Proceedings of the International Conference on Artificial Intelligence and Smart Systems (ICAIS-2021)
IEEE Xplore Part Number: CFP21OAB-ART; ISBN: 978-1-7281-9537-7

Convalescent Plasma Therapy: Data driven approach for finding the Best Plasma Donors

M N Noorshidha
Department of Computer Science
National Institute of Technology Puducherry
Karaikal, India
noorshidha@gmail.com

Dr.G.Aghila
Department of Computer Science
National Institute of Technology Puducherry
Karaikal, India
aghilaa@gmail.com

2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS) | 978-1-7281-9537-7/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICAIS50930.2021.9396012

***Abstract*—**Convalescent Plasma Therapy is an investigational therapeutic method recommended as a treatment strategy for COVID-19 as vaccines, and proper treatment methods were unavailable. The therapy involves transfusing antibody contained plasma from the COVID recovered individuals (donors) into critically affected patients. It can accelerate the recovery of the recipient. The effectiveness of antibodies is affected by the health and clinical history of donors, according to research. It implies the possibility of implementing Machine Learning Classification models for predicting the Eligible donors (who meet the threshold antibody level for donation) and Regression models to predict the antibody level value of a donor from the person's clinical history before conducting tests for the same. The proposed system can help the health authorities approach the most probably efficient donors for the therapy rather than wasting time and test kits on a random donor who may or may not be eligible. The results from various ML algorithms trained on a synthetic clinical history dataset are examined and assessed as significant to some degree. The system has to be validated against real data to arrive at reasonable conclusions. This paper demonstrates how a data-driven solution is more beneficial than the conventional methods for donor search.

Keywords—Convalescent Plasma therapy; Donor selection; ML classification algorithms; ML Regression algorithm.

I. INTRODUCTION

The world has been facing a medical crisis after an infectious virus got identified in December 2019 from Wuhan, China. The disease caused by this virus named COVID-19 was declared a pandemic by the World Health Organization (WHO) on 03 October 2020. As of 06 January 2021, 86.5 million confirmed cases and 1.87 million deaths have got reported worldwide. Apart from the general supportive care and mechanical ventilation for critical patients, no therapies or antiviral drugs were available. Although recently some vaccines have got developed, the majority of the countries are yet to avail themselves. However, researches and clinical trials for developing vaccines and medicines for COVID-19 are progressing worldwide. In this context, alternative investigational treatment methods like Convalescent Plasma (CP) therapy got employed in many countries.

The human body attacked by a virus produces proteins called antibodies in the blood to defend. These antibodies would stay in the blood for a certain period, fighting off the same virus if exposed to it again. The CP therapy can be briefed as a person who got recovered from a given virus (convalescent phase) will have enough antibodies that can be extracted and transfused into patients fighting the same disease. These antibodies would help the recipient's immune system to accelerate developing its antibodies. Hence, this therapy is called passive immunity.

Human blood constitutes 45% of blood cells (Red blood cells, white blood cells, and platelets) and 55 % of plasma. Plasma is the liquid yellowish part of the blood that contains the antibodies. The process of separating plasma from the blood is called Plasmapheresis. The process takes around two hours. Unlike the typical blood donation, plasma donation is a closed-loop system. The plasma is drawn from the donor into a machine using a centrifuge and then the plasma is separated from the blood. After collecting enough plasma, the blood returns to the body of the donor. The plasma would undergo many tests for identifying transmittable diseases before it gets transfused to an ill patient.

CP therapy can be traced back more than a century. It had been in use in past disease outbreaks like SARS, H1N1, H1N5, Ebola, etc. The studies show a statistically significant reduction in mortality rate, and the recovery was faster compared to non-therapy patients. Infusing a person with another one's plasma has some risk factors like uncertain rejections and allergic reactions that can lead to Multi-organ failure. That explains why Plasma Therapy is used only for critically ill patients with a lesser chance of survival. Hence, Plasma Therapy is the best option when no treatments are available.

A potential donor with consent can approach any of the plasma donation centers if he fulfills the eligibility criteria like (1) history of confirmed COVID positive test result (2) age between [18-55] (3) being symptomless for at least last 14 days etc. The donor will go through a screening process during which he gets inquired about his/her health history. Plasma donation selection criteria can be slightly modified according to local requirements and standards but should be in line with the World Health Organization (WHO) guidelines. Those who have fulfilled the initial criteria have to go through screening tests like (1) Antibody testing – to determine whether antibodies exist or not (2) Antibody titer test – to determine the presence and the antibody concentration in the body. The threshold antibody level for donation varies slightly in different countries.

In the initial period of the outbreak, there was no proper channel for inviting plasma donors. As the cases went high, the demand for plasma grew, and very few recovered patients were willing to donate. There were cases where the health authorities could not arrange a donor and hence the family was encouraged to do the job. People were in a desperate chase, trying every possible way to locate a donor. Even after finding a donor, chances are there for the donor to fail any of the eligibility criteria or not to have a threshold antibody level. Since, Plasma therapy is done only for critically ill patients, the above scenario can create stress and panic. An appropriate and efficient method is necessary for not just finding donors but the most efficient one. Studies and research show that the level of antibodies is influenced by many donor-related factors, such as the severity of the disease, age, and many more. It is not easy to set a rule-based system for the prediction of antibody level. Hence, the best and feasible way is to use data-driven methods.

The contribution of this research can be summarized as follows:

- An effort to mimic the data of plasma donors as the donor's clinical history data is not publicly available.
- Predict whether a person is Eligible for Plasma donation with a threshold antibody titer using ML Classification Algorithms.
- Predict the antibody titer/level value of donor using ML Regression Algorithms to discover the person with a higher level of antibody.
- Analyze different Classification and Regression algorithm results in the account of the donor selection task.

The rest of the paper is organized as follows. Section II is a Literature Survey of various researches in the context of CP therapy and ML. Section III explains how the donor data set is randomly created. In section IV we compare different ML algorithm results. Finally, in section V the paper is concluded.

II. Related Work

Fundamentals of convalescent Plasma therapy and its historical background [2] with a review of the recent Ebola outbreak are studied [3]. These papers have pointed out the advantages, uncertainties, and limitations of this therapy for COVID-19 or any other future similar situations indicating the employability. This therapy has got employed in many countries as an investigational therapy for COVID-19. A detailed review of a few such trials done so far studied by K. Rajendran et al. in [5] is analyzed. The paper has evaluated 5 CP therapy studies early in China: The Clinical details, Survival rate, Antibody titer changes after therapy, and limitations of the clinical trials gets examined in this paper. Another study in [6] concurs with it.

Eligibility criteria, Pre-screening tests, and plasma transfusion guidelines for CP therapy are discussed in [1]. The study is significant in determining who is eligible to donate. The study of antibody titer, and how it is impacted by various clinical factors form the ground of the proposed system. In [7], Xuemei Li et al. compare antibody response patterns and observes that IgG is the antibody that persists for a long period [8] and the level of IgG varies vigorously in severe and mild cases. According to Geoffrey J Gorse, Mary M Donovan, and Gira B Patel [9], the Coronavirus antibody is found increasing in older people than adults. At the same time, other factors such as height and weight don't have any significance in the antibody level variance [10]. The different serological tests available and interpreting the positive and negative values of these tests are described in [11].

All the above ideas are incorporated when a synthetic donor dataset is built using random number generation [12]. Blood glucose level prediction using data driven methods are attempted and implemented with better results by various researchers [16][17]. The task in hand of "predicting the best plasma donors in data driven method" is not found to have investigated in any previous researches. The domain of convalescent plasma therapy is quite less explored using data science. The major reason for this could be the unavailability of data. The idea of implementing classification and regression algorithms for a problem is demonstrated in [13] and the performance was measured using appropriate metrics chosen based on [14] and [15].

III. Proposed Work

Machine learning algorithms need data to learn. The algorithm would predict the outcome of a future data point based on the learning done. Since Convalescent Plasma therapy is an investigational therapy, the data is not publicly available. We need data about the clinical history of the donors (1) to predict his/her eligibility using a classification algorithm [A donor is considered eligible if he has a value above the threshold antibody titer value] (2) to predict the real value of antibody using a Regression Algorithm. We have attempted to mimic the required data of clinical history and screening test results.

A. Plasma Donor Dataset Feature Selection

Table I shows the details collected from a donor at a plasmapheresis center before the plasma donation. These details are straightforward enough to be considered as features for the Donor dataset generation. However, some manipulations are done for the sake of implementation simplicity. A few details like "Address" and Contact Details" from Table I are excluded due to their insignificance in the prediction task at hand. Similarly, other details like "Date of COVID-19 confirmation by PCR" and "Date of hospital admission" are also excluded even though these features may help in determining some useful information like days taken for the symptoms to appear. But it can be determined only if the patient is admitted to the hospital when symptoms first appear. In reality, the decision of admitting the patient to the hospital when tested positive or when symptoms appear are all specific to local health administrations. So, we cannot generalize any information based on these dates. "Symptoms" 11 of Table I can be fever, cough, tiredness, and many others in reality. Hence, it can be taken as separate independent Boolean valued features. Similarly, the features 12, 13,14, 15, 17, 18,

and 22 from Table 1 are all taken as Boolean valued. For simplifying the implementation, the following details are kept as Boolean values, although, in reality, they are multi-valued features: "Symptoms Others, Any other complication, Any comorbid illness" etc. Features 20 and 21 are also removed as there are multiple blood components and the blood itself can be transfused when the patient was sick. The impact of such transfusions on the patient's antibody level is unclear. The numbers 23 and 24 are also avoided but the Number of days since being symptomless is an important feature that can be calculated using these 2 features. Finally, the "COVID Ab Rapid Card test" is again a Boolean valued feature that says whether the person has an antibody in his plasma or not, and "COVID Ab titre" is the antibody titre/level test done for Immunoglobulin G (IgG) type antibody which is a real value. In total 22 features are considered for the final dataset generated as shown in Table II.

B. Plasma Donor Dataset Feature Creation

A dataset with samples of Plasma donor details having features as indicated in Table II is simulated. The values for each feature were generated randomly within a range of possible values. A few constraints are applied in the value generation based on the behaviour of antibody to these features as per [5][7][8][9][10]. Age is created between 18 and 55 with a mean value 35 randomly. Blood groups [O+/-, A+/-, B+/-, AB+/-] assigned with values 1 to 8 respectively are randomly generated with probabilities given as per the statistics of Blood Type distribution over the Asian population. In Table II Features 7 to 10 are all considered to be mild symptoms. A majority of the COVID cases are symptomless and hence each of these symptoms is given a 0.2 probability when randomly generated. A Similarly, 11 to 14 are considered critical but the probability of such critical condition is comparatively low.

TABLE I. PLASMA DONOR DETAILS COLLECTED AT THE PLASMAPHERISIS CENTRE

No	Donor Details	Description
1	Name	Name of the donor
2	age	As per guidelines [18 -55]
3	Sex	Sex (Male/Female)
4	Height	Height
5	Weight	Weight
6.	Blood group	8 common blood types are assigned
7	Address	Address of the donor
8	Contact Details	contact number and email address
9	Date of COVID-19 confirmation by PCR	Date on which the person was tested positive for COVID using swab test
10	Date of hospital admission	Date on which he is admitted to the hospital
11	Symptoms	Whether the person had any minor symptoms like fever, tiredness, cough, others etc.
12	Pneumonia	Whether the person had Pneumonia
13	Respiratory Failure	Whether the person had Respiratory Failure
14	Septic shock	Whether the person had got Septic shock
15	MODS	Whether the person had Multiple organ dysfunctionalities
16	No of days of ventilator support	Critical patients are provided with ventilator support
17	Any other complication	Any other complication
18	ICU admission	ICU admission denotes severity in sickness
19	No of days of ICU admission	No of days been in ICU
20	Blood component transfusion	If the person has got any kind of blood or blood component transfusions
21	No of units of blood transfused	Quantity of transfusion
22	Any co morbid illness	Whether the patient had any co-morbid illnesses
23	Date of complete resolution of symptoms	Date when the person went symptomless after being tested positive
24	Date of First and second Negative PCR	Two negative tests are required to be considered COVID-negative
25	COVID Ab Rapid Card test	Antibody test -Yes or No
26	COVID Ab titer	Antibody titer for IgG Antibody

TABLE II. PLASMA DONOR DATASET FEATURES AND DESCRIPTION

No	Donor Details	Type	Description
1	Name	string	P1, P2…; distinguish the datapoints
2	age	int	male range (18 to 55); female range (18 to32)
3	Sex	int	Male:0 and Female:1
4	Height	float	Height
5	Weight	float	Weight
6	Blood Type	int	Values from 1 to 8 given to the 8 common blood types
7	Symptom fever	bool	Fever=1; no fever=0
8	Symptom tiredness	bool	tiredness =1; no tiredness =0
9	Symptom cough	bool	cough =1; no cough =0
10	Symptom others	bool	other symptoms=1; no other symptoms=0
11	Pneumonia	bool	Pneumonia=1; no pneumonia =0
12	Respiratory failure	bool	Respiratory failure =1; no Respiratory failure =0
13	Septic shock	bool	Septic shock =1; no Septic shock =0
14	MODS	bool	If any of MODS then 1; else 0
15	No of days of ventilator support	int	If not taken ventilator support value 0 else the no of days spent in ventilator
16	ICU admission	bool	If admitted in ICU then 1 otherwise 0
17	No of days of ICU admission	int	If admitted in ICU the count of no of days spent in ICU
18	Co-morbidities	bool	If any of the co morbidities existed then 1 also 0
19	No of days since symptomless	int	No of days since symptoms did not show
20	Antibody	bool	If antibody exists then1 else 0
21	Antibody titer	float	Test value of antibody titer
22	Eligible	bool	If the donor is eligible with antibody value > 6.5 then 1 else 0

Based on the fact that aged people are more prone to critical symptoms such records are given a higher probability to have critical symptoms. Since only crucial cases might require ventilator support and ICU admission, more crucial looking records are given a higher probability of having a True value for these features. The Eligible column is derived from 21(antibody titer value) column with a 1 if the value is greater than 6.5 and 0 otherwise. This column can be taken as output feature in classification.

C. Visualization of Data

In this section, we will visualize the data, the distribution of some of the features, and correlations between the features. Fig. 1 is numbered from a to g. It has depicted the following details (a) denotes the “Gender” density: Females are of less percentage but Eligible candidates are present equally in both. (b) For “Age” more candidates are clustered near 35 and as we go up the Eligible candidates outnumbered Ineligible candidates and vice versa when moving towards lesser age. (c) the distributions of different “Blood Types” are as per Asian statistics. (d) “Symptom Fever” represents the distribution of all mild severity symptoms. Even though, fewer people are assigned this symptom more candidates having fever has become Eligible (e) ”Respiratory Failure” represents the severe symptoms and the plot shows that more people with severity have become Eligible (f) ”ICU Admission” also points to the severity, but the number of people admitted to ICU would be less but more of them are Eligible. A similar graph is obtained for “Number of days since symptomless” which implies how many days the donor had been symptomless up to the current date. More Eligible candidates are seen when the virus infection is recent.

Fig. 2 demonstrates a more detailed view of how the features are related to antibody titer in determining a sample Eligible. Age, when plotted against Antibody titer lets us observe that as age increases more points are cluttered above Antibody titer value 6.5 which is the threshold for being Eligible. Another feature “days of ventilator support” from Fig. 2 says that the majority of donors who have taken ventilator support (number of days greater than zero) have higher antibody levels signifying the idea that as severity increases the antibody level rises.

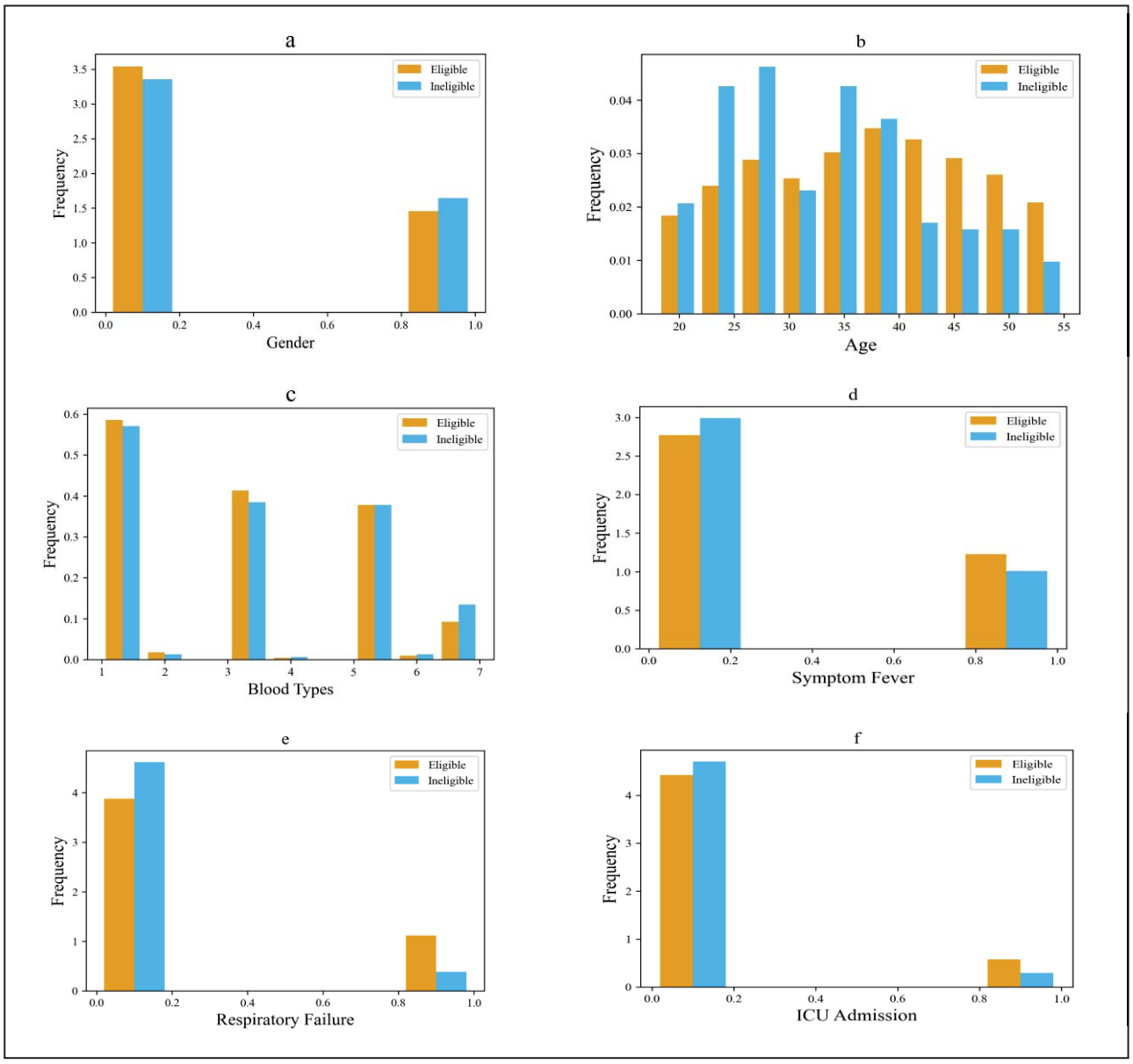

Fig. 1. Density distribution of various features a) Gender; (b) Age; (c) Blood Type; (d) Symptom fever; (e) Respiratory Failure;(f) ICU Admission;

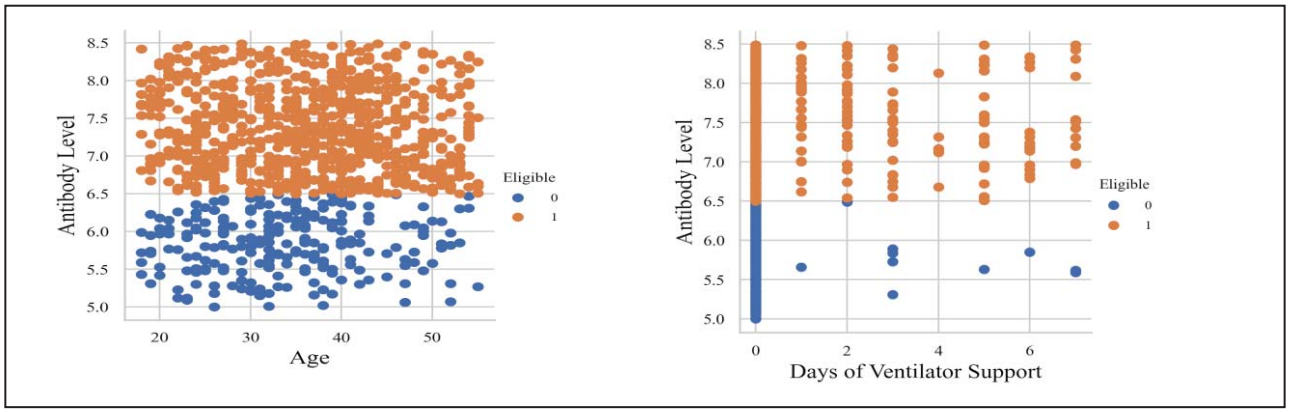

Fig. 2. Varition of antibody with different features

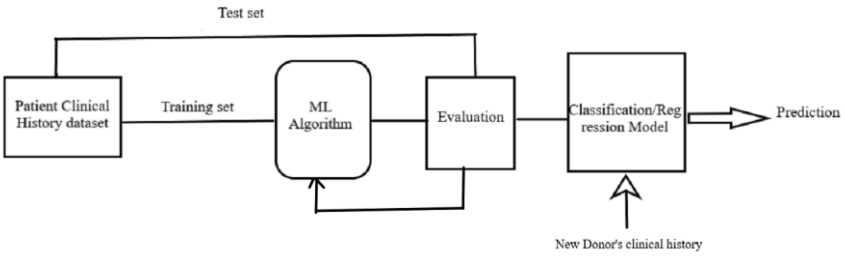

Fig. 3. Overview of the System

D. Classification Problem

Our primary goal is to determine whether a plasma donor is eligible for donating plasma (whether antibody level meets the threshold) from his basic clinical history before doing any serological tests. We can use classification algorithms to predict as the output feature is a categorical value. From Table II, features from 2 to 19 are taken as independent variables to predict the dependent variable Feature number 22 "Eligible". We have excluded columns 20 and 21, as they are the actual test values. As shown in fig. 3, a Classification model is built by training and evaluation. Then given a new user detail it can predict the class Eligible/Ineligible.

E. Regression Problem

While the classification algorithm can distinguish between an eligible and ineligible donor, ML Regression algorithms can be utilized to predict the antibody concentration value in a donor's body. Antibody concentration/titer value is a real value. It is the serological test value done at the donation time. From Table II, features from 2 to 19 are taken as independent variables to predict the dependent variable Feature 22 "Antibody titer". As shown in fig. 3, a Regression model is built by training and evaluation. Then given a new user detail it can predict a real value of antibody titer. If Regressor can predict the value of the antibody level, then among a group of donors the one with the higher value could be chosen as that can guarantee a better chance of survival for the recipient.

IV. Results Analysis

A. Classification Results

To solve the classification problem stated in Section III, we have used three conventional machine learning classification algorithms: Logistic Regression (LR), K Nearest Neighbours (KNN), Decision Tree (DT), ensemble methods: Random Forest (RF), Gradient Boosting (GB). The models are trained using the Donor dataset and evaluated using the K Fold cross-validation procedure. All the algorithms are evaluated using metrics relevant for Classification. The evaluation is performed using *sklearn* library. The metric values obtained when each of these algorithms are evaluated against our synthetic dataset are presented in Table III.

According to the results from Table III, the Accuracy of LR, DT, RF, and GB is satisfactory. Accuracy is the percentage of True predictions against total predictions. But since our dataset is imbalanced with more Eligible samples than Ineligible samples. Accuracy is not reliable in determining the best performance. Moreover, the False Positive and False Negative cases are equally important here. Hence F1 Score is also evaluated. F1 Score is the weighted average of *precision* and *recall* ranging from 0 to 1. The F1 scores of all the algorithms were acceptable but LR, RF, and GB have outperformed the other two. The Area Under Curve (AUC) values of LR, RF, and GB are closer to 1 indicates their edge over the other 2 algorithms. AUC can be defined as the degree of separability. Finally, the Matthews Co-relation Coefficient metric (MCC) that can give a balanced measure of performance with values ranging between [-1, 1] again proves that LR, RF, and GB are better. Being consistently giving better values for all the analyzed metrics, Gradient Boosting can be considered as the best approach for classification.

TABLE III. Performance Comparison of Classification Algorithm

Algorithm	Accuracy	F1-Score	RoC-AUC	MCC
LR	0.83±.011	0.89±07	0.83±.17	0.51±24
KNN	0.70±.13	0.81±.10	0.51±.17	0.01±.25
DT	0.77±.10	0.85±.07	0.70±18	0.37±28
RF	0.83±12	0.89±09	0.82±.17	0.52±31
GB	0.85±.10	0.90±07	0.82±17	0.56±26

B. Regression Results

The Regression, problem defined in section III, is addressed with three conventional regression algorithms: Linear Regression (LR), KNN Regression, Decision Tree (DT) Regression, and ensemble methods: Random Forest (RF) Regression, and Gradient Boosting (GB). The Algorithms were trained on the same synthetic dataset. The performance is evaluated using the K Fold cross-validation procedure. The evaluation is performed using *sklearn* library. The performance metrics analyzed are Mean Absolute Error (MAE), Mean Squared Error (MSE), R squared. MAE is the mean of the absolute value of the difference between real and obtained values. MSE is the average squared difference between actual and obtained values. R squared is the measure of how good fit the model is. The scores produced when the algorithms are evaluated with 1000, 10000, and 100000 samples of synthetic data are presented in Table IV.

TABLE IV. PERFORMANCE COMPARISON OF REGRESSION ALGORITHM

Algorithm	MAE			MSE			R Squared		
	1000	*10000*	*100000*	*1000*	*10000*	*100000*	*1000*	*10000*	*100000*
LR	0.57±.11	0.57±.04	0.58±.01	0.49±.17	0.51±.06	0.51±.02	0.28±.23	0.27±.05	0.27±.02
KNN	0.69±.17	0.66±.05	0.65±.01	0.75±.37	0.69±.08	0.66±.08	-0.07±.27	0.01±.11	0.05±.03
DT	0.63±.15	0.62±.04	0.61±.01	0.63±.27	0.57±.08	0.56±.02	0.06±.47	0.18±.10	0.19±.03
RF	0.47±.08	0.45±.02	0.45±.01	0.33±.01	0.29±.03	0.29±.0	0.51±.19	0.58±.04	0.59±.01
GB	0.53±.10	0.49±.04	0.53±.10	0.41±.14	0.35±.05	0.33±.01	0.39±.21	0.50±.04	0.52±.01

MAE scores say that KNN and DT make the predictions too far off from the actual value while RF has a comparatively less error than all others. MSE tells us which predictive model outperforms others as it says lower the value the better the model, and 0 means perfect prediction. RF stayed best among all others with comparatively closer values to 0, whereas KNN and DT again gave a poor performance. For R Squared, a value greater than 0.5 is acceptable, and the best score is 1. Among the algorithms we have used, RF has scored above 0.5. To summarize the comparison of algorithms, RF kept up marginally higher values for all the metrics. However, it is worth noticing that all the algorithms have improvised the performance as the sample quantity of data points scaled high up to 1,00,000.

C. *Discussion and Summary*

- The experiment indicates that the problem can be modeled as a real ML problem.
- The Performance of Gradient Boosting was the best we have found for the classification problem, but the performance of conventional Algorithms like Logistic Regression and Decision Tree was also comparable. As for the Regression problem, Random Forest Regressor gave acceptable results. However, the best performer algorithms may change in reality as we have done the whole experiment on a synthetic dataset. That explains any other ML algorithms and optimization techniques are not experimented. The problem needs a thorough analysis of the real data before experimenting and finalizing the best suitable algorithm.
- The reason for not getting an efficient performance, especially in Regression problem, can be justified with the facts such as 1) randomness incorporated in the data: 2) Non-inclusion of many clinical severity aspects like Co-Morbidities, Other Symptoms, etc. which can take different values in reality, the inclusion of unimpactful features like height, weight, etc. Consider the fact that additional explanatory variables can improve the fit of the models, and help achieve better R squared values.
- There exists a plethora of techniques for improvisation of performance of ML model like rescaling the input variables, feature selection and engineering, using different evaluation metric, evaluate a diverse suite of linear and nonlinear algorithms, Configuring and customizing each algorithm in a way well suited for the problem, Use of different ensemble techniques, etc.
- The purpose of the experiment is suggesting a better paradigm donor selection.

V. CONCLUSION

The paper points out the complexities and inconveniences in finding a donor for Convalescent Plasma therapy. It shows how the problem can be solved in a data-driven way. The solution is based on one classification model to predict whether the donor has the threshold antibody level for donation and a regression model that can predict which donor can have a better level antibody titer in his plasma based on his/her clinical history. The proposed data-driven method lets the concerned people find who is Eligible for donation and among the Eligible ones who has the efficient antibody level beforehand. As the hospital authorities has the clinical history of all the COVID cured patients, they can make use of the system to predict who has an efficient antibody level and make a smart approach by contacting them for donation if they consent. This can prevent wastage of time, cost, effort and test kits on every random donor, especially, in emergencies. We have simulated the data and built different Classification and Regression models. Although the results look satisfactory to some extent, valid conclusions cannot be drawn due to the unavailability of authentic data. Even though the system is proposed in the context of COVID-19, the methodology can be utilized for any future virus outbreaks too. Upon the availability of real data, the system could be implemented and optimized to produce the best results.

REFERENCES

[1] J.Epstein , and T.Burnouf, "Points to consider in the preparation and transfusion of COVID-19 convalescent plasma", Vox sanguinis, vol.115(6), pp. 485-487, August 2020.

[2] P.Marson, A.Cozza, and G.De Silvestro, "The true historical origin of convalescent plasma therapy", Transfusion and apheresis science", vol. 59(5), October 2020.

[3] G.Marano et al. “Convalescent plasma: new evidence for an old therapeutic tool?.” Blood transfusion = Trasfusione del sangue vol. 14(2) pp. 152-7, March 2016.

[4] P.Tiberghien et al., "Collecting and evaluating convalescent plasma for COVID-19 treatment: why and how?", Vox sanguinis, vol. 115(6) pp. 488-494, August 2020.

[5] K.Rajendran et al., "Convalescent plasma transfusion for the treatment of COVID-19: Systematic review. Journal of medical virology", Journal of Medical Virology, vol.92, pp. 1475-1483, September 2020.

[6] K.Duan et al., "Effectiveness of convalescent plasma therapy in severe COVID-19 patients", Proceedings of the National Academy of Sciences, vol. 117 (17), pp. 9490-9496, April 2020.

[7] X.Liu et al. “Patterns of IgG and IgM antibody response in COVID-19 patients.” Emerging microbes & infections, vol. 9(1) pp. 1269-1274, June 2020.

[8] H.Hou et al. “Detection of IgM and IgG antibodies in patients with coronavirus disease 2019.” Clinical & translational immunology, vol. 9,5, May 2020.

[9] GJ.Gorse, MM.Donovan, and G B.Patel "Antibodies to coronaviruses are higher in older compared with younger adults and binding antibodies are more sensitive than neutralizing antibodies in identifying coronavirus-associated illnesses.", Journal of medical virology,, vol. 92(5), pp.512-517, May 2020.

[10] B.Pawłowski, J.Nowak, B.Borkowska, D.Augustyniak, and Z.Drulis-Kawa, "Body height and immune efficacy: testing body stature as a signal of biological quality.",Proceedings. Biological sciences, vol. 284(1859), July 2017.

[11] D.Jacofsky, EM.Jacofsky, and M.Jacofsky, "Understanding Antibody Testing for COVID-19.", The Journal of arthroplasty, vol. 35(7S), pp. S74–S81, July 2020.

[12] Vishal, Python Random Module to Generate random Data [Guide]. Retrieved from https://pynative.com/python-random-module/

[13] F.Pedregosa et al., "Scikit-learn: Machine Learning in Python", arXiv,2018.

[14] H.David, "Assessing the Performance of Classification Methods", International Statistical Review, vol. 80(3), pp. 349-493, December 2012.

[15] C.Ferri, JH.Orallo, and R.Modroiu, “An experimental comparison of performance measures for classification.” Pattern Recognit. Lett, vol. 30, pp. 27-38, 2020.

[16] HF.Golino, et al., "Predicting increased blood pressure using machine learning.", Journal of Obesity, vol. 2014, January 2014.

[17] K. Plis and Razvan C. Bunescu and C. Marling and J. Shubrook and F. Schwartz,"A Machine Learning Approach to Predicting Blood Glucose Levels for Diabetes Management",AAAI Workshop: Modern Artificial Intelligence for Health Analytics, 2014.