**Supplementary Materials to “A Data Mining Approach to Investigate Food Groups related to Incidence of Bladder Cancer in the BLadder cancer Epidemiology and Nutritional Determinants International Study”**

**Supplementary Table 1** Included individual food items in the BLEND Dataset

**Supplementary Figure 1** Receiver Operating Characteristic Curves for Each Cross-validation Run

**Supplementary Figure 2** Importance Values of Input Variables after C5.0 in the BLEND Dataset

**Supplementary Table 1** Included individual food items in the BLEND Dataset

|  |  |  |
| --- | --- | --- |
| **Food Items** | **Food Items** | **Food Items** |
| **Milk and milk products** | **Eggs and egg products** | **Meat and meat products** |
| Milk in miscellaneous | Eggs in miscellaneous | *Meat in miscellaneous* |
| *Liquid milk* | Egg dishes | *Beef*  |
| Liquid milk in miscellaneous | Omelette | Beef in miscellaneous |
| Milk >4% fat |  | Beef tenderloin |
| Milk 3-4% fat |  | Beef Striploin |
| Milk 1-3% fat |  | *Veal* |
| Milk < 1% fat |  | *Pork* |
| *Processed milk* |  | Pork in miscellaneous |
| Chocolate-flavoured milk |  | Pork chump |
| Acidophilus milk |  | Pork Loin |
| *Cream* |  | *Mutton/Lamb* |
| *Yoghurt* |  | *Mammals* |
| Yoghurt in miscellaneous |  | Horse |
| Yoghurt >3% |  | *Poultry* |
| Yoghurt <1% |  | Chicken |
| Cheese |  | Duck |
| Cheese in miscellaneous |  | *Liver* |
| Soft cheese |  | *Kidney* |
| Processed cheese |  | *Other Offal* |
| Whey cheese |  | *Preserved meat* |
| *Imitation milk* |  | Preserved meat in miscellaneous |
| Imitation milk in miscellaneous |  | Ham |
| Non-dairy coffee creamer |  | Bacon |
| *Milk beverage powder* |  |  |
| Drinking chocolate powder |  |  |
| *Ices* |  |  |
| Dairy ice |  |  |
| Sorbet |  |  |

**Supplementary Table 1\_continued** Included individual food items in the BLEND Dataset

|  |  |  |
| --- | --- | --- |
| **Food Items** | **Food Items** | **Food Items** |
| **Fish and fish Products** | **Fats and oils** | **Grain and grain products** |
| Fish in miscellaneous | Fats and oils in miscellaneous | Grain in miscellaneous |
| *Clupeiformes* | *Butter* | *Wheat basic products* |
| Herring | *Margarine* | Oats basic products |
| Salmon and trout | *Animal fat* | Oatmeal |
| Whitefish | *Marine Oil* | *Barley basic products* |
| *Perciformes* | *Vegetable fats and oil* | Barley meal |
| Tuna | Vegetable fats and oil in miscellaneous | *Rice basic products* |
|  | Olive oil | Brown rice  |
|  | Sunflower oil | Polished rice  |
|  | Compound fats and oils | *Basic products of other cereals*  |
|  |  | *Pasta and noodles* |
|  |  | Pasta and noodles in miscellaneous |
|  |  | *Leavened bread* |
|  |  | Leavened bread in miscellaneous |
|  |  | Rye bread |
|  |  | Wheat bread |
|  |  | *Unleavened bread and crispbreads* |
|  |  | Crispbread |
|  |  | *Bread products* |
|  |  | *Fine bakery wares* |
|  |  | Savoury biscuits |
|  |  | Sweet biscuits and cookies |
|  |  | Currant bun |
|  |  | Dough cakes |
|  |  | Doughnut |
|  |  | Danish pastry |
|  |  | Mince pie |
|  |  | Baked cake |

**Supplementary Table 1\_continued** Included individual food items in the BLEND Dataset

|  |  |  |
| --- | --- | --- |
| **Food Items** | **Food Items** | **Food Items** |
| *Savoury cereal dishes* | **Pulses, seeds, kernels, and nut products** | **Vegetable and vegetable products** |
| Savoury pancake | Pulses, seeds, kernels, and nut products in miscellaneous | Vegetable in miscellaneous |
| Risotto | Pulses | *Leaf vegetable* |
| Savoury pie | Dried broad bean | Endive |
| *Breakfast cereals* | Lentil | Lettuce |
|  | Mung bean | Spinach |
|  | Soya bean | *Brassicas* |
|  | *Underground pulses* | Cauliflower |
|  | Peanut | Broccoli |
|  | *Seeds and kernels* | Red cabbage |
|  | Sunflower seed | Chinese cabbage |
|  | Pine nut | Cabbage |
|  | *Nuts* | Turnip tops |
|  | Walnut | Brussels sprout |
|  | Hazelnut | Kohlrabi |
|  | Hickory nut | Curly kale |
|  | Cashew nut | *Stalk vegetables* |
|  | Sweet chestnut | Rhubarb |
|  | *Pulse products* | Celery |
|  | Unfermented soya paste (tofu) | Fennel |
|  | Peanut butter | *Shoot vegetable* |
|  | *Nut and seed products* | Asparagus |
|  |  | Chicory |
|  |  | Globe artichoke |
|  |  | Bamboo shoot |
|  |  | *Onion family Vegetable* |
|  |  | Garlic |
|  |  | Onion |
|  |  | Leek |

**Supplementary Table 1\_continued** Included individual food items in the BLEND Dataset

|  |  |  |
| --- | --- | --- |
| **Food Items** | **Food Items** | **Food Items** |
| Garlic | Runner bean | *Berries* |
| Chives | Sweet corn | White grapes |
| *Tubers* | Okra | Strawberries |
| Tubers in miscellaneous | Avocado | Blueberries |
| New potato | Olive | *Citrus fruit* |
| Main crop | *Edible Fungi* | Orange |
| Sweet potato | Cultivated mushrooms | Grapefruit  |
| Yam | Truffle | *Miscellaneous* |
| *Root vegetable* | Morel | Banana |
| Carrot | *Seaweed* | Pineapple |
| Parsnip | *Vegetable mixes* | Kiwi |
| Swede | Vegetable mixes in miscellaneous | Melon |
| Beetroot | Mustard and cress | Watermelon |
| Radish | Pot herb | *Fruit mixture* |
| Parsley | *Vegetable products* | *Fruit products* |
| *Fruit vegetable* | Tomato puree | Fruit products in miscellaneous |
| Fruit vegetable in miscellaneous | Pickled gherkins | Dried mixed fruit |
| Tomato | Pickled onion |  |
| Aubergine | Sauerkraut |  |
| Sweet pepper | **Fruit and fruit products** |  |
| Chilli pepper | Fruit in miscellaneous |  |
| Cucumber | *Malaceious fruit* |  |
| Courgette | Apple |  |
| Cucurbita squash | *Prunus species fruit* |  |
| *Pod and seed vegetable* | Apricot |  |
| Pea | Peach |  |
| Wax beans | Plum |  |
| Broad bean | Sweet cherry |  |
| French bean | *Other stone fruit* |  |

**Supplementary Table 1\_continued** Included individual food items in the BLEND Dataset

|  |  |  |
| --- | --- | --- |
| **Food Items** | **Food Items** | **Food Items** |
| **Sugar and sugar products** | **Beverages (non-milk)** | Tea in miscellaneous |
| Sugar in miscellaneous | Beverage in miscellaneous | Herbal tea |
| *Sugar* | *Beers in malt* | Decaffeinated coffee |
| *Other sugar* | Beer - alcohol >5% | Caffeinated coffee |
| Fructose | Beer - alcohol 3-5% | Instant coffee |
| *Sugar substitutes* | Beer - alcohol 1-3% | Coffee in miscellaneous |
| Nutritive sweeteners | Beer - alcohol <1% | *Water*  |
| *Honey* | *Wine* | *Fruit juice* |
| Syrup | Wine - alcohol >9% | Fruit juice in miscellaneous |
| Fruit syrup | Wine - alcohol 5-9% | Orange juice |
| Jams, marmalades and spreads | *Fortified and liqueur* | Grapefruit juice |
| *Chocolate and chocolate products* | *Liqueurs* | Apple juice |
| Chocolate in miscellaneous | *Spirits* | Grape juice |
| Milk chocolate bar | Whisky | *Vegetable juice* |
| *Chocolate-coated confectionery bars* | Gin | Tomato juice |
| *Non-chocolate confectionery* | Vodka | *Other juice* |
| Boiled sweet | *Alcoholic mixed drinks* |  |
|  | Punch - alcohol >1% |  |
|  | *Carbonated soft drinks* |  |
|  | Carbonated soft drinks in miscellaneous |  |
|  | Soda water |  |
|  | Cola |  |
|  | Lucozade |  |
|  | *Non-dilution still drinks* |  |
|  | Still lemonade |  |
|  | *Infusion drinks* |  |
|  | Infusion drinks in miscellaneous |  |
|  | Black tea |  |
|  | Green tea |  |



AUC=0.690

AUC=0.692

AUC=0.695

AUC=0.691

AUC=0.698

AUC=0.699

AUC=0.681

AUC=0.685

**Supplementary Figure 1** Receiver Operating Characteristic Curves for Each Cross-validation Run

AUC=Area Under the Curve

AUC=0.694

AUC=0.701

2.35%

26.11%

27.89%

39.11%

56.05%

78.64 %

100%

2.47%

Importance values

Input variables

**Supplementary Figure 2** Importance Values of Input Variables after C5.0 in the BLEND Dataset

A: milk and dairy products; B: eggs and egg products; C: meats and meat products; D: fishes and fish products; E: fats, oils and their products; F: grains and grain products; G: pulses, seeds, kernels, nuts and their products; H: vegetables and vegetable products; I: fruits and fruit products; J: sugar and sugar products; K: beverages (non-milk)

The importance values range from 0% to 100%, where 0% indicates ‘unimportant’ and 100% indicates ‘extremely important’.