# Classification of toxicity of metallic NPs

Tung X. Trinh

4/2/2021

Reference: Trinh et al. "Curation of datasets, assessment of their quality and completeness, and nanoSAR classification model development for metallic nanoparticles." Environmental Science: Nano 5.8 (2018): 1902-1910.

Load necessary libraries:

```
library(openxlsx)
library(caret)
```

```
## Loading required package: lattice
```

```
## Loading required package: ggplot2
```

```
library(randomForest)
```

```
## randomForest 4.6-14
```

```
## Type rfNews() to see new features/changes/bug fixes.
```

```
##
## Attaching package: 'randomForest'
```

```
## The following object is masked from 'package:ggplot2':
##
##     margin
```

```
library(data.table)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:data.table':
##
##     between, first, last
```

```
## The following object is masked from 'package:randomForest':
##
##     combine
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(svDialogs)
library(shiny)
library(shinythemes)
library(shinydashboard)
```

```
##
## Attaching package: 'shinydashboard'
```

```
## The following object is masked from 'package:graphics':
##
##     box
```

Load dataset:

```
DataMetal <- read.xlsx("www/MetalESN.xlsx", sheet = 1, startRow = 1, colNames = TRUE,
                       rowNames = FALSE, detectDates = FALSE, skipEmptyRows = TRUE,
                       skipEmptyCols = TRUE, rows = NULL, cols = NULL, check.names = FALSE,
                       namedRegion = NULL, na.strings = "NA", fillMergedCells = FALSE)
DataMetal <- select(DataMetal, c("Toxicity",
                                 "Dose",
                                 "Assay",
                                 "Time",
                                 "Species",
                                 "Cancer",
                                 "Cell_Tissue",
                                 "Cell_line",
                                 "SSA",
                                 "Zeta",
                                 "HSize",
                                 "CoreSize",
                                 "Coating",
                                 "Shape",
                                 "Metal"))
```

Split data into train and test set (70/30 )

```
set.seed(1991)
split_size <- floor(0.70 * nrow(DataMetal))
in_rows <- sample(c(1:nrow(DataMetal)), size = split_size, replace = FALSE)
train <- DataMetal[in_rows, ]
test <- DataMetal[-in_rows, ]
```

Train Random Forest model:

```
train.control <- trainControl(method = "repeatedcv", number = 10, repeats = 5)
RFmodel <- train(Toxicity ~ ., data = train, method = "rf", ntree = 100, trControl = train.control)
print(RFmodel)
```

```
## Random Forest
##
## 1403 samples
##   14 predictor
##    2 classes: 'NON_TOXIC', 'TOXIC'
##
## No pre-processing
## Resampling: Cross-Validated (10 fold, repeated 5 times)
## Summary of sample sizes: 1262, 1262, 1263, 1263, 1263, 1262, ...
```

```
## Resampling results across tuning parameters:
##
##   mtry  Accuracy   Kappa
##     2   0.8238213  0.1845892
##    68   0.9154577  0.7138967
##   134   0.9158772  0.7215603
##
## Accuracy was used to select the optimal model using the largest value.
## The final value used for the model was mtry = 134.
```

Use RFmodel to predict test set:

```
predictions <- RFmodel %>% predict(test); predictions_train <- RFmodel %>% predict(train)
```

Get confusion matrix and performance of model:

```
CMatrix <- confusionMatrix(predictions, as.factor(test$Toxicity))
Performance <- data.frame(Parameter = row.names(as.data.frame(CMatrix$byClass)),
                          Value = as.data.frame(CMatrix$byClass))
colnames(Performance) <- c("Parameters", "Values")
CMatrix
```

```
## Confusion Matrix and Statistics
##
##            Reference
## Prediction  NON_TOXIC TOXIC
##   NON_TOXIC       480    25
##   TOXIC            14    83
##
##                Accuracy : 0.9352
##                  95% CI : (0.9125, 0.9535)
##     No Information Rate : 0.8206
##     P-Value [Acc > NIR] : <2e-16
##
##                   Kappa : 0.7709
##
##  Mcnemar's Test P-Value : 0.1093
##
##             Sensitivity : 0.9717
##             Specificity : 0.7685
##          Pos Pred Value : 0.9505
##          Neg Pred Value : 0.8557
##              Prevalence : 0.8206
##          Detection Rate : 0.7973
##    Detection Prevalence : 0.8389
##       Balanced Accuracy : 0.8701
##
##        'Positive' Class : NON_TOXIC
##
```

```
Performance
```

```
##                               Parameters    Values
## Sensitivity                  Sensitivity 0.9716599
## Specificity                  Specificity 0.7685185
## Pos Pred Value             Pos Pred Value 0.9504950
## Neg Pred Value             Neg Pred Value 0.8556701
```

```
## Precision                         Precision 0.9504950
## Recall                               Recall 0.9716599
## F1                                       F1 0.9609610
## Prevalence                       Prevalence 0.8205980
## Detection Rate             Detection Rate 0.7973422
## Detection Prevalence Detection Prevalence 0.8388704
## Balanced Accuracy       Balanced Accuracy 0.8700892
```

Make an interactive app for users:

```
ui <- dashboardPage(
  dashboardHeader(title = "QSAR metal NPs", titleWidth = 350),
  dashboardSidebar(
    width = 350,
    sidebarMenu(
      menuItem("Model", tabName = "Model", icon = icon("th")),
      menuItem("Introduction", tabName = "Introduction", icon = icon("dashboard"))
    )
  ),
  dashboardBody(
    tabItems(
      # Model tab content
      tabItem(tabName = "Model",
    h2("A web-based app for predicting cytotoxicity of metal nanoparticles (i.e., Ag and Au)"),
    h2("Reference: Trinh et al.  Environmental Science: Nano 5.8 (2018): 1902-1910."),
    # Boxes need to be put in a row (or column)
    fluidRow(
      box(
        height = 500,
        title = "Input properties of NPs:",
        selectInput("Metal", "Metal NPs:",
                    c("Au" = "Au",
                      "Ag" = "Ag")),
        selectInput("Shape", "Shape:",
                    c("Nanorod" = "Nanorod",
                      "Sphere" = "Sphere",
                      "Hollow" = "Hollow")),
        sliderInput("CoreSize", "Core diameter (nm):", 1, 100, 10),
        sliderInput("HSize", "Hydrodynamic diameter (nm):", 1, 300, 50),
        sliderInput("Zeta", "Zeta potential (mV):", -20, 20, 0)

      ),
      box(
        height = 500,
        title = "Input experimental conditions:",
        selectInput("Cell_line", "Cell line:",
                    c("HeLa" = "HeLa",
                      "HepG2" = "HepG2",
                      "BEAS-2B" = "BEAS-2B",
                      "A549" = "A549")),
        selectInput("Assay", "Toxicity assay method:",
                    c("MTS" = "MTS",
                      "MTT" = "MTT",
                      "AlamarBlue" = "Alamar Blue",
                      "NRU" = "NRU")),
```

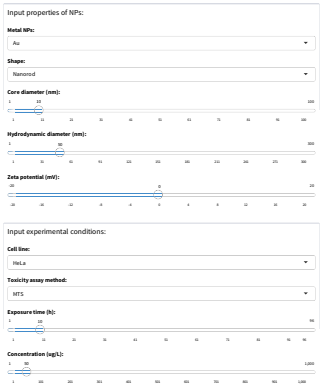
```
        sliderInput("Time", "Exposure time (h):", 1, 96, 10),
        sliderInput("Dose", "Concentration (ug/L):", 1, 10^3, 50),
      ),

      box(
        height = 500,
        title = "Model performance:",
        tableOutput("Performance")
      ),


      box(height = 120, title = "Predicted toxicity:", tableOutput("Prediction")),

      box(height = 360, title = "Summary of input:", tableOutput("SummaryInput"))
    )
  ),
  # Introduction tab content
  tabItem(tabName = "Introduction",
          h2("A web-based app for predicting cytotoxicity of metal nanoparticles (i.e., Ag and Au)"),

          column(
            br(),
            p("Dataset for model development was published in:",
              strong("Trinh et al.  Environmental Science: Nano 5.8 (2018): 1902-1910."),
              style="text-align:justify;color:black;
              background-color:papayawhip;padding:15px;border-radius:10px"),
            br(),
            width=8),
  )
  )
  )
)

server <- function(input, output) {

  output$Performance <- renderTable({
    Performance
  }, digits = 2)

  output$SummaryInput <- renderTable({
    data.frame("Descriptor" = c("Metal NPs",
                                "Core size (nm)",
                                "Hydrodynamic diameter (nm)",
                                "Zeta potential (mV)",
                                "Cell line",
                                "Toxic Assay",
                                "Exposure time (h)",
                                "Concentration (ug/L)"),
               "Values" = c(input$Metal,
                            input$CoreSize,
                            input$HSize,
                            input$Zeta,
                            input$Cell_line,
```

```
                              input$Assay,
                              input$Time,
                              input$Dose))
  }, digits = 2)

  output$Prediction <- renderTable({
    table1 <- data.frame("Toxicity" = "UNKNOWN",
                         "Dose" = input$Dose,
                         "Assay" = input$Assay,
                         "Time" = input$Time,
                         "Species" = "Human",
                         "Cancer" = 1,
                         "Cell_Tissue" = "Lung",
                         "Cell_line" = input$Cell_line,
                         "SSA" = 20,
                         "Zeta" = input$Zeta,
                         "HSize" = input$HSize,
                         "CoreSize" = input$CoreSize,
                         "Coating" = "None",
                         "Shape" = "Sphere",
                         "Metal" = input$Metal)

    table2 <- as.data.frame(predict(RFmodel, table1))

    data.frame("Observed.Toxicity" = table1[1,1], "Predicted.Toxicity" = table2[1,1])


  })
}

shinyApp(server = server, ui = ui)
```

```
##
## Listening on http://127.0.0.1:6735
```

A web-based app for predicting cytotoxicity of metal nanoparticles (i.e., Ag and Au)

Reference: Trinh et al. Environmental Science: Nano 5.8 (2018): 1902-1910.




Model performance:

| Parameters | Values |
|---|---|
| Sensitivity | 0.97 |
| Specificity | 0.77 |
| Pos Pred Value | 0.95 |
| Neg Pred Value | 0.86 |
| Precision | 0.95 |
| Recall | 0.97 |
| F1 | 0.96 |
| Prevalence | 0.82 |
| Detection Rate | 0.80 |
| Detection Prevalence | 0.84 |
| Balanced Accuracy | 0.87 |

Predicted toxicity:

| Observed.Toxicity | Predicted.Toxicity |
|---|---|
| UNKNOWN | NON_TOXIC |

Summary of input:

| Descriptor | Values |
|---|---|
| Metal NPs | Au |
| Core size (nm) | 10 |
| Hydrodynamic diameter (nm) | 50 |
| Zeta potential (mV) | 0 |
| Cell line | HeLa |
| Toxic Assay | MTS |
| Exposure time (h) | 10 |
| Concentration (ug/L) | 50 |