ISSN: 2366-1313

STATISTICAL MODEL IN MACHINE LEARNING TO PREDICTSEVERE RETINOPATHY OF PREMATURITY

G. Pavani **B. Jyothi** **Dr. M. Sandeep**
Assistant professor Assistant professor Associate professor
Dept of EEE
Sree Dattha Intitute of Engineering and Science

ABSTRACT
Using statistical analysis and logistic regression as a type of Generalized Additive Model (GAM) with pair-wise interaction terms (GA2M), we investigated the risk factors for severe retinopathy of prematurity. We talk about the trade-off between these machine learning techniques' accuracy and interpretability on clinical data in this process. We also confirm expert neonatologists' instincts regarding a few risk factors, such as gender, that were previously thought to be clinically insignificant in RoP prediction. In many middle-income countries, retinopathy of Prematurity (ROP), a vasoproliferative disorder of the immature retina in premature infants, is a significant cause of blindness. In high-income nations, risk factors like oxygen administration and blood oxygen saturationare closely monitored, resulting in a lower incidence of ROP.1 In developed nations, severe ROP is typically found in infants born at a very early Gestational Age (GA). Because there is insufficient awareness of the disease process's risk factors, a lack of skilled professionals, and/or a lack of essentialequipment for infant care, heavier, more mature babies can also develop ROP in developing nations.

KEYWORDS: interpretability of machine learning in healthcare, generalized additive model, logistic regression Retinopathy of Prematurity (RoP), neonatology

I INTRODUCTION

A specific type of artificial intelligence knownas machine learning enables systems to learn from data and recognize patterns without muchhuman intervention. Computers using machinelearning are shown patterns and data, allowing them to make their own decisions rather than being instructed. Algorithms for machinelearning serve a variety of purposes, includingassisting in the filtering of email, identifying objects in images, and analyzing large quantities of increasingly complex data sets. Machine learning systems are used by computers to automatically sort through emails, identify spam, recognize objects in pictures, and process large amounts of data. Techniques for machine learning are a growing area of study with numerous potential

ISSN: 2366-1313

applications. For healthcare professionals and health systems, machine learning technology will become increasingly important as patient data becomes more readily available formeaning extraction. Machine learning is especially useful in the healthcare industry because it can help us make sense of the enormous amounts of healthcare data that are generated each day in electronic health records.We can use machine learning in healthcare in amanner that is analogous to how algorithms formachine learning can assist us in discovering.patterns and insights that cannot be discoveredmanually. With the widespread acceptance of machine learning in healthcare, healthcare providers now have the opportunity to implement a more predictive strategy that results in the development of a more unified system with enhanced care delivery andpatient-centered processes.

The most common uses of machine learning inthe healthcare sector are to automate medical billing, provide clinical decision support, and develop clinical practice guidelines within health systems. Machine learning and healthcare concepts are used in numerous notable high-level applications in science and medicine. Data scientists at MD Anderson developed the first deep learning healthcare algorithm for predicting acute toxicities in patients receiving radiation therapy for head and neck cancers. Deep learning data can automatically identify intricate patterns in healthcare data and provide primary care providers with clinical decision support at the point of care within the electronic health record.

For the purposes of machine learning, large volumes of unstructured healthcare data account for nearly 80% of the data stored or "locked" in electronic health record systems. These are absolutely not data elements; rather, they are pertinent text files or data documents that contain patient data that, in the past, couldnot be analyzed without a human reviewing themedical records. Human language, also knownas "natural language," is extremely complicated, inconsistent, and full of jargon, ambiguity, and ambiguity. In the field of healthcare, machine learning frequently makes use of natural language processing (NLP) software to transform these documents intodata that is easier to use and analyze. Machinelearning relies on healthcare data for the majority of NLP-based deep learning applications.

Recent trends in machine learning applicationsin healthcare center on the interpretability of the models. In some healthcare applications, interpretability of the model is more importantthan accuracy, but there are many situations inwhich interpretability is preferred despite the loss of accuracy. One way to achieve the idealsituation in which the model has both high accuracy and high interpretability is to start with a simple model like a generalized additive model (GAM) and then make it more complex(and thus more accurate) like a GA2M (GAM with pairwise interactions), or to start with a complex model like XGBoost and try to interpret it locally with methods like LIME. Ann Lecun,

ISSN: 2366-1313

Facebook's head of AI research, and Rich Caruana, Microsoft's lead machine learning researcher on healthcare applications, organized a debate session at Neur IPS (NeuralInformation Processing Systems), one of the most significant conferences on machine learning. During this session at the "Interpretable ML Symposium's" conclusion, the need for additional case studies on theaccuracy versus interpretability trade-off in machine learning for healthcare was highlighted. 1][9][10] In this paper, we will examine a case study regarding Severe Retinopathy of Prematurity (Severe RoP), which, if left untreated, can result in newbornsgoing completely or partially blind. Stevie Wonders was a well-known musician who hada lot of RoP. Doctors only recently realized that oxygen could save premature infants in the 1950s. However, it took them a few more yearsto realize that excessive oxygen caused the vessels in the eyes to grow out of place.

Premature babies would go blind for life if thiscondition was treated with too much oxygen. The World Health Organization (WHO) has estimated that 15 million of the 130 million babies born each year are born prematurely, orbefore the 37th week of gestation.

Each year, approximately one million childrendie because of preterm birth complications. Among survivors, learning disabilities as well as visual and auditory impairments are common disabilities. Among these issues, retinopathy of prematurity (RoP) is the leading and most serious cause of disability. Retinopathy of Prematurity (RoP) was first described by Terry [11] as a developmental, vascular, and proliferative retinal disorder thataffects the retinas of premature newborns that have not yet undergone vascularization.

Along with cortical blindness, RoP is one of themost common causes of childhood blindness worldwide. RoP has been linked to a number ofrisk factors, and if severe RoP is not treated, itcan cause retinal detachment blindness. With alower birth weight and shorter gestational weeks, RoP becomes more severe. To stop thedisease from progressing, early diagnosis andappropriate treatment are essential. The International Classification for Retinopathy ofPrematurity (ICROP) is used to categorize the condition based on its progression and severity. The Newborn Clinic at Istanbul's Zeynep Kamil Woman and Child Diseases Hospital collected the clinical data for our analysis between 2011 and 2014. The diagnosis of retinal detachment, which is the most severe repercussion of RoP, serves as the starting point for classification. Every year, around 150000 babies in Turkey have birthweights below1500 grams. As we improve our prediction model, we will look at how the accuracy of ourinterpretable model improves as we combine numerical and categorical values and add additional interaction terms. We will also investigate the risk factors that are thought to cause or be correlated with severe RoP. It is more likely that these newborns will be diagnosed with severe RoP.

ISSN: 2366-1313

II. PROPOSED METHODOLOGY

Our research begins with data on retinopathy of premature (RoP) from Zeynep Kamil Maternity and Child Diseases Hospital from 2011 to 2014. It is a record of 102 variables across more than 5000 patients. ZK- RoP) For patients with birth weights below 1500 grams, we were interested in the same 20risk factors examined in the TR-RoP study [4]. We utilized their data for 1066 patients who weighed less than 2000 grams because the ZK-RoP study focused on newborns under 2000 grams. Among the 1066 newborns with birthweights below 2000 grams, 109 cases of severe RoP were examined. Our sample size was reduced to 385 because the important task was to predict who might develop severe ROPfrom those who had already been diagnosed with any type of RoP. We ran univariate and multivariate logistic regression machine learning algorithms, beginning with the generalized additive model (GAM) to predict severe RoP based on the same risk factors as inthe TR-ROP study [4]. We also added interaction terms to our multivariate analysis inorder to minimize type II error and minimize wrong diagnosis of a disease that could result in blindness. Out of 385 patients who were diagnosed with any type of RoP, we attemptedto construct a model that would predict the 109patients who we wanted to use the same method as the TR-RoP paper and check their results with a new set of data, ZK-RoP. We alsowanted to include accuracy analysis of these predictions, which the TR-RoP study didn't dobecause it wasn't geared toward ML users.We could use a number of ML algorithms, but we wanted to start with the simplest and move up to the more complex ones. We only focused on the multivariate logistic regression analysis in this paper. We will analyze with more complicated models like decision trees, k-means, and xgboost in further research.

2.1 Generalized Additive Model with Pairwise Interactions(GA2M):

GAM is short for generalized additive model. Before we discuss GAMs, let's first briefly review a common statistical model that you arelikely to be familiar with. It is the generalized linear model (GLM). A GLM is a very popular and flexible extension of the classical linear regression model. It enables youto model a target (or response) variable that is not normally distributed.

For example, if you are interested in predicting whether an incoming email is spam, or you wantto predict the number of people dining in a restaurant, the linear regression model is inappropriate. This is because the target variables for those applications violate the model's normality assumption. However, a GLM can handle both data types by assuming adistribution in the exponential

ISSN: 2366-1313

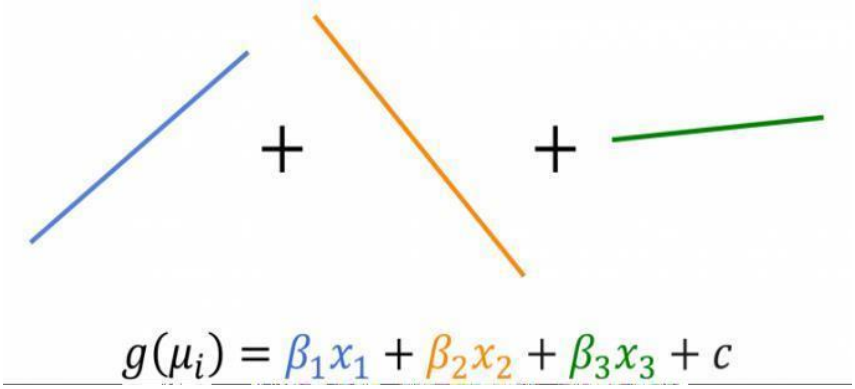

Fig 1: Linear models assume each componentfunction is linear. Source: Adapted from Servén Marín (2018)

For GAMs, on the other hand, the component functions can be arbitrary nonlinear smoothing functions that the model usually estimates by using spline functions. You can see this in Figure 2.

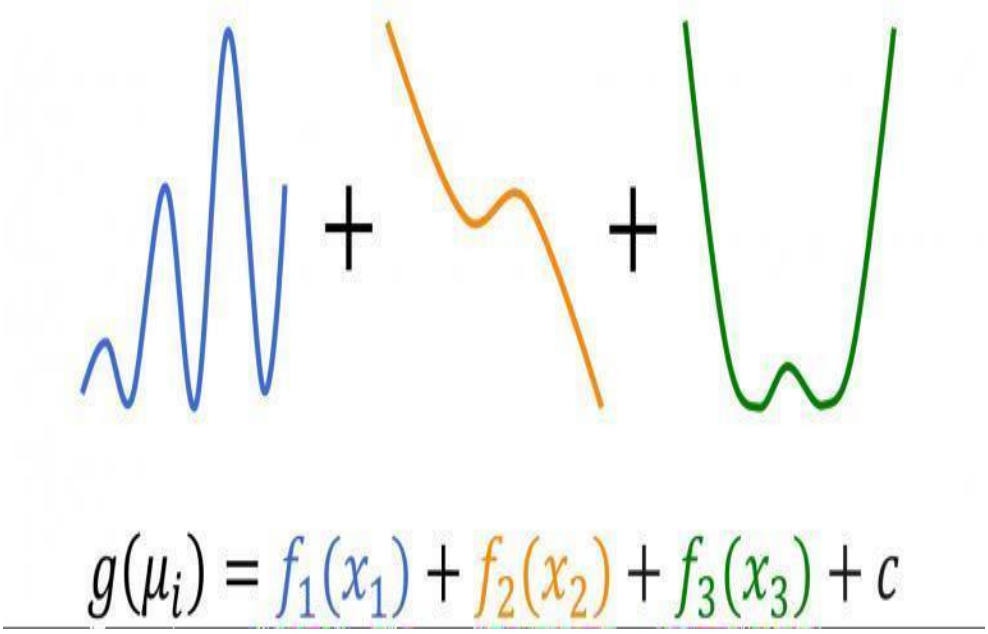

Fig 2: GAMs use smoothing componentfunctions to flexibly model nonlinearities.

Source: Adapted from Servén Marín (2018) Even with this flexibility, GAMs are stillinterpretable. This is an increasingly desirablefeature of machine learning models. This isespecially true for high-stakes applications suchas health care, finance, and mission-criticalsystems.Typically, you face a trade-off betweenmodel performance and model interpretability.

ISSN: 2366-1313

For the effects in the final model, the node's results include smoothing component plots for the spline terms and parameter estimates for the parametric terms. For example,the results include a smoothing component plotfor the Spline(Debtinc) term. Figure 3 displaysthis. Interpreting the results, you can see that theprobability of default is generally higher for an applicant who has a high debt-to-income ratio and the relationship is nonlinear.

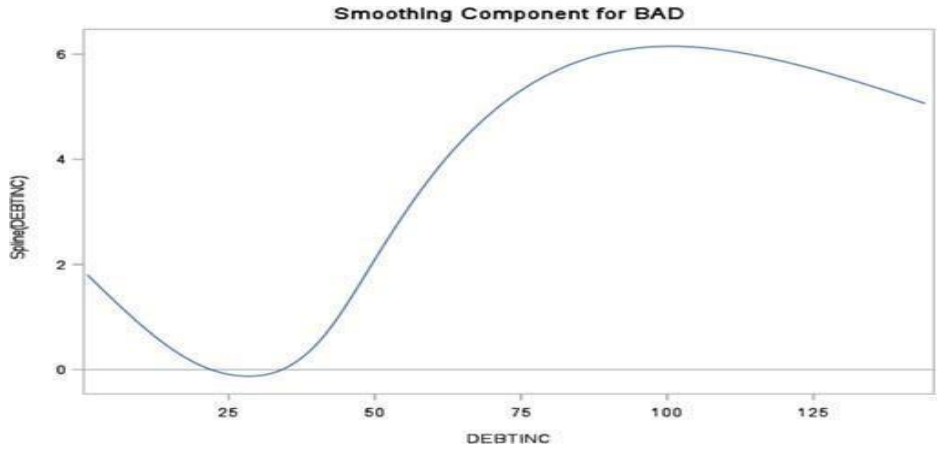

Fig 3: The probability of default is generally higher for higher debt-to-income ratios.

In addition, the ability to interpret the parametriceffects like you would with a logistic regressionmodel is another way in which the GAM's results are relatively easy to understand. For example, an applicant with more delinquent credit lines has a higher predicted probability ofdefault, on average. See Figure 4.

Effect	Parameter ↑	Estimate
DELINQ	DELINQ 0	-4.4301
DELINQ	DELINQ 1	-4.0737
DELINQ	DELINQ 2	-3.5835
DELINQ	DELINQ 3	-3.5960
DELINQ	DELINQ 4	-2.6697
DELINQ	DELINQ 5	-1.0057
DELINQ	DELINQ 6	-0.8645
DELINQ	DELINQ 7	-0.4681
DELINQ	DELINQ 8	-0.2449

ISSN: 2366-1313

Fig 4: More delinquent credit lines typicallycorrespond to a higher probability of default.

. With Model Studio, you can also easily trainother supervised learning models orcomparison.Let's compare the GAM to gradient boosting, logistic regression, neuralnetwork, and support vector machine (SVM)models. If you rank the algorithms by misclassification rate, you see that the GAM ranks second. Figure 5 displays the results.

Algorithm Name	Misclassification Rate ↑
Gradient Boosting	0.0481
GAM	0.0571
Logistic Regression	0.0586
SVM	0.0662
Neural Network	0.0782

Fig 5: The GAM outperforms some of the morecomplex models but maintains interpretability.

Even though the GAM has a slightly higher misclassification rate, it is more nterpretable than the

Fig 4: More delinquent credit lines typicallycorrespond to a higher probability of default.

With Model Studio, you can also easily trainother supervised learning models forcomparison.Let's compare the GAM togradient boosting, logistic regression, neuralnetwork, and support vector machine (SVM)models. If you rank the algorithms by misclassification rate, you see that the GAM ranks second. Figure 5 displays the results.

ISSN: 2366-1313

Algorithm Name	Misclassification Rate ↑
Gradient Boosting	0.0481
GAM	0.0571
Logistic Regression	0.0586
SVM	0.0662
Neural Network	0.0782

Fig 5: The GAM outperforms some of the morecomplex models but maintains interpretability.

Even though the GAM has a slightly higher misclassification rate, it is more interpretable than the champion gradient boosting mode. And it still outperforms other complex models such as SVM and neural network.

In this model we are trying to predict a categorical variable, whether the patient is diagnosed with severe RoP or not, we will use logistic regression instead of linear regression. Univariate logistic regression holds only one variable ln [Y / (1−Y)]=a + b 1 X 1 , while theformula for multivariate logistic regression is asfollows: ln[Y /(1− Y)]= a + b 1 X 1 + b 2 X 2+ b 3 X 3+... The generalized additive model (GAM) with only 1st order terms is equivalent to the equation above. If we want to add 2nd order terms, we would need to add them as follows: n[Y /(1− Y)]= a + b 1 X 1 + b 2 X 2 +b 3 X 3+ b 12 X 1 X 2 + b 13 X 1 X 3+ b 23 X2 X 3 +..or example, for a GAM with 6 first-order variables (X 1 ,X 2 ,X 3 ,X 4 ,X 5 ,X 6) , there would be 6*5/2=15 pairwise interaction terms.

It is suggested to add only a few and only start with the most influential pairwise interactions. Thus, as we make our model more complex, wewill add the pairwise interactions one by one.However, when we train a machine learning model, we typically face a tradeoff between accuracy and interpretability. GAMs provide a solution to this dilemma because they strike a nice balance between these two competing goals. With Model Studio, you can easily train GAMs alongside other common machine learning models to compare model performance,all without the need to write any code.

II CONCLUSION

In our research 12 risk factors are used in the multivariate linear regression: 10 binary(categorical) values and two numbers. First, weran the regression with numerical and categorical values separately. The multivariate logistic

ISSN: 2366-1313

regression analysis that would produce a binary prediction of severe RoP was thencarried out by combining the numerical andcategorical risk factors into a single equation. Last but not least, it is important to note that comparisons between GAM and GA2M andother machine learning methods that can beexplained, such as decision trees, require additional research. Perhaps our RoP predictionproblem and data were a better fit for other interpretable machine learning techniques givenour decision tree approach with cross-validationin SPSS, which produced a higher accuracy rateof 44% and only a few type II errors. Cross- validation on RoP data and various interpretableML techniques need to be compared in additional research.We hope to use Microsoft's interpretable ML library and the scikit-learn ML library as we expand our explainable ML work based on this paper. Both of these libraries promise high accuracy and high interpretability.

REFERENCES

[1] R. Caruana, Y. LeCun, The Great AI Debate "Interpretable ML Symposium" as part of NeurIPS- 2017. https://www.youtube.com/watch?v=93Xv8vJ2acI

[2] R. Caruana, Y. Lou, J. Gehrke, P. Koch, M.Sturm, and N. Elhadad. Intelligible Models forHealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission. In Proceedings ofthe 21th ACM SIGKDD InternationalConference on Knowledge Discovery and Data Mining, pages 1721–1730. ACM, 2015.

[3] O. Gottesman, F. Johansson, J. Meier, JDent, D. Lee, S. Srinivasan, L. Zhang, Y. Ding, D. Wihl, X. Peng, J. Yao, I. Lage, C. Mosch, L.. Lehman, M. Komorowski, A. Faisal, L. A. Celi, D. Sontag, and F. Doshi-Velez. Evaluating Reinforcement Learning Algorithms in Observational Health Settings. pp.1-16, 2018. https://arxiv.org/pdf/1805.12298.pdf

[4] A.Y. Bas, N. Demirel, E. Koc, D. Ulubas Isik, I.M. Hirfanoglu, T. Tunc, and TR-ROP Study Group. Incidence, risk factors and severity of retinopathy of prematurity in Turkey(TR-ROP study): a prospective, multicentre study in 69 neonatal intensive care units. Br J Ophthalmol. 102(12):1711-1716, 2018.

[5] Y. Lou, R. Caruana, J. Gehrke, and G. Hooker. Accurate Intelligible Models with Pairwise Interactions. KDD2013, August 11–14, 2013, Chicago, Illinois, USA. http://www.cs.cornell.edu/~yinlou/papers/lou-