

LUNG CANCER STAGE CLASSIFICATION USING RANDOM FOREST AND ARTIFICIAL NEURAL NETWORK

by

**Stanley T. Dalagan, Ella Joyce S. Paragas,
Aileen Joy V. Ramos, Clarissa Jewel B. Alota**

Central Luzon State University

Presented by:

Ella Joyce S. Paragas

LUNG CANCER

Lung carcinoma



Leading cause of cancer deaths for men and women



A malignant tumor characterized by uncontrolled cell growth in the tissues of the lungs

LUNG CANCER STAGE (Dependent Variable)

Stage
I

- tumor that has not yet spread to any lymph nodes

Stage II

- categorized by the tumor size

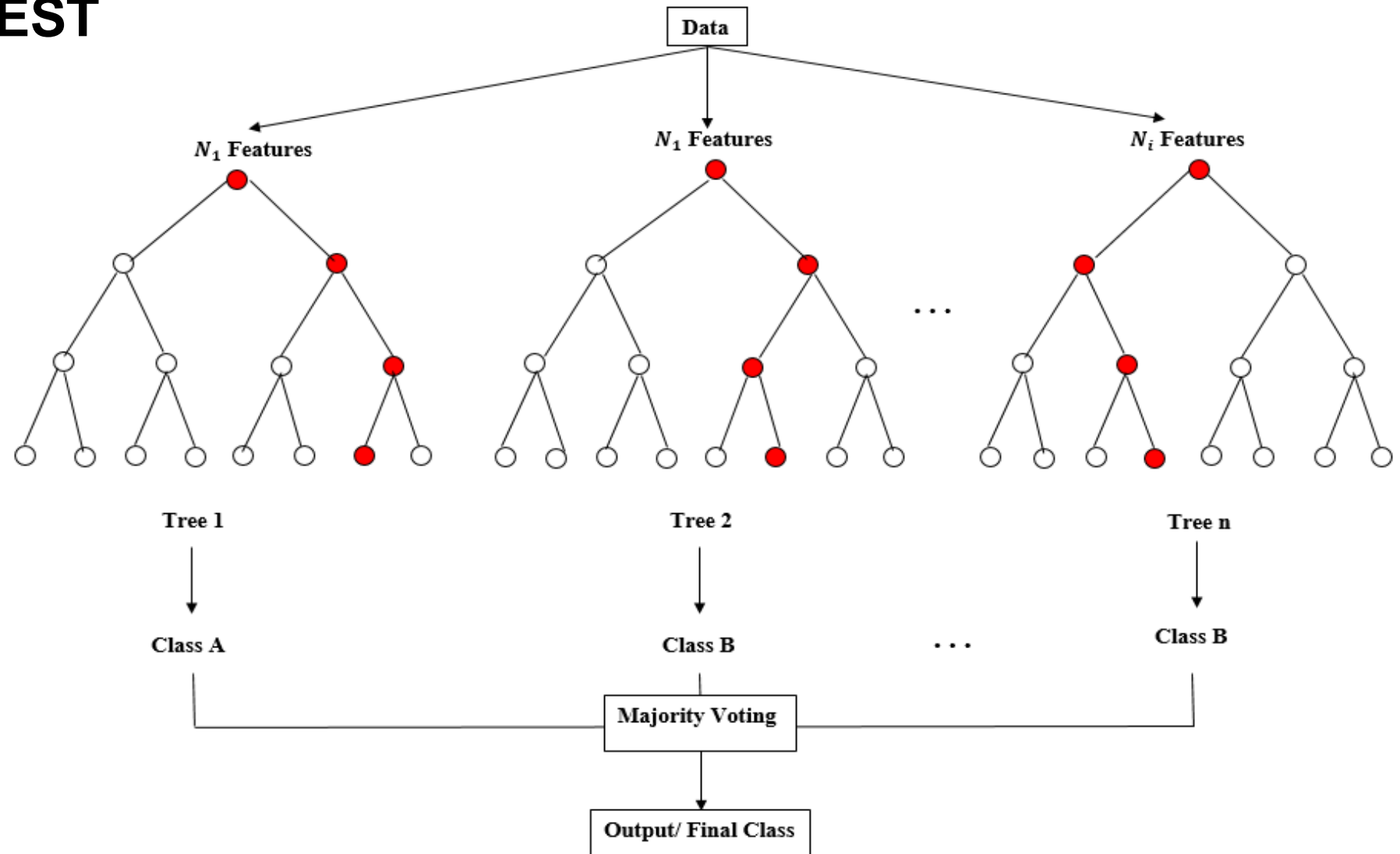
Stage
III

- based on the size of the tumor and which lymph nodes the cancer has spread to

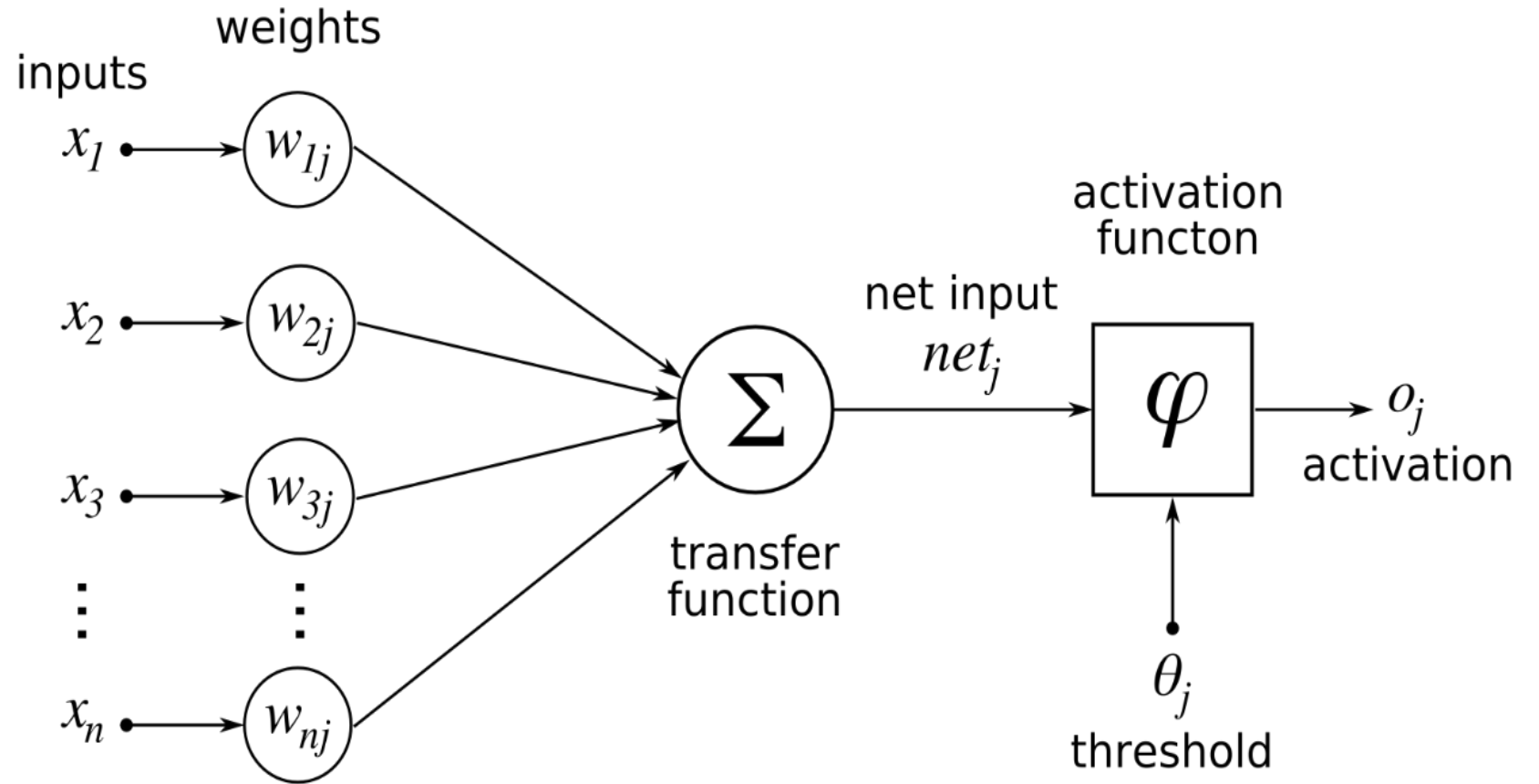
Stage
IV

- cancer has spread within the chest and outside the chest

RANDOM FOREST STRUCTURE



ARCHITECTURE OF ARTIFICIAL NEURAL NETWORK



DATA DESCRIPTION

National Cancer
Institute's Genomic
Data Commons Portal

4,086 observations
with 25 features

383 cases with
9 features

DATA DESCRIPTION

Sex

Number of years
smoke/smoking

Year of Birth

Primary diagnosis

Age at diagnosis

Site of origin

Average number of
cigarettes per day

Vital status

RESULTS

Male (56.7%)

1930 - 1949 (72%)

60-79 (79.3%)

Average of 3 sticks
per day

Smoked for 36 years on the
average

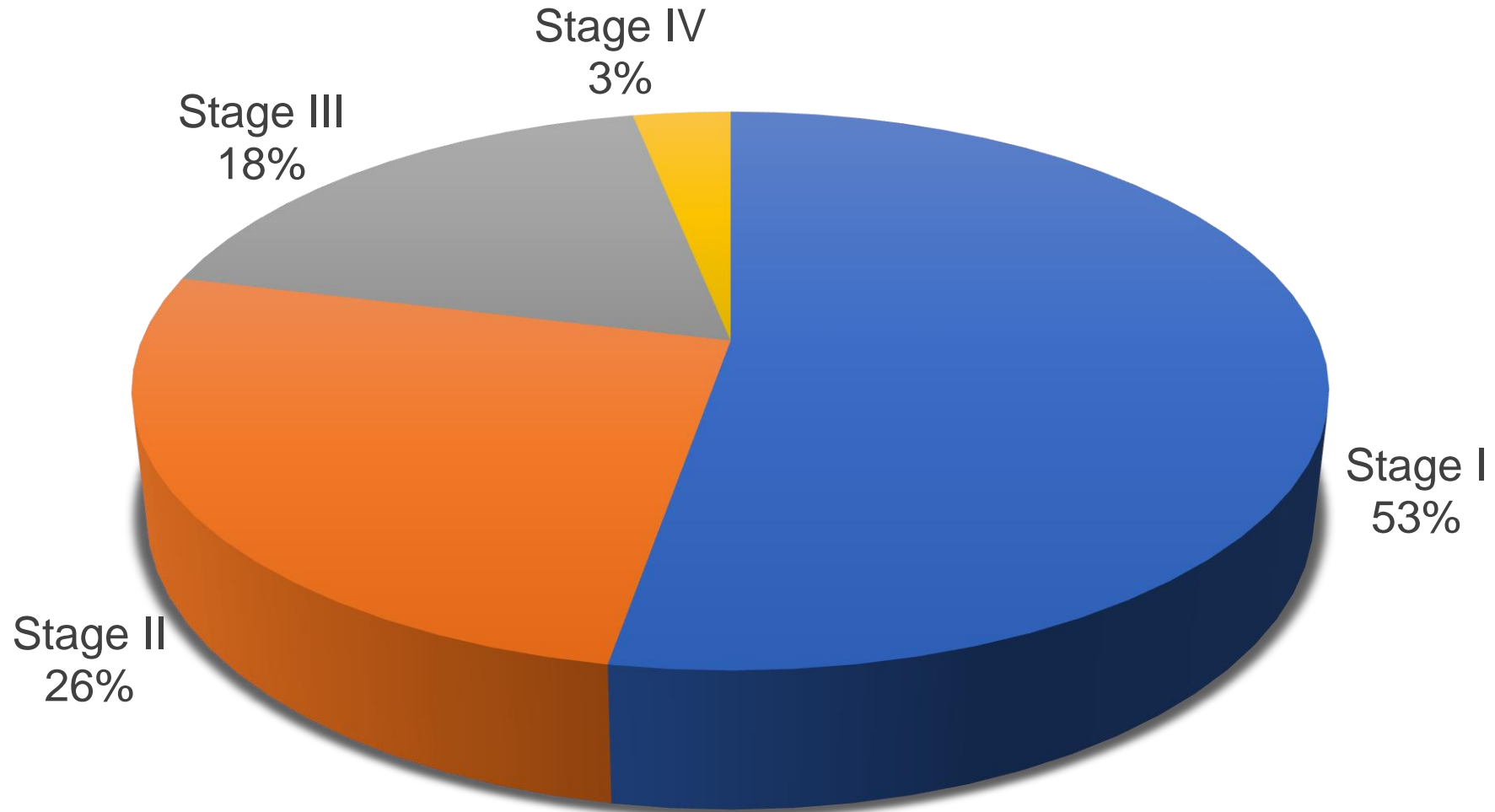
Squamous Cell Carcinoma(52.2%),
Adenocarcinoma (39.7%)

Upper lobe (53%)

Alive (61.6%)

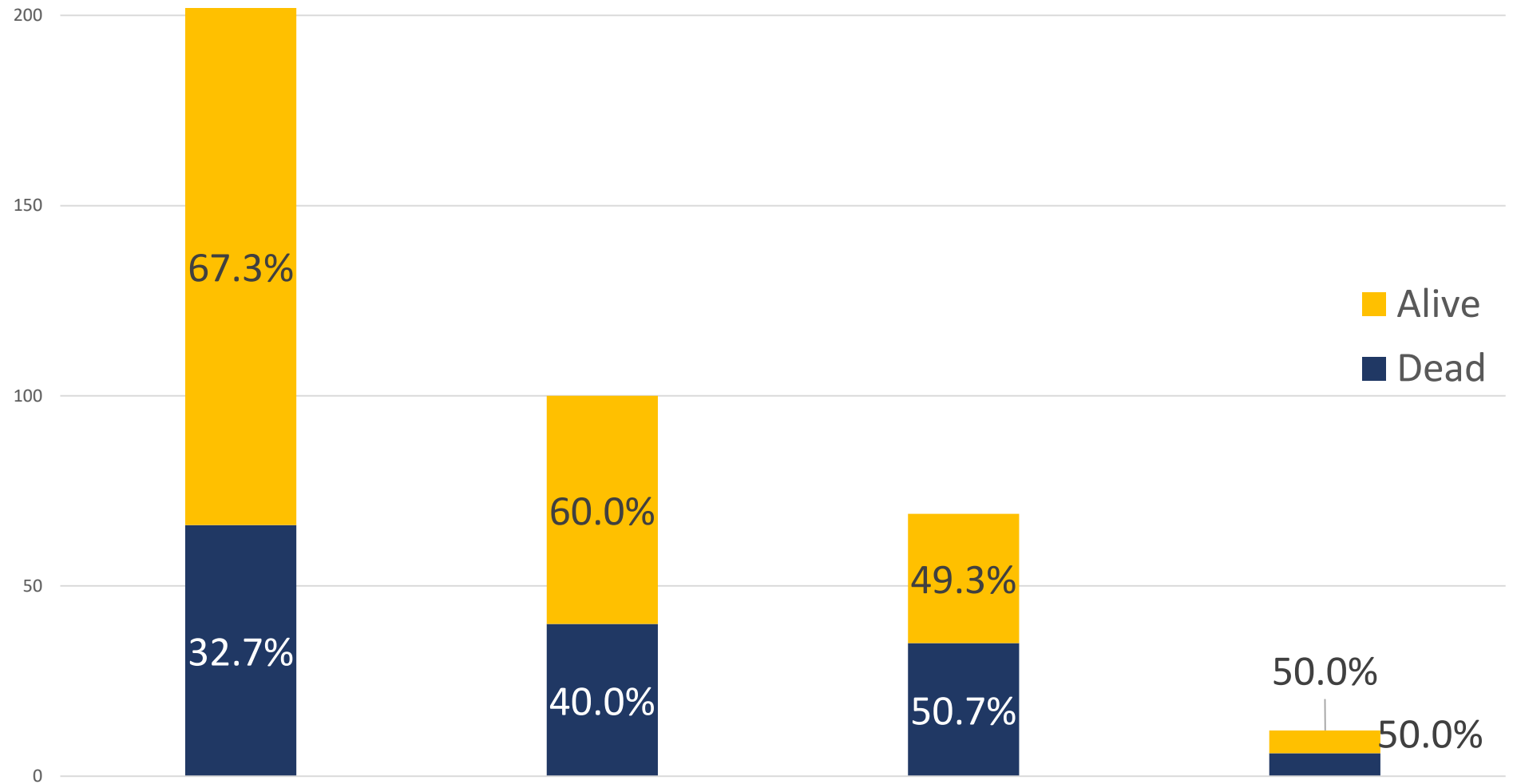
RESULTS

Lung Cancer Stages



RESULTS

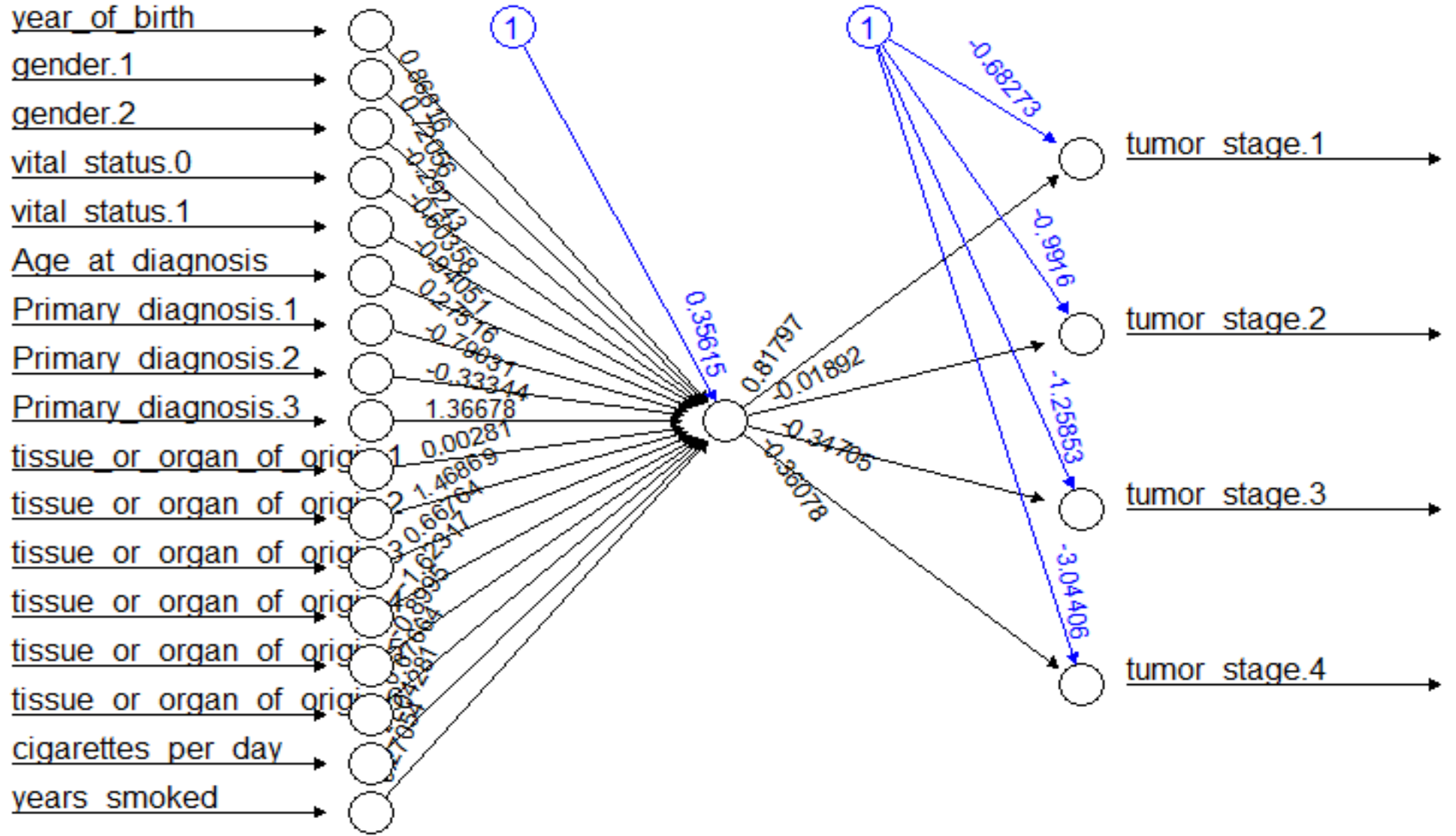
Frequency of tumor stage and percentage of vital status



RANDOM FOREST

FACTORS	MEAN GINI DECREASE	NUMBER OF TIMES USED IN THE MODEL
Age at diagnosis	39.23	2,031
Cigarettes smoked per day	33.83	1,887
Years smoked	32.08	1,816
Year of birth	30.60	1,792
Tissue/organ of origin	15.18	871
Primary diagnosis	9.73	655
Gender	7.66	529
Vital status	6.78	533

RESULTS



RESULTS

MODEL	ACCURACY		KAPPA	
	Training	Evaluation	Training	Evaluation
Random Forest (100 trees)	1	0.49	1	0.006
Neural Network (17-1-4)	0.7669	0.75	0.3783	0.3333

Thank you.