

BAYESIAN NETWORK STRUCTURE FOR PREDICTING TWO-YEARS SURVIVAL IN PATIENTS DIAGNOSED WITH NON-SMALL CELL LUNG CANCER (#158)

OBJECTIVES

Develop a Bayesian network structure to predict two-year survival by combining expert knowledge with a rigorous structure learning algorithm.

MATERIALS & METHODS

We retrospectively collected data of 545 non-small cell lung cancer (NSCLC) patients diagnosed between 2010 to 2013 and eligible for (chemo) radiotherapy treatment at Maastrro Clinic, Maastricht, The Netherlands.

Age, Gender, WHO performance status, TNM stage, tumor load, FEV, smoking status and chemotherapy type were the considered variables to predict the two-year survival status.

Continuous variables were categorized with cutoff at 25th and 75th percentile or expert's recommendations & literature. Missing values were imputed but patients with missing survival status and variables with above 25% missingness were excluded.

The structure learning process was bootstrapped (B=400) with the hill-climbing algorithm to identify the causal relationships (arcs) between the variables. Experts' knowledge was employed to restrict the algorithm from forming arcs in clinically implausible directions (blacklist), like age having a causal influence on gender. Arc strength was evaluated as the rate of occurrence over all the bootstrap runs, and only arcs with an occurrence rate above 70% were included in the final structure.

The resulting structure was assessed by computing the area under the curve (AUC) using a repeated (r=50) 5-fold cross-validation technique and generating calibration plots.

DISCLOSURE

No conflicts of interest to declare.

INTRODUCTION

Lung cancer is among the leading cause of cancer morbidity and mortality worldwide [1].

These statistics actuate more knowledge and understanding of the patient condition and disease characteristics for better patient management and treatment outcomes.

However, the available information is more than humans can process, which calls for support from Predictive models such as Bayesian networks (BN) that can intricately represent causal relations between these variables [2, 3].

RESULTS 1

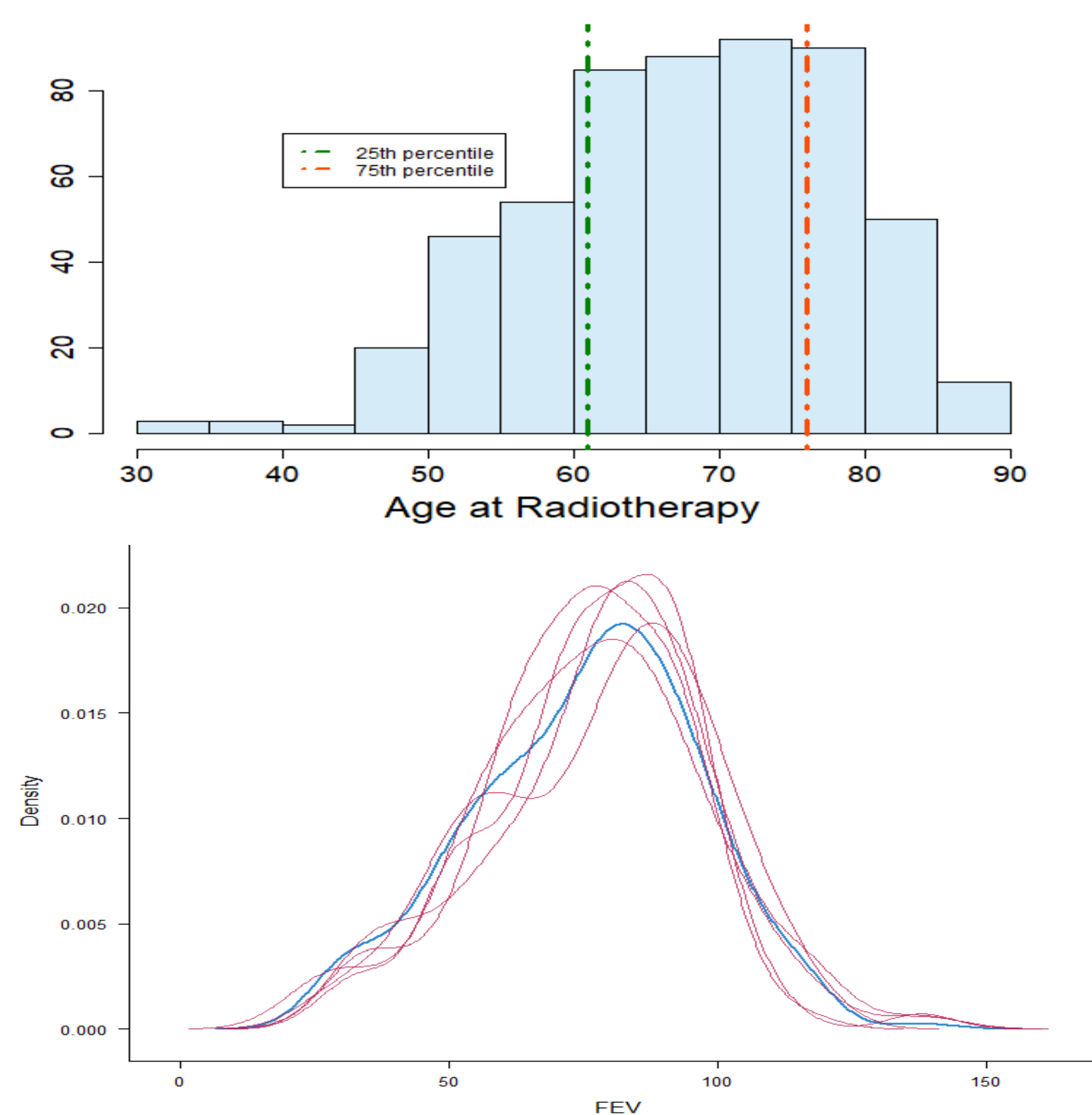


Figure 3: Histogram of age with cutoff values for discretization and Kernel density estimates for the observed data (blue) and the imputed data (red) for forced expiratory volume (FEV) to assess imputations.

CONTACT

biche.osong@maastro.nl
www.linkedin.com/in/bich

RESULTS 2

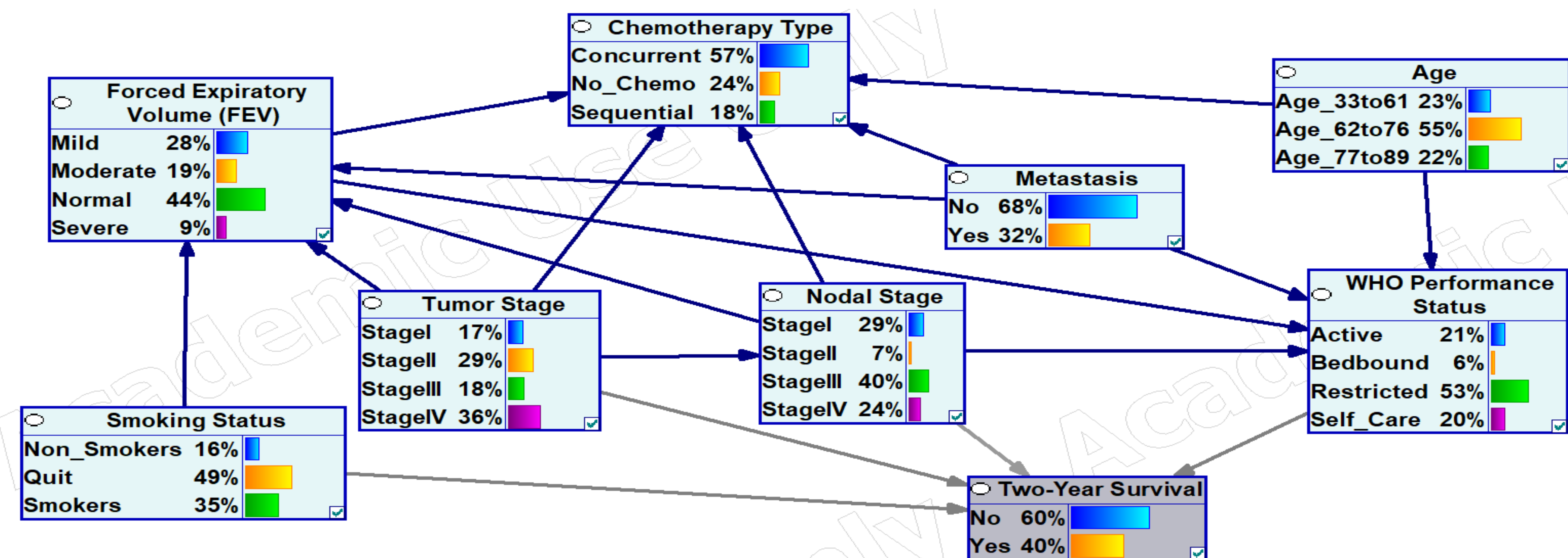


Figure 1: Resulting Bayesian network structure to predict two-year survival from the expert-algorithm method. Arrows indicate the direction of the causal-effect relationships. Grey arrows indicate a direct parental link to the outcome of interest. Chemo = Chemotherapy

Conclusion

A combination of expert knowledge with a rigorous structure learning algorithm produce a clinically plausible structure with optimal performance.

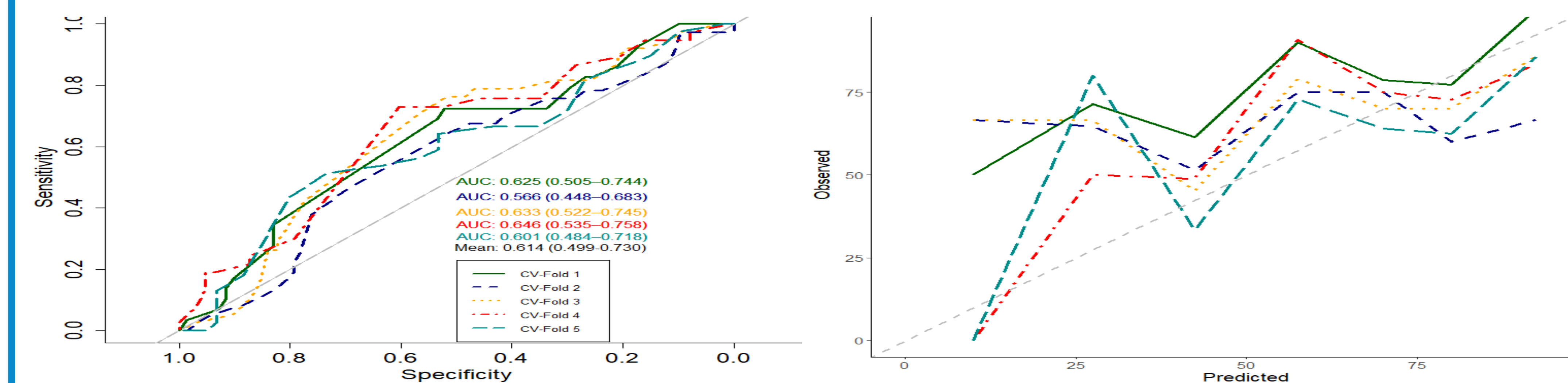


Figure 2: The area under the curve (AUC) and calibration plot of the Bayesian network structure. The colored lines represent the performance of each fold, and the gray line represents poor AUC and ideal calibration, respectively.

REFERENCES

- [1] Torre L.A, et al. Lung cancer statistics. *Lung cancer and personalized medicine*, pages 1–19, 2016.
- [2] Crealy Louise. Algorithm performs better than doctors at predicting mental health outcomes, 2018.
- [3] Jose M Pena. Finding consensus bayesian network structures. *Journal of Artificial Intelligence Research*.