

**Flavivirus Genomic Data**

```

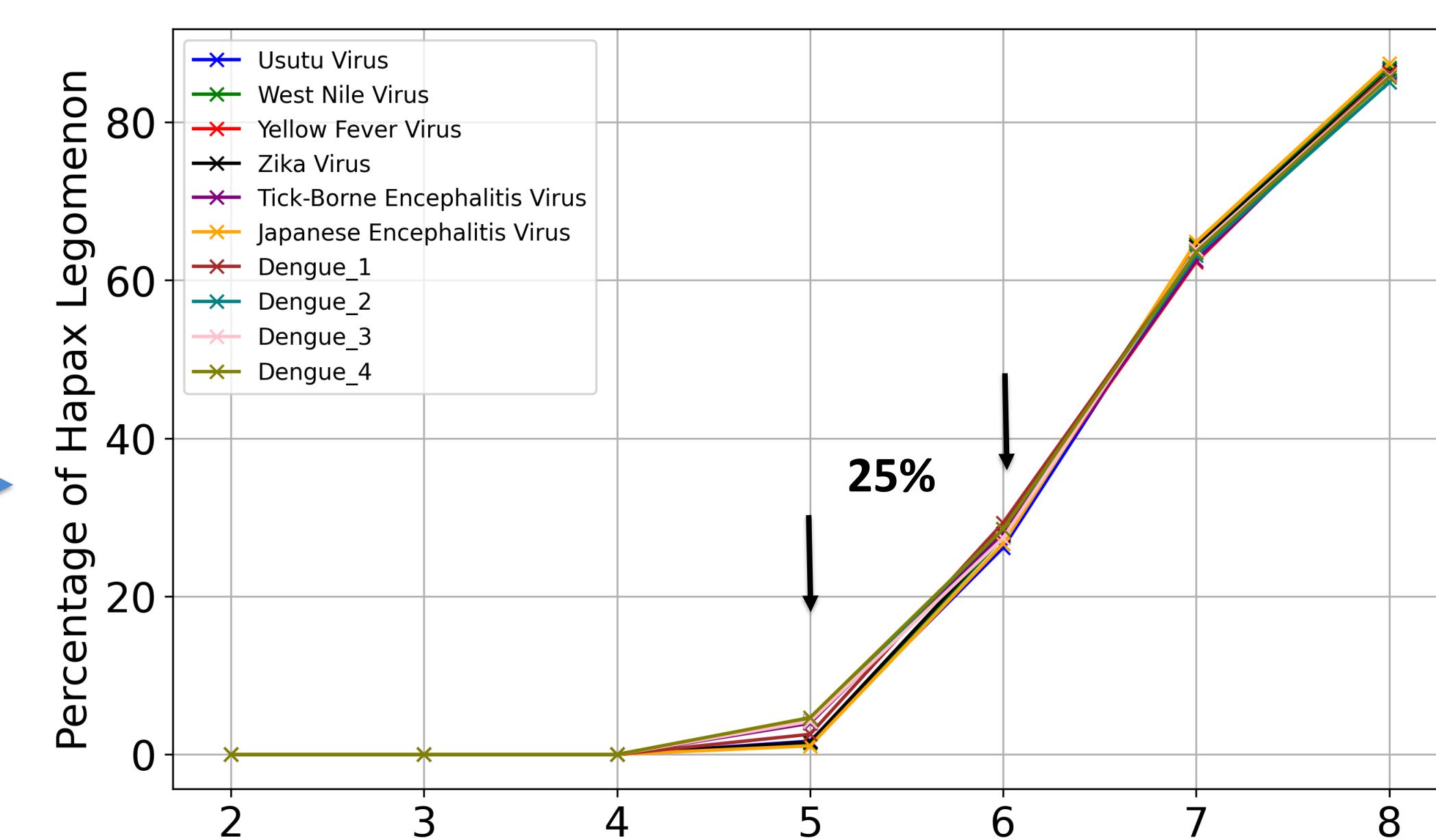
CATGCCATTGCGCTGCTGCC
CTGC GGCAATGTGCCCTCGC
CGCTATCCATTGCGCTGCTG
CCCTCTCGCCGCTATCACCA
CGKG CATCARAGCAGGGCTC
GTGATGYGAACCACGGCAT
  
```

X Ambiguous Characters

$\geq 95\%$  Genomic Coverage

Ambiguous Characters

Preprocessing



Zipfian Distribution and  
Hapax Legomenon Analysis

Kmer =5 Encodings (Total 1024)		Kmer =6 Encodings (Total 4096)	
K-mer	Encoding	K-mer	Encoding
TGGAA	1	AAAGAA	5
AAGAA	3	GCGCGC	4092
GAAAA	4	CGTAA	4093
CGCGC	1024	TAAGTA	4096

Selection and Encoding of  
Universal K-mer Libraries

Input for DiCNN

Dual Input Architecture

