



Predictive Analysis of Genetic Disease Haemophilia-A based on Machine Learning Classification Algorithm

Dillip Narayan Sahu¹, Vijay Pal Singh^{2*}

¹Lecturer, Department of MCA, School of Computer Science, Gangadhar Meher University (GMU), Odisha, India

²Associate Professor, Department of Computer Science, OPJS University, Rajasthan, India

*Corresponding Author: Dr. Vijay Pal Singh

Abstract: India with nearly two lakh cases is estimated to possess the second highest number of patients with haemophilia, a lifelong bleeding disorder that forestalls blood from clotting. Hemophilia-A affects 1 in 5,000 male births within the U.S., and approximately 400 babies are born with hemophilia annually. Around 4 lakh people worldwide live with hemophilia. Hemophilia occurs in about 1 of each 5,000 male births. Supported recent study that used data collected on patients receiving care in federally funded hemophilia treatment centers during the amount 2012-2018, about 20,000 as many as 33,000 males within the us live with the disorder. It is very difficult to cure this kind of disease but can be handle with early diagnosis and proper treatment. The purpose of this paper is to establish some predictive models using Machine Learning algorithms by taking a real time Haemophilia dataset. In this paper, we have shown some real-time experiments and observations with the help of some Machine Learning algorithms, and also shown a clear picture on the predictive analysis on medical diagnosis of the Haemophilia using Machine Learning algorithms using which patients may get accurate data so as to diagnose better for their early treatment.

Keywords: Algorithm, Classifier, Haemophilia, Machine Learning, Prediction.

I.INTRODUCTION

Hemophilia is an uncommon disorder in the course of which the blood doesn't clot properly way to a deficiency in blood-clotting proteins. In case you've were given haemophilia, you will bleed for extended intervals of some time following a twist of fate than in case your blood clots generally[1][2]. Small cuts are usually now not an full-size deal. in case you've got a severe case of the disorder, the exceptional problem is internal bleeding, particularly on your knees, ankles, and elbows. Internal bleeding can harm your organs and tissues, placing your life in jeopardy. Hemophilia is generally always resulting from a mutation. The diminished coagulation component is replaced on a daily basis as a part of the remedy. Newer healing procedures are now being used that don't involve clotting elements[3][4].

II.LITERATURE SURVEY

Analysis

Many humans who've haemophilia or have had relations with the disease could request that their baby boys be tested as quickly as possible after delivery. Approximately one-0.33 of infants with haemophilia have a very specific mutation that nobody else in their own family has. If a new child exhibits positive haemophilia symptoms, a physician might also behavior a haemophilia take a look at. Doctors could use blood exams to envision if the blood became clotting properly to shape a prognosis. If it does not, they'll perform coagulation element testing, also known as element assays, to training session what is inflicting the bleeding. these blood exams would display the kind and degree of haemophilia[5][6].

Causes

When an individual bleeds, the body usually gathers blood cells into a clot to prevent the bleeding. Clotting factors are proteins within the blood that help platelets form clots by interacting with them. When a coagulation factor is lacking or the amount of the coagulation factor are low, haemophilia develops. Hemophilia A may be a quite haemophilia that's present at birth. Hemophilia is usually inherited, which suggests that an individual is born with it (congenital)[7]. Low coagulation factor type is employed to classify congenital haemophilia. Hemophilia A is that the most frequent kind, which is linked to a coffee level of factor 8. hemophilia B, which is related to a coffee level of factor 9, is that the next most frequent form. Hemophilia A may be a quite haemophilia that's acquired. Hemophilia can develop in people that haven't any case history of the disease. Acquired haemophilia is that the medical term for this condition[8].



Acquired haemophilia may be a sort of haemophilia that develops when the system targets coagulation factor 8 or 9 within the blood.

Prenatal (earlier than start) checking out with amniocentesis or villus sample are often finished if haemophilia runs in the own family. when the toddler is born, a sample of blood from the duct is generally examined. Inside the first six months of existence, best multiple infants are diagnosed with haemophilia. That is frequently thanks to the very reality that they're not likely to maintain an injury that would result in bleeding[9][10]. After a circumcision, there could also be bleeding, that could reason aprognosis. If a little one bruises without problems and bleeds excessively whilst wounded as they grow antique and more energetic, a physician can also don't forget haemophilia. Hemophilia is caused by a mutation or trade in a single a number of the genes that provides commands on the way to make the coagulation component proteins that help grume . this transformation or mutation can motive the clotting protein to prevent operating or move lacking totally. The X chromosome consists of those genes.

Remedy

Hemophilia is best treated by replacing the missing blood coagulation factor, which allows the blood to clot normally. Commercially produced factor concentrations are infused (administered through a vein) to realize this. People with haemophilia can learn to administer these infusions on their own to halt bleeding episodes and, by doing so on a daily basis (called prophylaxis), even prevent most bleeding episodes. High-quality medical aid from doctors and nurses who are well-versed within the disease can help avert significant complications[11]. Visiting a comprehensive Hemophilia Treatment Center is usually the simplest option for treatment (HTC). HTCs not only provide look after all aspects of haemophilia, but also give health education to assist persons with the illness stay healthy[12].

Prevention

Patient education helps to avoid morbidity and death linked with acute bleeding, also as screening for carrier moms and affected families[13]. Medical emergency identification bracelets or similar devices can assist in quickly identifying victims within the event of a hemorrhage/trauma, for instance.

Who is suffering from This

Hemophilia affects roughly one out of each 5,000 males born. consistent with a replacement study supported data collected on patients getting care in federally supported haemophilia treatment centres from 2012 to 2018, about 20,000 to 33,000 boys within the us have haemophilia. hemophilia A is fourfold more common than hemophilia B , and almost half those that have it have the severe variant. Hemophilia may be a blood disease that affects people of all races and ethnicities[14].

Machine Learning algorithms can be a better tool to predict early diagnosis of this disease.

III.EXPERIMENTS AND OBSERVATIONS1

ZeroR Classifier Output === Run information ===

Scheme: weka.classifiers.rules.ZeroR

Relation: haemophilia

Instances: 3772

Attributes: 30

ZeroR predicts class value: negative

=== Summary ===

Correctly Classified Instances	3481	92.2853 %
--------------------------------	------	-----------

Incorrectly Classified Instances	291	7.7147 %
----------------------------------	-----	----------

Kappa statistic	0
-----------------	---

Mean absolute error	0.0729
---------------------	--------

Root mean squared error	0.1904
-------------------------	--------

Relative absolute error	100 %
-------------------------	-------

Root relative squared error	100 %
-----------------------------	-------

Total Number of Instances	3772
---------------------------	------

IV.EXPERIMENTS AND OBSERVATIONS2

Bayes Classifier Output=== Run information ===

Scheme: weka.classifiers.bayes.BayesNet -D -Q weka.classifiers.bayes.net.search.local.K2 -- -P 1 -S BAYES -E weka.classifiers.bayes.net.estimate.SimpleEstimator -- -A 0.5

Class(4):

LogScore Bayes: -29286.076203253557



LogScore BDeu: -29611.200986222455
 LogScore MDL: -29816.503686177814
 LogScore ENTROPY: -29178.26323628693
 LogScore AIC: -29333.26323628693

V.EXPERIMENTS AND OBSERVATIONS3

NaiveBayes Classifier Output==== Run information ===

Scheme: weka.classifiers.bayes.NaiveBayes

Test mode: 10-fold cross-validation

==== Classifier model (full training set) ===

age				
mean	51.8081	52.324	50.3348	41.9457
std. dev.	20.3449	19.5296	18.7207	2.4674
weight sum	3480	194	95	2
precision	4.9348	4.9348	4.9348	4.9348
sex				
F	2266.0	146.0	70.0	2.0
M	1078.0	43.0	23.0	2.0
[total]	3344.0	189.0	93.0	4.0

==== Stratified cross-validation ===== Summary ===

Correctly Classified Instances	3594	95.281 %
Incorrectly Classified Instances	178	4.719 %
Kappa statistic	0.6008	
Mean absolute error	0.0357	
Root mean squared error	0.1382	
Relative absolute error	48.9161 %	
Root relative squared error	72.5471 %	
Total Number of Instances	3772	

VI.EXPERIMENTS AND OBSERVATIONS4

Bagging Classifier Output==== Run information ===

Scheme: weka.classifiers.meta.Bagging -P 100 -S 1 -num-slots 1 -I 10 -W weka.classifiers.trees.REPTree -- -M 2 -V 0.001 -N 3 -S 1 -L -1 -I 0.0

Test mode: 10-fold cross-validation

Bagging with 10 iterations and base learner

weka.classifiers.trees.REPTree -M 2 -V 0.001 -N 3 -S 1 -L -1 -I 0.0

==== Stratified cross-validation ===== Summary ===

Correctly Classified Instances	3754	99.5228 %
Incorrectly Classified Instances	18	0.4772 %
Kappa statistic	0.9675	
Mean absolute error	0.0049	
Root mean squared error	0.0451	
Relative absolute error	6.7318 %	
Root relative squared error	23.6792 %	
Total Number of Instances	3772	

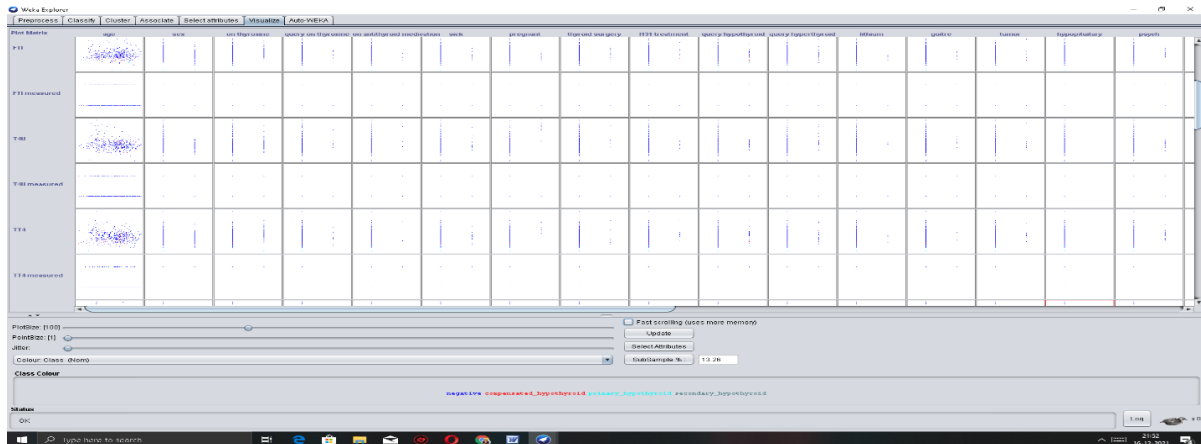


Fig.1. Visualization of graph in different cases

VII.CONCLUSION

We used the machine learning technique to assess, detect, and predict the haemophilia disease in four independent experimental data. Based on the findings of the preceding experiments, machine learning technologies are without a doubt an excellent way to forecast and detect haemophilia disease at an early stage, before the first stage patient's conditions become unsatisfactory. It has been discovered that applying different algorithms in Machine Learning to increase accuracy is a great way to detect and forecast haemophilia disease. Machine Learning algorithms with a high accuracy rate are efficient and acceptable. We used observations to determine the acceptability of a certain domain within the machine learning model using five distinct machine learning classifier methods. The accuracy level using the machine learning classification model Bagging is extremely much satisfactory, having a good accuracy rate of 92.35 percent then are going to be an honest option within the field of medical sciences to predict early diagnosis of haemophilia diseases, according to the study of the above real-time medical dataset implementation and several observations.

REFERENCES

- [1] Hoffbrand, A. V., Higgs, D. R., Keeling, D. & Mehta, A. B. Postgraduate Haematology. Seventh edn. (John Wiley and Sons, Inc., 2016).
- [2] Lee, C. A., Berntorp, E. & Hoots, K. Textbook of Hemophilia. Third edn. (John Wiley & Sons, Ltd., 2014).
- [3] Peters, R. & Harris, T. Advances and innovations in haemophilia treatment. *Nat. Rev. Drug Discov.* 17, 493–508 (2018).
- [4] Kitazawa, T. et al. A bispecific antibody to factors IXa and X restores factor VIII hemostatic activity in a hemophilia A model. *Nat. Med.* 18, 1570–1574 (2012).
- [5] Schiavoni, M. et al. Status of recombinant factor VIII concentrate treatment for hemophilia a in Italy: characteristics and clinical benefits. *Front. Med.* 6, 261 (2019).
- [6] Lenting, P. J., Denis, C. V. & Christophe, O. D. Emicizumab, a bispecific antibody recognizing coagulation factors IX and X: how does it actually compare to factor VIII? *Blood* 130, 2463–2468 (2017).
- [7] Doss, C. G. In silico profiling of deleterious amino acid substitutions of potential pathological importance in haemophilia A and haemophilia B. *J. Biomed. Sci.* 19, 30 (2012).
- [8] Gyulkhandanyan, A. et al. Analysis of protein missense alterations by combining sequence- and structure-based methods. *Mol. Genet. Genom. Med.* 8, e1166 (2020).
- [9] Hamasaki-Katagiri, N. et al. A gene-specific method for predicting hemophilia-causing point mutations. *J. Mol. Biol.* 425, 4023–4033 (2013).
- [10] Markoff, A., Gerke, V. & Bogdanova, N. Combined homology modelling and evolutionary significance evaluation of missense mutations in blood clotting factor VIII to highlight aspects of structure and function. *Haemophilia* 15, 932–941 (2009).
- [11] Sengupta, M. et al. In silico analyses of missense mutations in coagulation factor VIII: identification of severity determinants of haemophilia A. *Haemophilia* 21, 662–669 (2015).
- [12] Singh, V. K., Maurya, N. S., Mani, A. & Yadav, R. S. Machine learning method using position-specific mutation based classification outperforms one hot coding for disease severity prediction in haemophilia 'A'. *Genomics* 112, 5122–5128 (2020).
- [13] Shen, B. W. et al. The tertiary structure and domain organization of coagulation factor VIII. *Blood* 111, 1240–1247 (2008).
- [14] Smith, I. W. et al. The 3.2 Å structure of a bioengineered variant of blood coagulation factor VIII indicates two conformations of the C2 domain. *J. Thromb. Haemost.* 18, 57–69 (2020).
- [15] Nisthal, A., Wang, C. Y., Ary, M. L. & Mayo, S. L. Protein stability engineering insights revealed by domain-wide comprehensive mutagenesis. *Proc. Natl Acad. Sci.* 116, 16367–16377 (2019).
- [16] Ahmed, R. P. et al. Identification of 32 novel mutations in the factor VIII gene in Indian patients with hemophilia A. *Haematologica* 90, 283–284 (2005).
- [17] Higuchi, M. et al. Molecular characterization of mild-to-moderate hemophilia A: detection of the mutation in 25 of 29 patients by denaturing gradient gel electrophoresis. *Proc. Natl Acad. Sci. USA* 88, 8307–8311 (1991).
- [18] Liu, M., Murphy, M. E. & Thompson, A. R. A domain mutations in 65 haemophilia A families and molecular modelling of dysfunctional factor VIII proteins. *Br. J. Haematol.* 103, 1051–1060 (1998).
- [19] Kessel, A. & Ben-Tal, N. Introduction to Proteins: Structure, Function, and Motion. (CRC Press, 2010).
- [20] Breiman, L. Random forests. *Mach. Learn.* 45, 5–32 (2001).
- [21] He, H., Bai, Y., Garcia, E. A. & Li, S. in 2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence). 1322–1328 (IEEE).
- [22] Dieterich, T. G. in Proceedings of the First International Workshop on Multiple Classifier Systems 1–15 (Springer-Verlag, 2000).