A Probabilistic Approach to Extract Qualitative Knowledge for Early Prediction of Gestational Diabetes

Athresh Karanam* Alexander L. Hayes* Harsha Kokel

David M. Haas Predrag Radivojac Sriraam Natarajan

“Risk of Gestational Diabetes *increases* as Body Mass Index *increases*.”

"Risk of Gestational Diabetes *increases* as Body Mass Index *increases*."

$$\mathrm{BMI} \underset{\sim}{\overset{M}{+}} \mathrm{GestDiab}$$

“Risk of Gestational Diabetes *increases* as Body Mass Index *increases*.”

$$\text{BMI} \overset{M}{\prec} + \text{GestDiab}$$

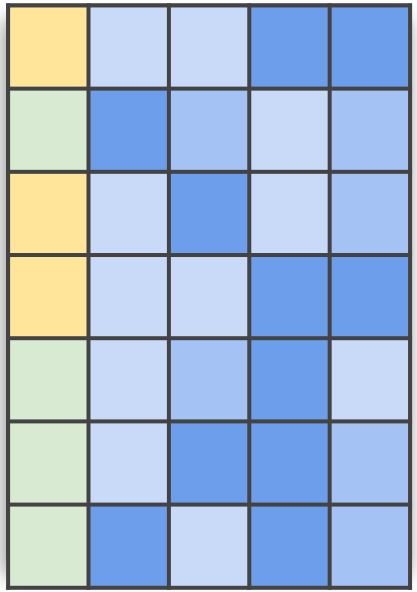
Human **interpretable**

Aligns with how people think about risk

Concisely expresses a trend over all variable-and-value combinations

Refine or repair models when data is noisy or sparse

y, x_1, x_2, x_3, x_4



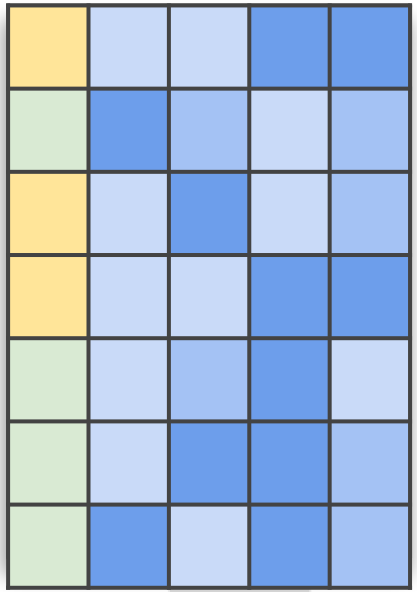
$$x_1 \overset{M+}{\prec} y$$

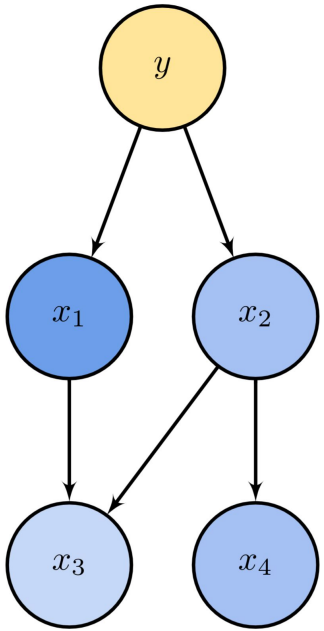
$$x_2 \overset{M+}{\prec} y$$

$$x_3 \overset{M-}{\prec} x_2$$

$$x_4 \overset{M-}{\prec} x_2$$

y, x_1, x_2, x_3, x_4





$P(Y, \boldsymbol{X})$

$x_1 \overset{M+}{\prec} y$

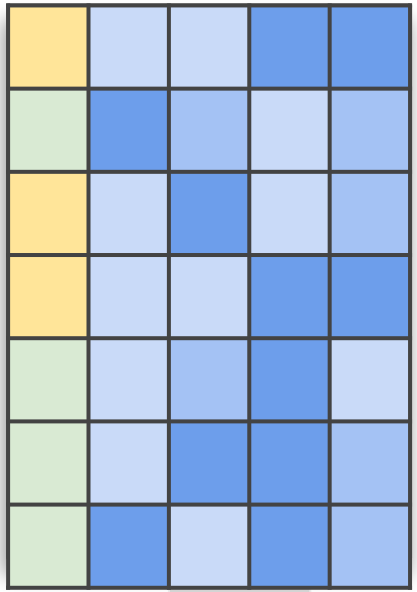
$x_2 \overset{M+}{\prec} y$

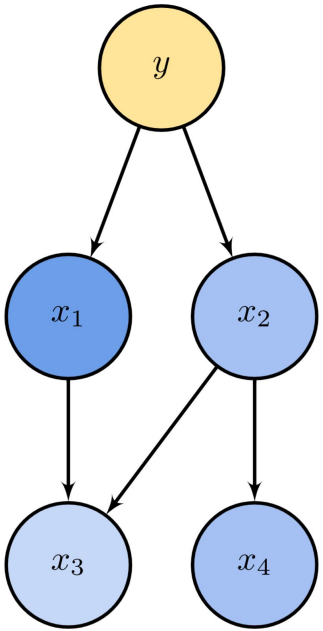
$x_3 \overset{M-}{\prec} x_2$

$x_4 \overset{M-}{\prec} x_2$

Altendorf et al. (2005) UAI, Yang and Natarajan (2013) MLKDD

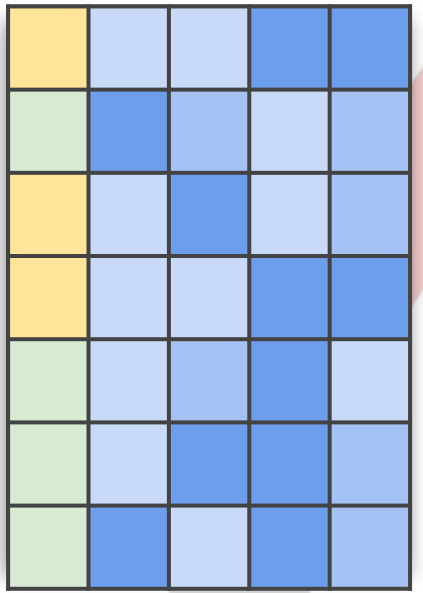
y, x_1, x_2, x_3, x_4



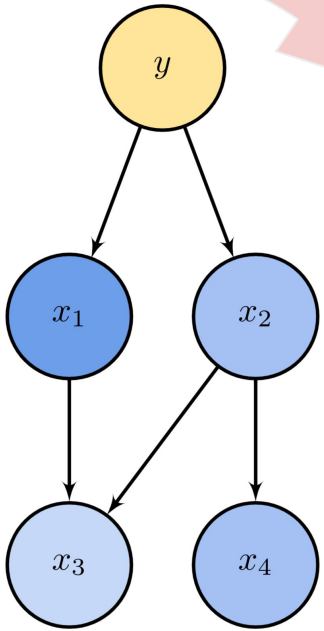


$P(Y, \boldsymbol{X})$

$$x_1 \overset{M+}{\prec} y$$
$$x_2 \overset{M+}{\prec} y$$
$$x_3 \overset{M-}{\prec} x_2$$
$$x_4 \overset{M-}{\prec} x_2$$



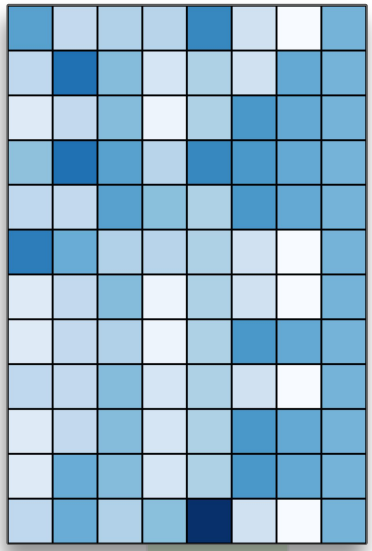
y, x_1, x_2, x_3, x_4



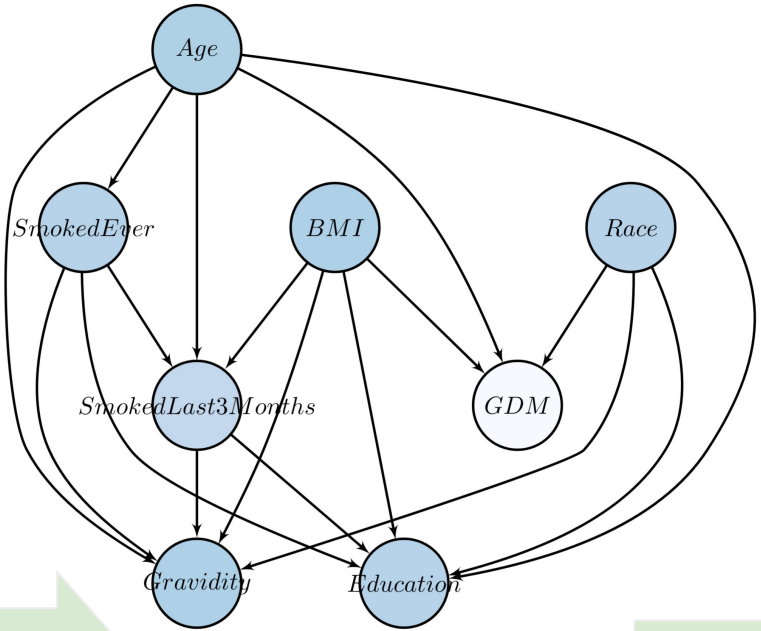
$P(Y, \boldsymbol{X})$

$$x_1 \overset{M+}{\prec} y$$
$$x_2 \overset{M+}{\prec} y$$
$$x_3 \overset{M-}{\prec} x_2$$
$$x_4 \overset{M-}{\prec} x_2$$

$GDM, Age, BMI, ...$



PC Algorithm



$P(Y, \boldsymbol{X})$

QuaKE

$BMI^{M^+}_{\preceq} GDM$
$Age^{M^+}_{\preceq} GDM$
$Race^{M^+}_{\preceq} GDM$
$Education^{M^+}_{\preceq} GDM$
$Gravidity^{M^+}_{\preceq} GDM$
$SmokedLast3Months^{M^+}_{\preceq} GDM$
$SmokedEver^{M^+}_{\preceq} GDM$
$Age, BMI^{S^+}_{\preceq} GDM$
$Age, SmokedLast3Months^{S^+}_{\preceq} GDM$
$BMI, SmokedEver^{S^+}_{\preceq} GDM$
$Education, SmokedLast3Months^{S^+}_{\preceq} GDM$
$BMI, Gravidity^{S^+}_{\preceq} GDM$
$BMI, SmokedLast3Months^{S^+}_{\preceq} GDM$
$Age, SmokedEver^{S^+}_{\preceq} GDM$
$BMI, Education^{S^+}_{\preceq} GDM$
$Education, SmokedEver^{S^+}_{\preceq} GDM$
$Age, Education^{S^-}_{\preceq} GDM$
$BMI, SmokedLast3Months^{S^-}_{\preceq} GDM$
$Age, SmokedEver^{S^-}_{\preceq} GDM$
$BMI, Gravidity^{S^-}_{\preceq} GDM$
$Gravidity, SmokedEver^{S^-}_{\preceq} GDM$
$Education, SmokedEver^{S^-}_{\preceq} GDM$
$Age, Gravidity^{S^-}_{\preceq} GDM$

Does QuaKE help uncover rules that align with prior knowledge?

Qualitative **K**nowledge **E**xtraction	"Data Alone" Baseline
Average 5-fold Precision with Expert: 0.923	*Average 5-fold Precision with Expert*: 0.636

Can QuaKE help uncover QI statements when prior knowledge is uncertain?

	Prior Knowledge	QuaKE
$\mathrm{BMI} \overset{M+}{\prec} \mathrm{GestDiab}$	✔	✔
$\mathrm{Education,Smoked3months} \overset{S+}{\prec} \mathrm{GestDiab}$	?	✔

A Probabilistic Approach to Extract Qualitative Knowledge for Early Prediction of Gestational Diabetes

Athresh Karanam* Alexander L. Hayes* Harsha Kokel

David M. Haas Predrag Radivojac Sriraam Natarajan

https://starling.utdallas.edu/papers/QuaKE/