JSRT JOURNAL
PAGES: 13-20
12/15/23
JOURNAL OF SCIENTIFIC RESEARCH AND TECHNOLOGY(JSRT) VOLUME-1 ISSUE-9 DECEMBER
Registered under MSME Government of India ISSN: 2583-8660

Leverage Machine Learning To Infer Proof of the Nipah Influenza

Dr. Shubhangi D C[1], Dr. Baswaraj Gadgay[2], S. Anita[3]

[1]*Professor, Department of Computer Science, Visvesvaraya Technological University CPGS Kalaburagi, Karnataka, India, drshubhangipatil1972@gmail.com*
[2]*Professor, Department of Electronics and Communication, Visvesvaraya Technological University CPGS Kalaburagi, Karnataka, India, baswaraj_gadgay@vtu.ac.in*
[3]*Student, Department of Computer Science, Visvesvaraya Technological University CPGS Kalaburagi, Karnataka, India, anitaswamy1995@gmail.com*

ABSTRACT

Nipah virus is highly fatal virus which spreads from bats to humans & other animals. Due to the fatality of the virus, the aim of this effort is to detect and identify it as soon as possible by understanding the efficiency of machine learning. From the perspective of medicine, the Nipah virus is not treatable using vaccines or medications that have been shown effective. In the field of medicine, machine learning algorithms are crucial for employing ML predictors to isolate the virus in dubious and urgent cases. This technique will produce numerical results to show if a patient has Nipah virus infection or not. Since there is currently no vaccine for the Nipah virus, care must be taken because "prevention is better than cure." To improve model accuracy, more machine learning methods, like Random forest and Decision tree are being applied.

Keywords— Nipah, Random Forest[RF], Decision tree[DT] , Restricted Boltzmann Machine(RBM)

I. INTRODUCTION

A set of numerical phylogenetic indicators derived from computations using DNA sequence allows for the classification of closely related viruses. To estimate role of natural selection in development of a given sequence, we look for such markers that deviate considerably from what would be predicted for random mutations. In Asia, pig farmers have been hit hard by outbreaks of Nipah virus, a zoonotic contagious virus. Over the course of hundreds of years, infectious diseases including SARS, swine flu, MERS, avian flu, Nipah, & Henipa have caused global public health disasters. Infections may linger dormant for years before reemerging, infecting hundreds of millions of individuals worldwide. NiV, a novel zoonotic paramyxovirus, is closely related to Hydra virus. An epidemic of illness in pigs and people in Malaysia led to the discovery of this virus around the end of 1998 or early 1999. System employs ML models like Random forest & Decision tree to make Nipah virus forecasts. Such models allow us to forecast geographically dispersed cases of viral infection. That we could be prudent in this matter. Foreground prediction is handled through Random Forest & Decision tree classifier methods. Six symptoms related to Nipah virus are included in dataset being utilized. Furthermore, user's sickness history may be used to inform system's recommendations for safety measures.

II. RELATED WORK

Mathematical models[1] that have been built to investigate viral zoonoses in wild animals and point out where further work needs to be done in this field. Author presented an ODE-based mathematical model for characterizing host-pathogen interactions. developed a novel host-pathogen model[2] which takes into account seasonal migration & reproduction as well as seasonal changes in the host's surroundings.[10] Three main conclusions can be drawn from this model's analysis.Nipah virus[12] infections in Malaysia & following epidemics in India and Bangladesh have very different epidemiology & medical aspects. In this paper[13] These two viruses were proposed to comprise a novel genus within the Paramyxoviridae family. Nipah virus, like Hydra virus, is rare amongst paramyxoviruses since it could affect or even fatally illness a wide variety of host animals.

III. PROPOSED SYSTEM

we utilize deep learning through high-level programming. In order to make reliable predictions, data must be sent across many different levels. Here, the work is carried out in a certain order, from one layer to the next, in order to eliminate duplication in layered structure. High-yield strategies for diagnosis

treatment have been the subject of a lot of research. In the same vein, we use a Restricted Boltzmann Machine (RBM) to determine if Nipah virus is present in patient's system. It's structured like a bipartite graph, with a layer for intra-node communication. Using Random forest and decision tree for comparison study purpose.
Nipah virus dataset is used for this study purpose.
In the first round of data preparation, we shall select just 3 characteristics from data file. The adjacent dataset is divided in training & testing set, with, say, 20% testing data & 80% training data.
Projected Algorithm

Step-1: Pre-process dataset
Step-2: Divide dataset into training & testing samples
Step-3: Apply classifiers RF and DT. For solving regression problems, MSE for data branches from each node.

Step-4: Predict NiVD infection

$$MSE = \frac{1}{N}\sum_{i=1}(fi - yi)^2 \quad \text{.......} (1)$$

Where N is total number of observations, fi is predicted value, and yi is observed value at observation i.
Using this technique, you can determine how far off your predictions are for every node in forest, giving you better idea of which path to take. Here, fi is result provided by the decision tree, & yi is data point's value at node in question.
Utilizing Gini index—a measure for determining how nodes upon decision tree branch—is common practice when running Random Forests upon classification data.

$$Gini = 1 - \sum_{i=1}^{C} (p_i)^2 \quad \text{.......} (2)$$

Gini of every branch upon node is calculated using class & likelihood to establish most probable branch. Number of classes, c, is denoted by, while frequency, pi, of class being seen in dataset is denoted by.
A decision tree's branching structure may also be calculated using entropy.

$$Entropy = \sum_{i=1}^{C} -p_i * \log_2(p_i) \quad \text{.......} (3)$$

When deciding which way node must branch off in network, entropy considers likelihood of various outcomes. Logarithmic function employed in its calculation makes it more complex than Gini index.

Feature Extraction: In feature extraction equations related with shape aspects

$$Solidity = \frac{AreaofNucleus}{AreaofConvexhull} \quad \text{.......} (4)$$

$$Convexity = \frac{Perimeterofconvexhull}{PerimeterofNucleus}$$

$$Circularity = \frac{(PerimeterofNucleus)^2}{4x \prod x(AreaofNucleus)} \quad \text{.......} (5)$$

Color qualities make up second set of attributes. Hematologists knowledge suggests that, in along with form characteristics,
Random Forest: Covariate selection and explanatory power evaluation are two of its main applications. Using random selections of response variable & covariates, several decision trees are built, & predictions are averaged utilizing this machine learning technique. When compared to using a single decision tree, overfitting is greatly reduced when many trees are constructed & averaged. Python was used for linear modeling & random forests.
Random Forest Algorithm:

Step 1: The data or training set may be split into random samples.
Step 2: Each piece of training information will be used by the algorithm to create a decision tree.
Step 3: Voting process will be carried out by averaging tree of decisions.
Step 4: In the end, choose forecast outcome that received most votes and use that one.

Pseudocode:

```
To generate c classifiers:
for i=1 to c do
Randomly sample the training data D with
replacement to produce D,
Create a root node, Ni containing Di
call BuildTree (Ni)
end for
BuildTree(N):
if N contains instances of only one class then
return
else
Randomly select x % of the possible splitting
features in N
select the feature F with the highest information
gain to split on
Create f child nodes of N, N1.....Nf, where F as
f possible values(F1,.....,Ff)
for i=1 to f do
set the contents of Ni to Di, where Di is all
instances in N that match
F
Call BuildTree (Ni)
end for
end if
```

Ensemble is the term for these merged models. The ensemble employs 2 strategies:
Bagging: Bagging refers to the process of creating a new training subset from existing sample training data by swapping out some of the samples. The ultimate result is decided by a simple majority.
Boosting: Boosting is the process of combining many low-performing learners into a single high-performing learner via the use of sequential model creation. BOOST, XG BOOST, ADA BOOST, etc.
Decision Tree Algorithm:

- **Step-1:** According to S, the whole nipah virus dataset may be found at tree's root node.
- **Step-2:** Use Attribute Selection Measure (ASM) to zero down on most useful characteristic of data collection..
- **Step-3:** Separate S in subsets containing candidate values for most desirable characteristics.
- **Step-4:** Best attribute should be generated as node of decision tree.
- **Step-5:** Construct more decision trees utilizing filtered datasets from step -3 in a recursive fashion. Keep going till you reach viral prediction stage wherein you're no longer categorize prediction-nodes, at which point you will have leaf node.

IV. SYSTEM ARCHITECTURE

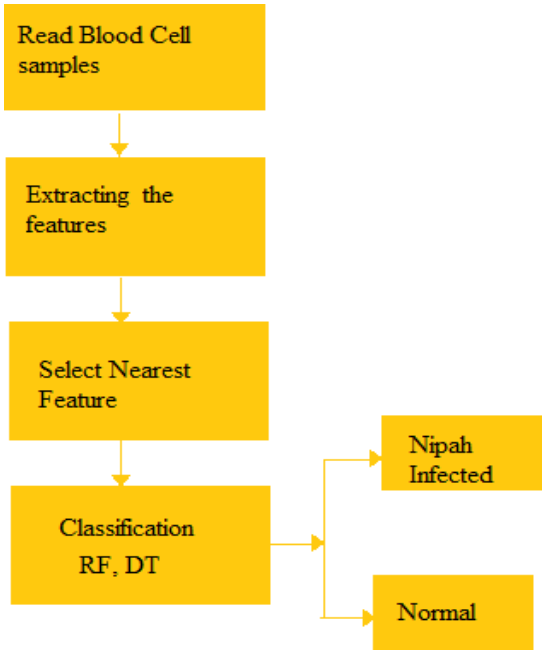


Figure 1: System Architecture.

Architecture Description: In the above architecture diagram, the system accepts the blood samples. Extracts the features, then applies classification such as RF and DT. Then the system predicts whether the patient is nipah virus infected or not.

V. RESULTS AND DISCUSSIONS

Fractal dimension and Shannon entropy were used to analyze the Nipah virus's glycoprotein and nucleoprotein sequences. The phylogenetic analysis is based on the variation in the atomic number of nucleotides. The major findings of conventional phylogeny analysis are reflected in the classification, although the classification is better at distinguishing between closely related strains. Glycoprotein sequence GC pair content as a function of fractal dimension.Differential development of cell& di-nucleotide entropy in nucleoprotein. Extrapolating from Nipah & Spanish flu-like viruses, low fractal dimension in nucleoprotein sequence may serve as a diagnostic indicator of these pathogens. There are two main flaws in Shannon's method. To begin, it can't be used to evaluate how various scales of diversity distributions stack up against one another. Second, it isn't able to evaluate subsets of diversity distributions against the whole set.
To overcome this problem purpose we are using ML algorithms in our system.

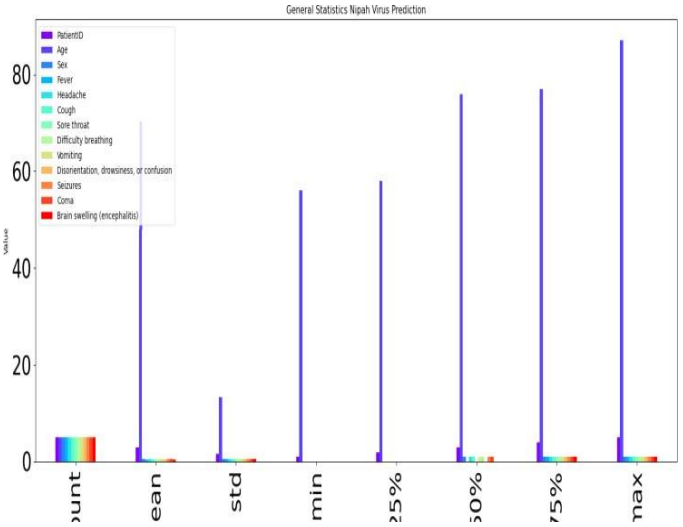
VI. IMPLEMENTATIONS

	PatientID	Age	Sex	...	Seizures	Coma	Brain swelling (encephalitis)
0	1	56	1	...	1	1	1
1	2	77	1	...	0	1	1
2	3	58	0	...	0	0	0
3	4	76	0	...	1	0	0
4	5	87	1	...	1	1	0

[5 rows x 13 columns]

Figure 2: Nipahvirus dataset

The patient dataset is collected in terms of patient ID, age,sex…



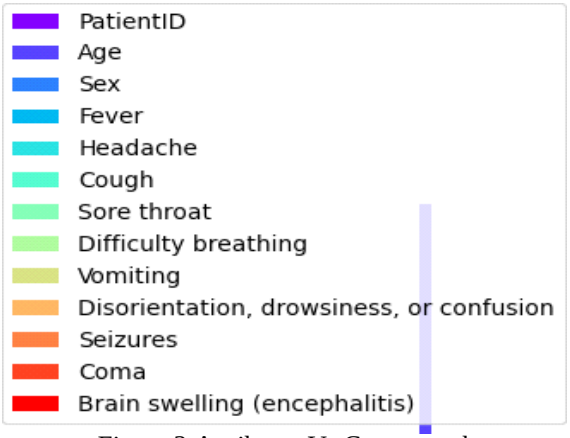


Figure 3:Attributes Vs Count graph

Statistics analysis of nipah virus attributes

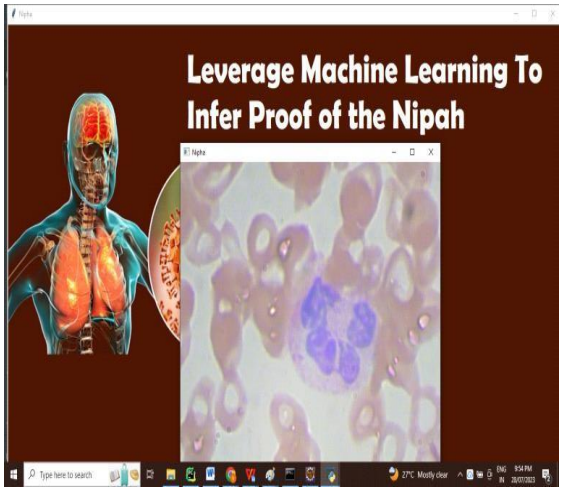


Figure 4: Blood Cell Sample

Patient blood cell sample is taken for testing purpose

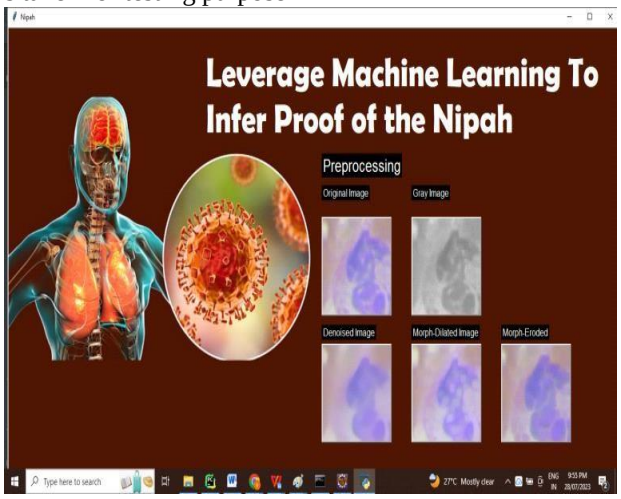
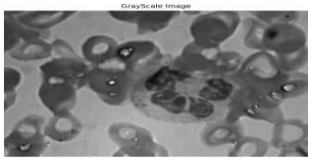


Figure 5: Preprocessing

It performs preprocessing which converts the image into grayscale and using morphological technique denoise the image



(a) Grayscale Image

(b) Threshold Image

(c) Segmented Image

Figure 6: Segmentation is process of dividing a picture into smaller pieces, called regions or segments.

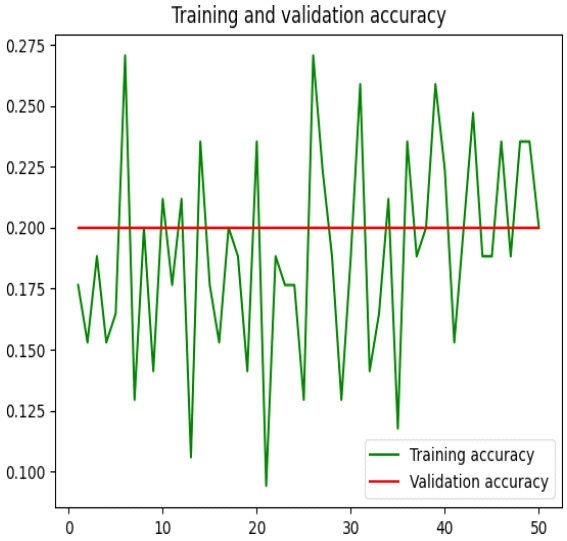


Figure 7:Training and Validation Accuracy

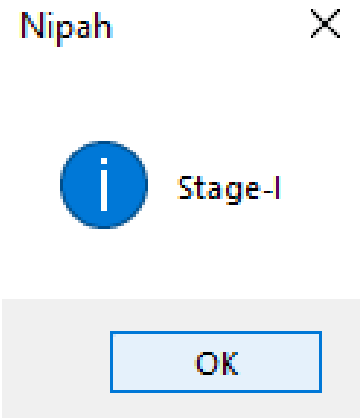


Figure 8: Prediction

(d) CONCLUSION

Date palm tree sap is the primary vector for the transmission of Nipah virus from bats to people. If more people knew about these issues, maybe the focus would move from treatment to prevention. Random forest and Decision tree are two machine learning approaches utilized to improve the accuracy of the models in this research. The model may be improved in the future by using IoT.

REFERENCES

1. L.J.S. Allen, V.L. Brown, C.B. Jonsson, S.L. Klein, S.M. Laverty, K. Magwedere, J.C. Owen and P. Van Den Driessche, ―Mathematical modeling of viral zoonoses in wildlife‖, Natural Resource Modeling, vol. 25, no. 1, pp. 5-51, 2012.
2. M.H.A. Biswas, ―Optimal control of Nipah virus (NiV) infections: a Bangladesh scenario‖, Journal of Pure and Applied Mathematics: Advances and Applications, vol. 12, no. 1, pp. 77-104, 2014.
3. Dr. Rekha J Patil, Indira Mulage and Nishant Patil 2023. Smart Agriculture Using IoT and Machine Learning. Journal of Scientific Research and Technology. 1, 3 (Jun. 2023), 47–59.
4. Computer Engineering, University of Porto, Portugal on June 28–29, 2012.
5. M.H.A. Biswas, L.T. Paiva and M.D.R. de Pinho, ―A SEIR model for control of infectious diseases with constraints‖, Mathematical Biosciences and Engineering, vol. 11, no. 4, pp. 761-784, 2014.
6. Naveen Kumar, Hanumanth, Nagareddy and Jyothi Patil 2023. Flight Delay Prediction Based On Aviation Big Data And Machine Learning. Journal of Scientific Research and Technology. 1, 7 (Oct. 2023), 58–67.
7. Parimala 2023. Machine Learning Technique For Prediction Of Wind And Rainfall Using Underwater Measurement. Journal of Scientific Research and Technology. 1, 6 (Sep. 2023), 124–135. DOI:https://doi.org/10.5281/zenodo.8330249.
8. K.B. Chua, ―Nipah virus outbreak in Malaysia‖, J. Clin. Virol. Apr., vol. 26, no. 3, pp. 265-275, 2003.
9. Dr. Raafiya Gulmeher and Umama Aiman 2023. A Novel Approach To Unveiling Employee Attrition Patterns using Machine Learning Algorithms. Journal of Scientific Research and Technology. 1, 6 (Sep. 2023), 234–241.
10. R. Breban, J.M. Drake, D.E. Stallknecht and P. Rohani, ―The role of environmental transmission in recurrent avian influenza epidemics, PLoS Comput. Biol., vol. 5, no. 4, pp. 1-11, 2009.
11. W.H. Fleming and R.W. Rishel, ― Deterministic and stochastic optimal control. applications of mathematics‖, no. 1, Springer Verlag, New York, 1975.
12. Shubam Sharma and Prof. Vinod Sharma 2023. Comparison of machine learning techniques in the diagnosis of erythematous squamous disease. Journal of Scientific Research and Technology. 1, 4 (Jul. 2023), 1–9.
13. K.B. Chua, W.J. Bellini, P.A. Rota et al., ―Nipah virus: a recently emergent deadly paramyxovirus, Science‖, vol. 288, no. 5470, pp. 1432- 5, 2000.
14. T.W. Geisbert, K.M. Daddario-DiCaprio, A.C. Hickey, M.A. Smith, Y.P. Chan et al., ―Development of an acute and highly pathogenic nonhuman primate model of Nipah virus infection‖, PLoS ONE, vol. 5, no. 5, pp. 1-12, 2010.