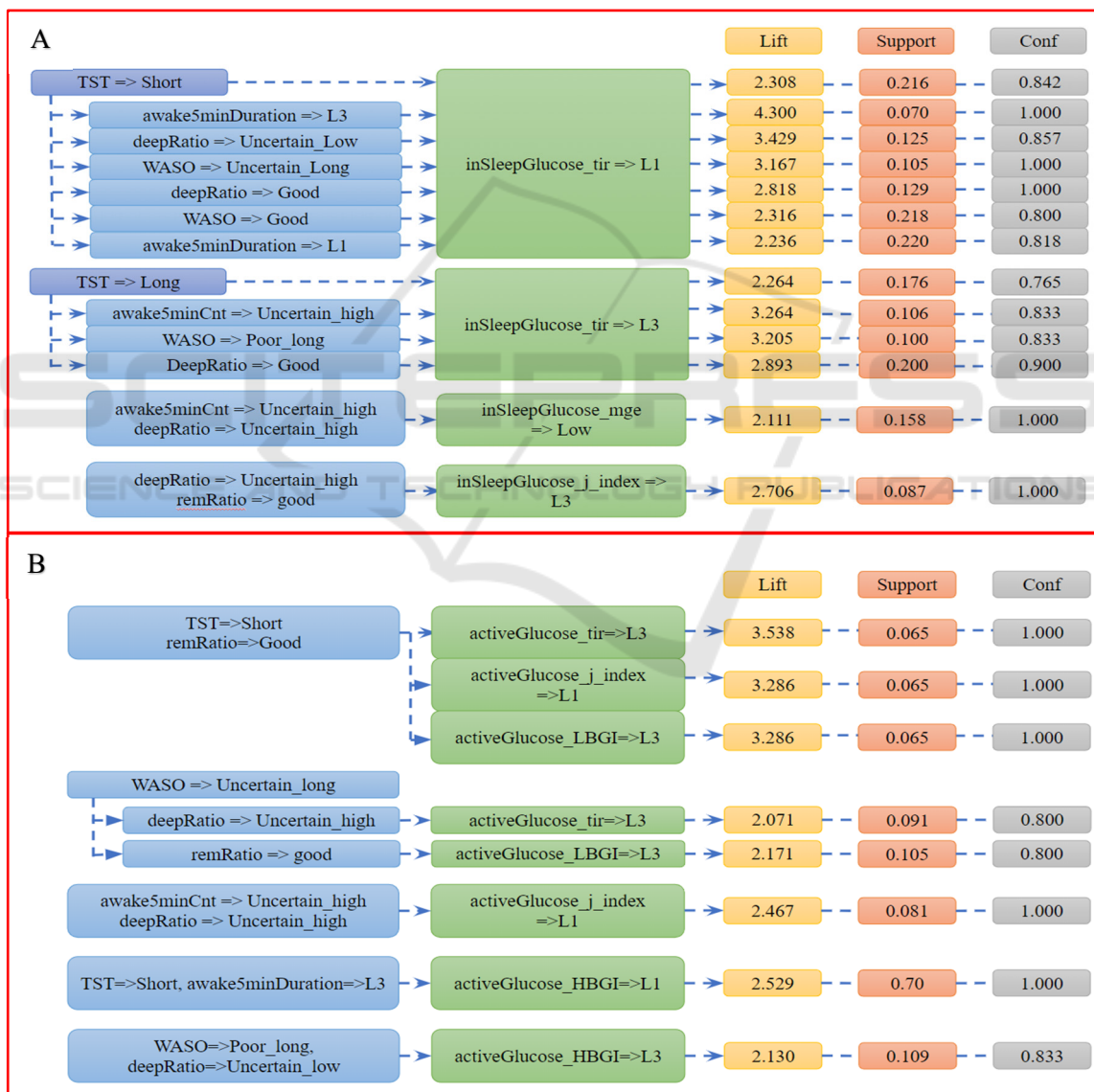


Figure 3: Contrast sets were distributed into two groups. Group A reveals the relationship between sleep metrics and interstitial glucose during nocturnal sleep. Group B reveals the relationship between sleep metrics and interstitial glucose the following day.

Given a dataset D , let $Y = \{Y_1, Y_2, \dots, Y_n\}$ be a set of consequences and $X = \{X_1, X_2, \dots, X_n\}$ be a set of antecedents that consist of 1 or more attributes. *Support*, *confidence*, and *lift* are some of the measures widely used to assess the quality of the generated rules. Support for $X_1 \rightarrow Y_1$ is the percentage of example in D containing $X_1 \cup Y_1$. Confidence is the ratio of example X_1 in total X that contributes to Y_1 . Lift is the threshold defined by the user to find the contrast set. Long rules with too many features in the antecedents are often hard to interpret and could be redundant. Therefore, we limit the length of the rules to 2 features. We consider a contrast set as validate if it meets the requirements: support ≥ 0.02 , confident ≥ 0.75 , and lift ≥ 2 . After contrast sets generation, rules which share the same attribute were grouped together (Figure 3 & 4) for better interpretation.

4 RESULTS

4.1 Case 1: Antecedence = Sleep of Day N, Consequence = Glucose of Day N

The contrast set mining results show connection between the sleep quality at night and the glucose patterns the next day, as well as the sleep quality at night and the concurrent glucose characteristics during sleep. The identified rules are presented in Figure 3. TST is a major factor that associates to affect the TIR of glucose both in sleep and during active periods. It is shown that the time the glucose levels stayed within the normal range was shorter when the participants slept less than 7 hours at night. In contrast, people who had long sleep periods (> 9 hours) spent a longer time in normal glucose level

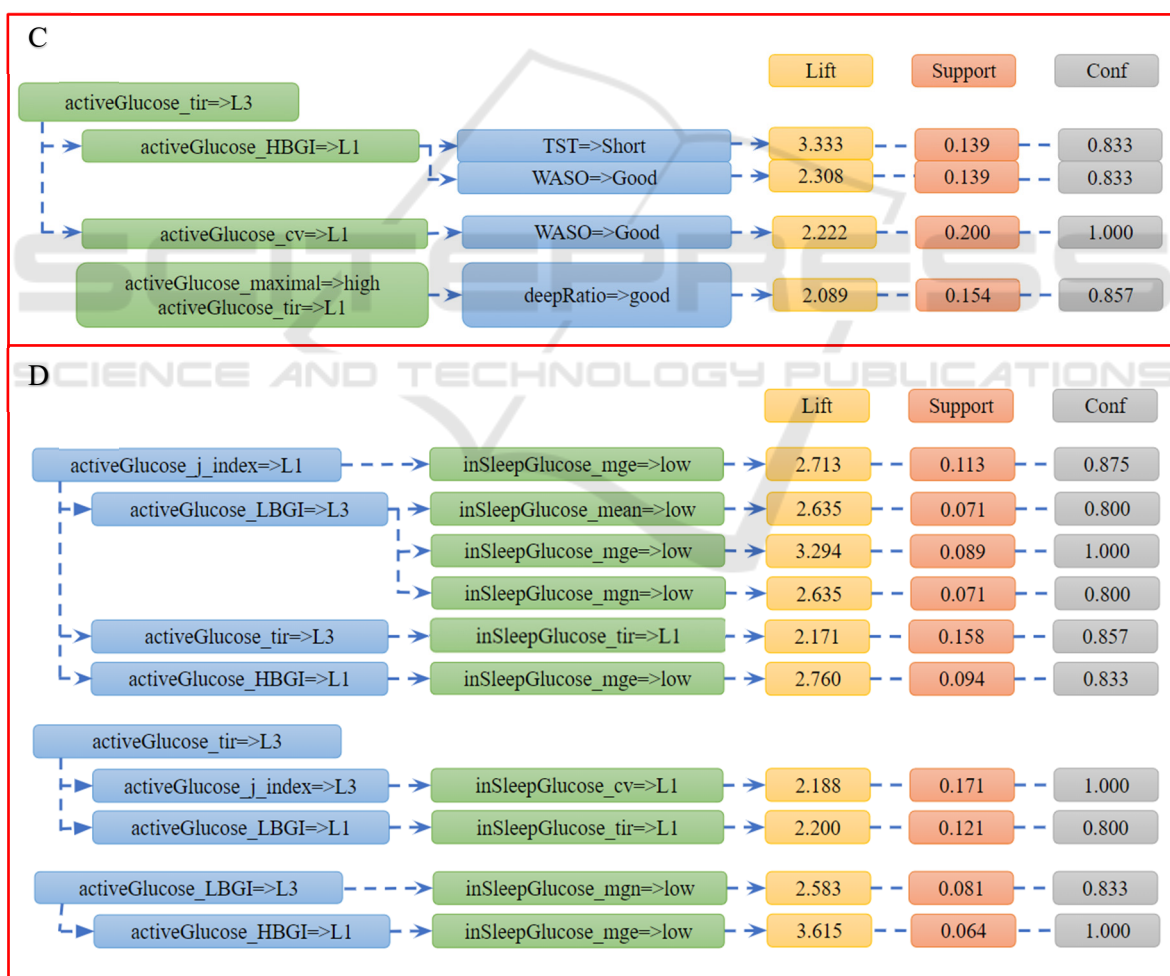


Figure 4: The contrast sets show how interstitial glucose during the day connected with the following night sleep (group C) and interstitial glucose level during sleep (group D).

range throughout the night. Under the confounding effect of other sleep metrics, TST was also associated with the TIR of glucose level during active time the next day. For example, short TST combined with a good ratio of REM sleep were associated with higher TIR the next day.

In addition to sleep duration, sleep continuity and sleep structure as characterised by the ratio of sleep stages, were also found to be associated with glucose regulation during sleep and during active time the next day. Long awake5minDuration strengthens the association between short TST and low TIR during sleep, with the lift increased from 2.31 to 4.30. However, the same combination was associated with a low level of HBGI the next day. Long WASO, together with the confounding effect of sleep structure, were associated with higher LBG1 and HBGI the next day, indicating impaired glucose regulation.

The associations between sleep structure (i.e., the ratio of deep sleep and REM sleep) and glucose regulation were complex and were often confounded by sleep duration and sleep continuity.

4.2 Case 2: Antecedent = Glucose of Day N, Consequent = Sleep of Day N+1

There is a vicious cycle between sleep and glucose levels when one affects the other and the impact repeats days and nights. While finding the rules between glucose levels at day and sleep metrics at nights, we found that a long time in range for glucose level helps to keep sleep WASO within good range (less than 20 minutes of awakening per night). People with long *tir* glucose are highly likely to sleep less than 7 hours the following night. In addition, long time in range link with good WASO while short time in range link with good deep sleep ratio.

It is noticeable that the relationship between active glucose and in-sleep glucose are more complex and extremely difficult to interpret. Interestingly, all the right-hand-side components belong to the “low” section of the features and the rules are mainly associated with glucose *mge* and *mgn*. From figure 4, there is a contrast set with 2 completely opposite components: if active glucose time in range is high, then in sleep glucose time in range is low. Several studies have suggested multiple pathways in the possible connection between sleep quality and the fluctuation of glycaemic variability. But whether daytime glucose impacts in-sleep glucose remains a research gap that needs further examination.

5 DISCUSSIONS

In this study, we analysed the potential connections between sleep and glucose in the series of consecutive days. Our finding revealed that there are bi-direction relationships between TST and glucose TIR. Total sleep time is the feature that dominated most of the contrast sets. Short sleep time was associated with the fluctuation of in-sleep glucose, whereas long sleep time was associated with more stable glucose regulation. On the flip side, an association was found between glucose fluctuation and sleep continuity, which is consistent with previous findings (Griggs et al., 2022). Our result indicated that long TIR and low HBGI correlate to shorter WASO. This echoes findings in prior studies that long WASO occurred on days with high J index, high HBGI, and less time in hypoglycemia (Griggs et al., 2020).

This study has several limitations that should be addressed in future studies. First, the size of the dataset is limited. Future studies should collect data from a larger cohort and potentially cover a wide spectrum of demographic characteristics. Furthermore, not only nocturnal sleep, but also napping may be linked to glucose regulation, as suggested in previous lab-based studies (Kothari et al., 2021). Since napping is a common habit of the young population and people in tropical areas, this line of research is likely to generate new insights. Another promising aspect for future examination is the impact within and between personal variations in sleep patterns. With a large dataset, the impact of interpersonal differences should be studied when each person has different lifestyles and circadian rhythms. Finally, with the abundance of rules found by STUCCO, it is necessary for a post-mining method to select quality rules.

6 CONCLUSIONS

We have designed a data mining pipeline featuring the contrast set mining algorithm STUCCO to identify reciprocal associations between sleep and glucose levels. The finding highlights the total sleep time as an important sleep metric associated with glucose regulation the next day, which is characterised by rules with high lift and confidence. Reversely, associations were also found between the glucose fluctuation in wake time and the continuity of the subsequent sleep at night. Compared to sleep->glucose relation, glucose has a weaker association with nocturnal sleep as the lift of the identified rules

was lower. These findings added to the existing knowledge looking at the glucose profile in the normoglycemic population and helped generate actionable insights for the holistic management of sleep and metabolic health. With a better method to remove abundant rules and interpret information, this contrast set mining algorithm can be applied to a recommendation system for adjusting human behaviour.

ACKNOWLEDGEMENTS

This study was supported by JSPS KAKENHI Grant Number 21K17670.

REFERENCES

- Bay, S. D., & Pazzani, M. J. (2001). Detecting Group Differences: Mining Contrast Sets. *Data Mining and Knowledge Discovery*, 5, 213–246. <https://doi.org/10.1023/A:1011429418057>
- Bertrand, L., Cleyet-Marrel, N., & Liang, Z. (2021). Recognizing Eating Activities in Free-Living Environment Using Consumer Wearable Devices. *Eng. Proc.*, 6, 58.
- Bosy-Westphal, A., Hinrichs, S., Jauch-Chara, K., Hitze, B., Later, W., Wilms, B., . . . Muller, M. J. (2008). Influence of partial sleep deprivation on energy balance and insulin sensitivity in healthy women. *Obesity facts*, 1(5), 266–273. <https://doi.org/10.1159/000158874>
- Cauter, E. V., Blackman, J. D., Roland, D., Spire, J. P., Refetoff, S., & Polonsky, K. S. (1991). Modulation of glucose regulation and insulin secretion by circadian rhythmicity and sleep. *J Clin Invest.*, 88(3), 934-942. <https://doi.org/10.1172/JCI115396>
- Donga, E., Dijk, M. v., Dijk, J. G. v., Biermasz, N. R., Lammers, G.-J., Kralingen, K. W. v., . . . Romijn, J. A. (2010). A single night of partial sleep deprivation induces insulin resistance in multiple metabolic pathways in healthy subjects. *The Journal of clinical endocrinology and metabolism*, 95(6), 2963–2968.
- Frank, S. A., Roland, D. C., Sturis, J., Byrne, M. M., Refetoff, S., Polonsky, K. S., & Van Cauter, E. (1995). Effects of aging on glucose regulation during wakefulness and sleep. *The American journal of physiology*, 269(6 Pt1)(E1006–E1016). <https://doi.org/10.1152/ajpendo.1995.269.6.E1006>
- Gottlieb, D. J., Punjabi, N. M., Newman, A. B., Resnick, H. E., Redline, S., Baldwin, C. M., & Nieto, F. J. (2005). Association of sleep time with diabetes mellitus and impaired glucose tolerance. *Archives of internal medicine*, 165(8), 863–867. <https://doi.org/10.1001/archinte.165.8.863>
- Griggs, S., Barbato, E., Hernandez, E., Gupta, D., Margevicius, S., Grey, M., & Jr., R. L. H. (2022). Glucose and unstructured physical activity coupling during sleep and wake in young adults with type 1 diabetes. *Scientific reports*, 12(1)(5790). <https://doi.org/10.1038/s41598-022-09728-2>
- Griggs, S., Redeker, N. S., Jeon, S., & Grey, M. (2020). Daily variations in sleep and glucose in adolescents with type 1 diabetes. *Pediatric diabetes*, 21(8), 1493–1501. <https://doi.org/10.1111/pedi.13117>
- Iyegha, I. D., Chieh, A. Y., Bryant, B. M., & Li, L. (2019). Associations between poor sleep and glucose intolerance in prediabetes. *Psychoneuroendocrinology*, 110(104444). <https://doi.org/10.1016/j.psyneuen.2019.104444>
- Jiawei, Y., Xiaoling, J., Zhilei, S., Shuzhen, L., Hao, H., Peiyun, L., . . . Liegang, L. (2017). Relationship of Sleep Duration With All-Cause Mortality and Cardiovascular Events: A Systematic Review and Dose-Response Meta-Analysis of Prospective Cohort Studies. *Journal of the American Heart Association*, 6(9)(e005947). <https://doi.org/10.1161/JAHA.117.005947>
- Kothari, V., Cardona, Z., Chirakalwasan, N., Anothaisintawee, T., & Reutrakul, S. (2021). Sleep interventions and glucose metabolism: systematic review and meta-analysis. *Sleep medicine*, 78, 24-35. <https://doi.org/10.1016/j.sleep.2020.11.035>
- Li, M., & Bao, Y. (2018). Methods for Interpreting Continuous Glucose Monitoring Graphs. In W. Jia (Ed.), *Continuous Glucose Monitoring*. https://doi.org/10.1007/978-981-10-7074-7_5
- Liang, Z. (2022). Mining associations between glycemic variability in awake-time and in-sleep among non-diabetic adults. *Front Med Technol*. <https://doi.org/10.3389/fmedt.2022.1026830>
- Liang, Z., & Chapa-Martell, M. A. (2018). Validity of consumer activity wristbands and wearable EEG for measuring overall sleep parameters and sleep structure in free-living conditions. *Journal of Healthcare Informatics Research*, 1-27.
- Liang, Z., & Chapa-Martell, M. A. (2019). Accuracy of Fitbit wristbands in measuring sleep stage transitions and the effect of user-specific factors. *JMIR Mhealth Uhealth*, 7(6), e13384. <https://doi.org/10.2196/13384>
- Lou, P., Zhang, P., Zhang, L., Chen, P., Chang, G., Zhang, N., . . . Qiao, C. (2015). Effects of sleep duration and sleep quality on prevalence of type 2 diabetes mellitus: A 5-year follow-up study in China. *Diabetes research and clinical practice*, 109 (1), 178–184. <https://doi.org/10.1016/j.diabres.2015.04.012>
- Ohayon, M., Wickwire, E. M., Hirshkowitz, M., & al., e. (2017). National Sleep Foundation's sleep quality recommendations: first report. *Sleep Health*, 3(1), 6-19.
- Reutrakul, S., Hood, M. M., Crowley, S. J., Morgan, M. K., Teodori, M., Knutson, K. L., & Cauter, E. V. (2013). Chronotype is independently associated with glycemic control in type 2 diabetes. *Diabetes care*, 36(9), 2523–2529. <https://doi.org/10.2337/dc12-2697>
- Tasali, E., Leproult, R., Ehrmann, D. A., & Cauter, E. V. (2008). Slow-wave sleep and the risk of type 2 diabetes in humans. *Proceedings of the National Academy of Sciences of the United States of America*,

- Wang, H., Leng, J., Li, W., Wang, L., Zhang, C., Liu, H., . . . Yang, X. (2017). Sleep duration and quality, and risk of gestational diabetes mellitus in pregnant Chinese women. *Diabetic medicine : a journal of the British Diabetic Association*, *34*(1), 44–50. <https://doi.org/10.1111/dme.13155>
- Watson, N. F., Badr, M. S., Belenky, G., Bliwise, D. L., Buxton, O. M., Buysse, D., . . . Tasali, E. (2015). Recommended Amount of Sleep for a Healthy Adult: A Joint Consensus Statement of the American Academy of Sleep Medicine and Sleep Research Society. *Sleep*, *38*(6):843-4.
- Xiang, A. H., Peters, R. K., Kjos, S. L., Marroquin, A., Goico, J., Ochoa, C., . . . Buchanan, T. A. (2006). Effect of pioglitazone on pancreatic beta-cell function and diabetes risk in Hispanic women with prior gestational diabetes. *Diabetes*, *55*(2), 517–522. <https://doi.org/10.2337/diabetes.55.02.06.db05-1066>
- Xiao, Q., Keadle, S. K., Hollenbeck, A. R., & Matthews, C. E. (2014). Sleep duration and total and cause-specific mortality in a large US cohort: interrelationships with physical activity, sedentary behavior, and body mass index. *American journal of epidemiology*, *180*(10),997–1006.
- Yoda, K., Inaba, M., Hamamoto, K., Yoda, M., Tsuda, A., Mori, K., . . . Yamada, S. (2015). Association between Poor Glycemic Control, Impaired Sleep Quality, and Increased Arterial Thickening in Type 2 Diabetic Patients. *PLoS One*, *10*(4)(e0122521). <https://doi.org/10.1371/journal.pone.0122521>
- Zielinski, M. R., Kline, C. E., Kripke, D. F., Bogan, R. K., & Youngstedt, S. D. (2008). No effect of 8-week time in bed restriction on glucose tolerance in older long sleepers. *Journal of sleep research*, *17*(4), 412–419. <https://doi.org/10.1111/j.1365-2869.2008.00673.x>
- Zuraikat, F. M., Makarem, N., Redline, S., Aggarwal, B., Jelic, S., & St-Onge, M. P. (2020). Sleep Regularity and Cardiometabolic Health: Is Variability in Sleep Patterns a Risk Factor for Excess Adiposity and Glycemic Dysregulation?. *Current diabetes reports* , *20* (8), 38. <https://doi.org/10.1007/s11892-020-01324-w>