# Appendices

## Appendix A: R Studio Code

install.packages("ggplot2")

install.packages("tidyr")

install.packages("VGAM")

install.packages("MASS") #

install.packages("ROCR")

install.packages("mltest")

install.packages("gclus") # For plotting scatterplot matrices and parallel coordinates

install.packages("GGally") # Txtends ggplot2 for very customized matrix plot

install.packages("gpairs") # To display the generalized pairs plot

install.packages ("lattice") # For fdata visualization, delivers great plots especially for multivariate plots

install.packages("caTools")

install.packages("party")

install.packages("dplyr")

install.packages("magrittr")

install.packages("rpart.plot")

install.packages("rpart")

install.packages("randomForest")

install.packages("caret")

install.packages("randomForestSRC")

install.packages("sensitivity")

install.packages("class")

install.packages("nnet")

install.packages("starter")

install.packages("verification")

library(verification)

library(starter)

library(class)

library(ggplot2)

library(lattice)

library(tidyr)

library(gclus)

library(VGAM)

library(MASS)

library(ROCR)

library(mltest)

library(GGally)

library(gpairs)

library(caTools)

library(party)

library(randomForest)

library(dplyr)

library(magrittr)

library(rpart)

library(rpart.plot)

library(caret)

library(randomForestSRC)

library(sensitivity)

library(MASS)

library(nnet)

library(caTools)

data<-read.csv("Phase\_field\_updated\_model\_data\_21Aug2023.csv", header=T)

colSums(is.na(data))

# Changing string variables to numerical

data$Synthesis\_Route<-as.factor(data$Synthesis\_Route)

levels(data$Synthesis\_Route)<-c("ST\_6","ST\_4","DC","IM","DMD")

data$Hot\_Cold\_Working<-as.factor(data$Hot\_Cold\_Working)

levels(data$Hot\_Cold\_Working)<-c("CW","CR")

data$Multiphase<-as.factor(data$Multiphase)

levels(data$Multiphase)<-c("MgZn2","Im","gamma-Mg29Cu6","epsilon-Mg6Mn","delta-MgZn","beta-Mg17Al12","alpha-Mg","Al8Mn5","Al2CuMn","Al11Mn4")

data$IM\_Structure<-as.factor(data$IM\_Structure)

levels(data$IM\_Structure)<-c("Al8Mn5","alpha-Mg","beta-Mg17Al12","Al8Mn6","Al8Mn7","beta-Mg2Si","epsilon-Mg3Al2Zn3","eta-Mg(Zn,Al,Cu)2","gamma-Mg29Al7","gamma-Mg2Si","Im","Mg2Zn11","MgZn2 Mg2Cu","S-Al2CuMg","T-Mg32(Al,Zn,Cu)49")

data$Structure<-as.factor(data$Structure)

levels(data$Structure)<-c("HCP","BCC","Laves","quasicrystal","gamma-FCC","S\_Laves","FCC","CCP")

data$Phases<-as.factor(data$Phases)

levels(data$Phases)<-c("alphaMg","beta-Mg17Al12","Al8Mn5","Al8Mn6","Im","Al8Mn7","gamma-Mg2Si","gamma-Mg29Al7","epsilon-Mg3Al2Zn3","T-Mg32(AlZnCu)49","eta-Mg(ZnAlCu)2","S-Al2CuMg","beta-Mg2Si","T-Mg32(AlZnCu)49","eta-Mg(ZnAlCu)2","Mg2Cu","MgZn2","Al8Mn5","alpha-Mg,MgZn2,Mg2Zn11","Mg2Cu","Al8Mn5","Im,MgZn2,Mg2Zn11,Mg32(AlZn)49,Mg17Al1","gamma-Mg29Cu6","delta-MgZn","epsilon-Mg6Mn")

str(data)

summary(data)

ggplot(data,aes(x=dHmix,fill=Microstructure,color=Microstructure)) + geom\_histogram(binwidth = 2,color="black") + labs(x = "Enthalpy of Mixing",y = "Frequency", title = "Microstructure w.r.t. Enthalpy of Mixing")

multiphasetable <- table(data$Multiphase)

pct<-round(multiphasetable/sum(multiphasetable)\*100)

lbls1<-paste(names(multiphasetable),pct)

lbls<-paste(lbls1, "%", sep="")

microstructuretable <- table(data$Phases)

pct<-round(microstructuretable/sum(microstructuretable)\*100)

lbls1<-paste(names(microstructuretable),pct)

lbls<-paste(lbls1, "%", sep="")

# Machine Learning

set.seed(100)

#100 is used to control the sampling permutation to 100.

index<-sample(nrow(data),0.75\*nrow(data))

train<-data[index,]

test<-data[-index,]

# Correlation among numeric data

ggcorr(train) # Non-numeric data ignored

ggplot(train,

# define axes

aes(x=VEC,

y= Atom\_Size\_Diff,

# define to color by Phases

color = Phases)) +

# define plot type

geom\_point() +

# define size of points and themes

geom\_point(size=8.0)+

theme(text = element\_text(size = 20))

ggplot(train,

# define axes

aes(x=VEC,

y= Elect\_Diff,

# define to color by Phases

color = Phases)) +

# define plot type

geom\_point() +

# define of points and themes

geom\_point(size=8.0)+

theme(text = element\_text(size = 20))

ggplot(train,

# define axes

aes(x=VEC,

y= Atom\_Size\_Diff,

# define to color by Phases

color = IM\_Structure)) +

# define plot type

geom\_point() +

# define black and white (bw)

geom\_point(size=1.0)+

theme(text = element\_text(size = 10))+

stat\_ellipse()

ggplot(train,

# define axes

aes(x=VEC,

y= Atom\_Size\_Diff,

# define to color by Phases

color = Structure)) +

# define plot type

geom\_point() +

# define black and white (bw)

geom\_point(size=8.0)+

theme(text = element\_text(size = 30))

ggplot(data,

# define axes

aes(x=VEC,

y= UTS,

# define to color by Phases

color = Multiphase)) +

# define plot type

geom\_point() +

# define of points and themes

geom\_point(size=8.0)+

theme(text = element\_text(size = 20))

ggplot(train,

# define axes

aes(x=VEC,

y= Yield,

# define to color by Phases

color = Phases)) +

# define plot type

geom\_point() +

# define of points and themes

geom\_point(size=8.0)+

theme(text = element\_text(size = 30))

## # Fit the LDA model

# Fit LDA model

fit.LDA = lda(Synthesis\_Route ~ Yield+UTS+E+VEC+dHmix, train)

fit.LDA

lda\_model <- lda(Phases ~ Yield + UTS + E + VEC +dHmix, train)

lda\_model

lda\_pred <- predict(lda\_model, test)

lda\_pred

test$Yield

lda\_pred$Yield

lda\_acc <- mean(lda\_pred == test$Phases)

lda\_acc

# Fit Random Forest model

rf\_model <- randomForest(Synthesis\_Route ~ Yield + UTS + E + VEC + dHmix, data = train)

rf\_model

rf\_pred <- predict(rf\_model, test)

rf\_pred

rf\_acc <- mean(rf\_pred == test$Synthesis\_Route)

rf\_acc

# Fit artificial neural network model

ann\_model <- nnet(Multiphase ~ Yield + UTS + E + VEC + Atom\_Size\_Diff+ dHmix, data = train, size = 10, maxit = 1000)

ann\_model

ann\_pred <- predict(ann\_model, test, type = "class")

ann\_pred

ann\_acc <- mean(ann\_pred == test$Multiphase)

ann\_acc

confusionMatrix <- table(Predicted = ann\_pred, Actual = test$Multiphase)

print(confusionMatrix)

# Print the performance of the models

cat("Performance of LDA:", lda\_acc, "\n")

cat("Performance of Random Forest:", rf\_acc, "\n")

cat("Performance of k-nearest neighbours:", knn\_acc, "\n")

cat("Performance of artificial neural network:", ann\_acc, "\n")

## Appendix B: Data Structure and Summary of Data

> data<-read.csv("Phase\_field\_updated\_model\_data\_21Aug2023.csv", header=T)

> colSums(is.na(data))

Mg Al Cu

0 0 0

Mn Zn UTS

0 0 0

Yield E Num\_of\_Elem

0 0 0

Density\_calc dHmix dSmix

0 0 0

dGmix Tm n\_Para

0 0 0

Atom\_Size\_Diff Elect\_Diff VEC

0 0 0

Synthesis\_Route Hot\_Cold\_Working Homogenization\_Temp

0 0 0

Homogenization\_Time Annealing\_Temp Annealing\_Time

0 0 0

Quenching Multiphase IM\_Structure

0 0 0

Phases Structure

0 0

> str(data)

'data.frame': 60 obs. of 29 variables:

$ Mg : num 91.2 91.2 91.2 91.2 91.2 91.2 91.2 91.2 91.2 91.2 ...

$ Al : num 8.3 8.3 8.3 8.3 8.3 8.3 8.3 8.3 8.3 8.3 ...

$ Cu : num 0 0 0 0 0 0 0 0 0 0 ...

$ Mn : num 0.15 0.15 0.15 0.15 0.15 0.15 0.15 0.15 0.15 0.15 ...

$ Zn : num 0.35 0.35 0.35 0.35 0.35 0.35 0.35 0.35 0.35 0.35 ...

$ UTS : int 270 270 270 270 270 270 270 270 230 230 ...

$ Yield : int 145 145 145 145 90 90 90 90 150 150 ...

$ E : num 47.5 47.5 47.5 47.5 35.5 ...

$ Num\_of\_Elem : int 4 4 4 4 4 4 4 4 4 4 ...

$ Density\_calc : num 1.8 1.8 1.8 1.8 1.8 ...

$ dHmix : num -656 -656 -656 -656 -656 ...

$ dSmix : num 81.3 81.3 81.3 81.3 81.3 ...

$ dGmix : num -128595 -128595 -128595 -128595 -128595 ...

$ Tm : num 651 651 651 651 651 ...

$ n\_Para : num 6.31 6.31 6.31 6.31 6.31 ...

$ Atom\_Size\_Diff : num 1.04e-06 1.04e-06 1.04e-06 1.04e-06 1.04e-06 ...

$ Elect\_Diff : num 0.0803 0.0803 0.0803 0.0803 0.0803 ...

$ VEC : num 2.08 2.08 2.08 2.08 2.08 ...

$ Synthesis\_Route : Factor w/ 5 levels "ST\_6","ST\_4",..: 5 5 5 5 4 4 4 4 1 1 ...

$ Hot\_Cold\_Working : Factor w/ 2 levels "CW","CR": 2 2 2 2 2 2 2 2 2 2 ...

$ Homogenization\_Temp: int 400 400 400 400 400 400 400 400 400 400 ...

$ Homogenization\_Time: int 1440 1440 1440 1440 1440 1440 1440 1440 1440 1440 ...

$ Annealing\_Temp : int 200 200 200 200 200 200 200 200 200 200 ...

$ Annealing\_Time : int 30 30 30 30 30 30 30 30 30 30 ...

$ Quenching : num 100 100 100 100 100 100 100 100 100 100 ...

$ Multiphase : Factor w/ 10 levels "MgZn2","Im","gamma-Mg29Cu6",..: 4 4 4 4 4 4 4 4 4 4 ...

$ IM\_Structure : Factor w/ 15 levels "Al8Mn5","alpha-Mg",..: 2 2 2 2 3 3 3 3 4 4 ...

$ Phases : Factor w/ 20 levels "alphaMg","beta-Mg17Al12",..: 6 7 3 14 6 7 4 14 6 7 ...

$ Structure : Factor w/ 13 levels " HCP,Laves,quasicrystal",..: 4 4 4 4 5 5 5 5 5 5 ...

> summary(data)

Mg Al Cu Mn

Min. :20.00 Min. : 0.000 Min. : 0.000 Min. : 0.000

1st Qu.:50.00 1st Qu.: 5.300 1st Qu.: 0.000 1st Qu.: 0.150

Median :91.20 Median : 8.300 Median : 0.000 Median : 0.260

Mean :76.74 Mean : 8.442 Mean : 4.818 Mean : 4.806

3rd Qu.:92.06 3rd Qu.:12.500 3rd Qu.:12.500 3rd Qu.:12.500

Max. :95.34 Max. :20.000 Max. :20.000 Max. :20.000

Zn UTS Yield E

Min. : 0.000 Min. :210.0 Min. : 90.0 Min. : 35.51

1st Qu.: 0.350 1st Qu.:231.5 1st Qu.:125.0 1st Qu.: 49.31

Median : 2.500 Median :270.0 Median :145.0 Median :100.00

Mean : 5.193 Mean :322.1 Mean :182.9 Mean : 94.56

3rd Qu.:12.500 3rd Qu.:394.0 3rd Qu.:230.0 3rd Qu.:123.00

Max. :20.000 Max. :717.0 Max. :450.0 Max. :151.00

Num\_of\_Elem Density\_calc dHmix dSmix

Min. :3.000 Min. :1.769 Min. :-688.2 Min. : 45.36

1st Qu.:3.750 1st Qu.:1.798 1st Qu.:-667.4 1st Qu.: 72.06

Median :4.000 Median :1.808 Median :-656.3 Median : 81.33

Mean :4.117 Mean :2.120 Mean :-622.2 Mean :158.26

3rd Qu.:5.000 3rd Qu.:2.613 3rd Qu.:-553.0 3rd Qu.:313.88

Max. :5.000 Max. :3.745 Max. :-416.1 Max. :434.77

dGmix Tm n\_Para Atom\_Size\_Diff

Min. :-684310 Min. :645.7 Min. :0.000000 Min. :4.834e-07

1st Qu.:-494287 1st Qu.:650.3 1st Qu.:0.001279 1st Qu.:1.040e-06

Median :-128595 Median :652.0 Median :4.731995 Median :1.249e-06

Mean :-249570 Mean :688.5 Mean :3.335107 Mean :1.174e-06

3rd Qu.:-114018 3rd Qu.:751.3 3rd Qu.:6.309166 3rd Qu.:1.320e-06

Max. : -72043 Max. :812.1 Max. :6.340518 Max. :1.715e-06

Elect\_Diff VEC Synthesis\_Route Hot\_Cold\_Working

Min. :0.05920 Min. :2.000 ST\_6:16 CW:16

1st Qu.:0.07214 1st Qu.:2.054 ST\_4: 5 CR:44

Median :0.08034 Median :2.077 DC :17

Mean :0.11114 Mean :2.148 IM : 8

3rd Qu.:0.17912 3rd Qu.:2.293 DMD :14

Max. :0.18960 Max. :2.576

Homogenization\_Temp Homogenization\_Time Annealing\_Temp Annealing\_Time

Min. :150.0 Min. : 60 Min. :200.0 Min. : 30

1st Qu.:400.0 1st Qu.: 210 1st Qu.:200.0 1st Qu.: 30

Median :400.0 Median :1440 Median :300.0 Median : 30

Mean :394.2 Mean : 863 Mean :261.7 Mean : 39

3rd Qu.:400.0 3rd Qu.:1440 3rd Qu.:300.0 3rd Qu.: 30

Max. :500.0 Max. :1440 Max. :350.0 Max. :120

Quenching Multiphase IM\_Structure Phases

Min. : 8.333 beta-Mg17Al12:33 Al8Mn6 :12 Al8Mn7 :10

1st Qu.: 100.000 epsilon-Mg6Mn:12 eta-Mg(Zn,Al,Cu)2: 9 gamma-Mg2Si :10

Median : 100.000 Im : 5 Al8Mn7 : 8 Mg2Cu :10

Mean : 170.417 delta-MgZn : 4 gamma-Mg29Al7 : 8 Al8Mn5 : 5

3rd Qu.: 100.000 MgZn2 : 3 alpha-Mg : 4 Im : 3

Max. :1000.000 gamma-Mg29Cu6: 3 beta-Mg17Al12 : 4 gamma-Mg29Cu6: 3

(Other) : 0 (Other) :15 (Other) :19

Structure

HCP,BCC,quasicrystal :20

quasicrystal,BCC,Laves ,FCC,CCP: 6

HCP,BCC,Laves : 5

HCP,BCC,Laves,quasicrystal : 4

quasicrystal,HCP,S\_Laves,BCC : 4

quasicrystal,Laves : 4

(Other) :17