Supplement Table 1: List of top ranking (top 10) mutations. First four rows are header lines; fifth line is for column headings. Columns 1 to 9 provide mutation information, such as the chromosome (CHROM), genomic position (POS), unique identifier (ID), reference allele (REF), alternative alleles (ALT) identified in different lineages, sequence quality score (QUAL), filtering out (FILT) criteria for low quality mutations, any information (INFO) and format of the mutation, GT – Genotype. Genotype data are given for two samples, one for USA and the other from India. Missing information is denoted by period "." symbol. If more than one alternative alleles exist, those are comma-separated in ALT column. The nucleotide symbols in REF and ALT columns are: A – Adenine; C – Cytosine; G – Guanine; T – Thymine; R – G or A (purine); Y – C or T (pyrimidine); K – G or T; M – A or C; S – G or C; W – A or T; B – G or T or C; D – G or A or T; H – A or C or T; V – G or C or A.

##fileformat=VCFv4.2

##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">

##reference=file:///xtdisk/apod/licp/Virus/ref/2019-nCoV.fa

##contig=<ID=2019-nCoV,length=29903>

#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	2019-nCoV/USA-AZ1/2020	SARS-CoV-2/human/IND/GMCKN318/2020
2019-nCoV	23403	2019-nCoV_23403	A	R,G	.	.	.	GT	0	2
2019-nCoV	14408	2019-nCoV_14408	C	A,T,Y	.	.	.	GT	0	2
2019-nCoV	3037	2019-nCoV_3037	C	T,Y	.	.	.	GT	0	1
2019-nCoV	241	2019-nCoV_241	C	T,Y	.	.	.	GT	0	1
2019-nCoV	25563	2019-nCoV_25563	G	T,C,R,K	.	.	.	GT	0	0
2019-nCoV	1059	2019-nCoV_1059	C	T,Y	.	.	.	GT	0	0
2019-nCoV	28881	2019-nCoV_28881	G	A,T,R	.	.	.	GT	0	0
2019-nCoV	28882	2019-nCoV_28882	G	A,T,R	.	.	.	GT	0	0
2019-nCoV	28883	2019-nCoV_28883	G	A,S,C	.	.	.	GT	0	0
2019-nCoV	8782	2019-nCoV_8782	C	T,Y	.	.	.	GT	1	0

Supplementary Figure 1: Flowchart of the computational frame-work

```
┌─────────────────────────────────────┐   ┌──────────────────────┐
│ VCF data-file (File name "2019-      │   │ Variant annotation   │
│ nCoV_total.vcf") of 10,261 (rows)    │   │ file for 10,034      │
│ and 20,746 SARS-Cov-2                │   │ mutations            │
│ samples/lineages (columns)           │   │                      │
└─────────────────────────────────────┘   └──────────────────────┘
```

Perl Script to prepare the Data Matrix X (a binary matrix, where 1 denotes presence of mutation and 0 otherwise) and Y – Vector of factor/class variable

Split the Data

$4/5^{th}$ training set

$1/5^{th}$ testing set

Select most discriminating mutations (**42**) by applying *varSelRF* R package

Apply the classifiers on test set and report performance metrics

Build two separate classifiers by applying *randomForest* and rpart R packages

Report classifiers' performance metrics by cross-validation