

Machine Learning Models to Predict Epidemiological Trends from Pathology Data

Felipe Ramos Teixeira^{1*}, Rafael Guilherme Santos²

¹ Universidade Federal do Rio de Janeiro (UFRJ), Brazil

² Universidade Federal de Minas Gerais (UFMG), Brazil

ABSTRACT

Pathology and pathologists play an essential role in treating various contagious diseases and analyzing drug expansion and clinical analysis tasks. The workflow of Pathologists is complex, specific and limited by human evaluation. Supportive clinical decision-making tools help transform the pathology field with the help of machine learning tools. It will improve the predictive reads compared to manual reads, automatic approaches for scoring clinical trial slides, discovery of innovative markers and recognizing the pathology of cancer. The machine learning models must require authentication and certification before they are applied for any pathological analysis as they impact individual life and health. Pathological images are immense in resolution and different from the normal image; machine learning models required to solve these problems need an understanding of the biological structure and guidelines to prevent model failure that is supposed to predict life-threatening diseases. Machine learning models like deep learning (DL) models give promising results in detecting the images. Machine learning models help in detecting the histopathological picture of tissues, and the deep learning model is one of the models that shows optimized outputs. Consequently, these images are different from natural images and for analyses of these images, there is a need to demonstrate a machine to recognize them.

Keywords: Machine learning Models (MLM), Predict Epidemiological Trends (PPT), Pathology Data (PD), Smart PLS Algorithm Model

INTRODUCTION

The branch in which we study the occurrence of various diseases in people belonging to various groups and the reason for the occurrence of those diseases within people. In this information, various methods and techniques are planned and then evaluated to avoid illness among various individuals. It also acts as a guide for managing all those patients in whichever diseases have occurred. Just like the achievements of clinics and pathology departments, the epidemiology of a particular disease is considered a necessary element of all essential information. This thing has the specific character of accumulation of data and then elaboration of that data. It is an essential jargon on behalf of technical terms. No doubt, its presentation represents that it is a short story but in actuality, it is providing all the essential information that is related to the basic knowledge of epidemiology, all its methods, and all its techniques[1, 2]. The only idea that circulates commonly in the minds of people is that it is the thing that is mostly epidemiological questions are considered as Worth answering. This brief introduction gives a clue about the well-known characteristics of epidemiological proceedings. One of the most well-known characteristics that is distinctive about epidemiology is measuring the level of consequences of a disease that are related to a specific population that is seen at greater risk. The population that is in the danger zone is the group of people that are either healthy or sick. But these people come under the category of well-known cases if they belong to those groups of diseases that have been studied already and now are under discussion to control those cases[3]. Let's try to understand it with the help of an example if a regular or general observer is trying to understand the level of consultation of disease patients with the expert practitioner about their issue of deafness. Then all those people are included in that list which can be seen with the issue of hearing. Patients who are on that list on behalf of their problem but later move towards another location will not be included in that list and can consult the doctor of that specified

area. After the change of their location, they will not come under the category of those people who are at risk[4]. The necessity and importance of considering all those people who are at risk is of great significance and it can be elaborated with the help of the following two examples. Perhaps these examples help you to understand the main objective of elaborating this topic. In a study that is all about the injured people who were injured in an accident and then they were admitted to hospitals later on. The study represents that most people are old and after the results of this study, the author came to the point that all those people who are age 60 or more than it are at greater risk of accidents and then they will come under the category of patients. Another study in which a survey was performed to notice the accidents that happen during hang gliding[5]. It was declared that flying should not be done during the time of 11 am and 3 pm because it is the time when 73% of accidents are noticed. Both of these observations are all about the same logical error, which is known as the floating numerator. As there is no connection found between several cases and the appropriate that's population it is at risk. These results are nothing fixed. These can vary according to the situation. The number of accidents to patients and hang gliders may vary[6]. Number of people who are at risk keep on varying. So, one can come to know to this point that the inferences of epidemiological research are not concluded completely and purely from clinic-based data [7]. Some of the models that are all about machine learning for the prediction of analytics involve linear regression. In this model that variable is predicted that is that is continuously dependent on independent variables that may be one or more. Decision trees play the role of a tree in this regard. A tree-like structure is used for representing decisions and all their outcomes that are possible. To diagnose an image-based disease deep learning technique is utilized. Deep learning is a beneficial method by which several characteristics and patterns are studied and represent that it is one of the most powerful techniques[8]. It has become

possible only through the efforts of deep learning to analyze models to diagnose diseases at early stages. Machine learning approaches come under that category in which outcomes are predicted and integrate the characteristics of patients by analyzing the patients at a larger scale. This technique is applied widely in epidemiology research that is completely clinical and is based on diabetes. This approach has become noble and popular in all fields which are other than biomedical research, machine learning terminology techniques, and findings of various research[9]. These outcomes may have no interaction with all those researchers that are related to diabetes. The main aim of this study is to utilize the machine learning techniques effectively. All those techniques in which machine learning is involved by using those learners that are tree-based. They produce the decision trees that seem very helpful for guiding all clinic-related interventions. All these models have a higher level of sensitivity and specific nature than conventional retroversion models for the prediction of risks[10]. By the use of neural chain networks and deep learning methods machine learning approaches can be applied can be implicated in medical image data more specifically to recognize and staging of some sensitive diseases like diabetic retinopathy and skin ulcers[11].

A machine learning model is a program that can identify patterns or make predictions based on previously undetermined data. For example, in natural language processing, machine learning models can assess and accurately discern the intent behind previously unknown words or word combinations. In image recognition, a machine learning model can be trained to recognize vehicles and dogs. A machine learning model that has been 'trained' on a big dataset may be able to perform such tasks. During training, the machine learning algorithm is tweaked to detect specific patterns or outputs from the dataset that are relevant to the job. This method creates a machine learning model, which is often a computer program with predefined rules and data structures.

Research Objective

The main purpose of this article is to understand the epidemiological trends from that data that is related to pathology by the application of machine learning approaches. Hence, this technique is proving very beneficial in this regard. The research study determines that Machine Learning models to Predict Epidemiological Trends from Pathology Data. The research paper divided into five specific research chapters first portion represent the introduction related to the machine learning models and pathology data. The first section also describes the objective of research study. The second portion represent the literature review the third section describe methodology included method of research also that fourth section represent result and its descriptions. The last section summarized overall research study and explain the recommendations about topic.

LITERATURE REVIEW

Researchers claim that the machine learning approach is used in clinical research programs to predict the trends related to epidemiological basis from large data reports. Diabetes research has become advanced by using a machine learning approach to explain the information related to this disease[12].studies suggest that lung cancer identification and its prevalence rate is determined using a big data model .This model uses the data of EHR and predicts the number of cancer patients in a specific area. the model is fully optimized to provide help to physicians to determine patients who are at higher risk

of lung cancer[13].studies show that different breast cancer types are effectively treated using NAC. BC is a type of breast cancer that gives pathological responses This response varies in various breast cancer patients. by using [a predictive model based on machine learning for the prediction of BC patient pathological responses[14].studies explain that health-related problems in SSA are solved using deep learning algorithms in health sectors. the detection of different disease types at an early phase is possible through the help of deep learning technology This technology analyzes big data and provides early diagnostic reports related to the patients[15].studies explain that from diagnosing diseases to recommending treatment, the use of AI in the health field is a great advancement. AI is specialized to detect medico legal problems in the health sector. The physician's decision-making ability when combined with the AI approach then the best possible solution for any disease treatment is developed[16].also, AI is specialized to complete tasks much faster than humans. The ability of AI-based health models to predict clinical data and to draw conclusions makes it the most suitable technology for the medical sector. Various AI prediction models are used in the health sector that serve different functions. like the AI prediction model is specialized to predict the outcomes of clinical data with higher precision[17].scholars studies reveal that most clinical research programs use machine learning technology to achieve precise results Clinical decision-making programs in health sectors are made using machine learning to improve the intervention providing services of these sectors[18].studies elaborate that in animal populations, many disease outbreaks occur that reduce the number of specific animal populations.to determine the disease outbreak in animals the use of an animal surveillance prediction model is preferred This model predicts temporal reasons behind the onset of disease in the animal population. the one significant feature of this model is that it predicts the epidemiological trends related to the animal population[19].studies suggest that using deep learning algorithms to read clinical data reports is becoming common in pathology labs. The pattern detection process of clinical reports is possible through DL algorithmic approach. The AI techniques that have gained popularity in clinical research labs are based on the neural network approach. The in-depth knowledge and information are obtained using the neural network approach of AI[20] Studies show that ANN is used for identifying the temporal statistics related to DSS of cancer patients. the data obtained from SEER were analyze to using AI algorithms .also, the prognostic models are developed to understand the process of clinical decision making and to calculate the patient expected chances of patient survival[21].researchers explained that CRC patient condition is determined using a simplified prediction model .the patients of CRC faces LNM problems that is identified through the results statistic of prediction model. This model provides guide to carry out the treatment decision against CRC[22].studies explain that the identification of gastric cancer at early stage is possible by measuring the tumor size and its location. RDA model is used that uses neural network analysis for predicting the pathological information related to the patient. The RDA model is regarded as the best predictive model for early diagnosis of gastric cancer[23].studies show that oral cancer is related to tumor cell formation in the oral cavity. the outcomes and results related to the presence of oral cancer in patients are predicted using the ML model. Molecular biomarkers are used in the working mechanism of ML models to improve the accuracy of predicted results[24].studies claim that the major burden on the health sector is because of the high occurrence rate of chronic diseases. CDs are highly dangerous and

require costly treatments. Using predictive models increases the chances of understating CDs. To classify various types of CD, predictive models are the most efficient[25].studies predict that with the help of ML in the health sector the burden on physicians has been greatly reduced. Most medical devices operate using ML-based algorithms. ML models are specifically designed to aid physicians in their clinical work so that physician and provide timely treatment [prediction to the patient[26].researchers studies suggest that dengue fever occurrence in brazil and its neighboring areas is predicted using models that are based on epidemiological variables. These diverse models are developed to track the presence of dengue fever in distant geographical Areas of Brazil. The most commonly used models for dengue fever diagnosis is ML model .these models uses algorithmic approach to trace the diversity of dengue fever in Brazilian states[27].studies explain that self-supervised model based on deep learning approach provide insight about cancer using the generalized biomarkers[28].studies reveal that interstitial pneumonia is a infectious diseases that .the histopathological basis of this disease is identified using dep learning model to explains the prognosis of this disease.to develop deep learning based pathological models human expertise and AI approaches are combined. The combination of two approaches is used in medical imaging systems to study other types of complex diseases. also, the combination of the male approach with the AI-based concept has given a new dimension to the use of AI in medical sectors[29].

Researchers suggest that during the period of the COVID-19 pandemic, most of the world population was affected by this delayed virus. The exact estimation of the number of people affected by COVID-19 during the pandemic was made using the LSTM model. The one big advantage of this model is that it accurately predicts the active cases of the coronavirus around the world[30].studies claim that melanoma patients are provided with immunotherapy-based treatments to treat the disease. the effectiveness of this therapy against melanoma is determined by using ML algorithms. By estimating the response of patients to the immunotherapy the ML algorithm-based clinical models predict patient recovery status[31].

Methodology

The research determines that Machine Learning models to Predict Epidemiological Trends from Pathology Data. The research based on primary data analysis for determine the research used smart PLS software and generate result included descriptive statistic, the correlation coefficient also that explain the smart PLS Algorithm model between them. The data related to medical growth each year by 48% and this outburst data must be used optimally to face the challenges in the healthcare field to improve this sector. It can only be possible with machine learning and artificial intelligence-based technologies. Machine learning (ML) technology based on artificial intelligence enables the computer to experience-based learning without explicit approaches. Machine learning models are developed by data collection and instruction via approaches related to statistics and mathematics. It predicts the newly previously unseen data in the absence of explicit programs. Currently, a tremendous revolution in AI and ML credit goes to advancement in computer programs and accessibility. Earlier, these technologies were only available in science fiction but, now they are part of every other field. Presently, these technologies are integrated into a wide range of health disciplines and medical practices and keep increasing gradually. The progressive role

of ML+AI in health-related sectors, FDA guidelines demonstrated that both ML+AI show promising results for patient health care. The pathology laboratories need to have advanced technology and non-stop development tools related to ML and AI. These technologies integrated with the advancement in healthcare fields. For instance, expectedly 70% health care decision depends on the laboratory analysis. The algorithms of machine learning technology spatially and temporarily help in controlling the spread of infectious diseases and also predict the progression of diseases. Machine learning algorithms can evaluate large, difficult sets of information that are not possible to evaluate by humans. This feature makes the ML technology suitable for the detection of different contagious diseases, caused by various parameters such as population size, atmospheric conditions and behaviour individual. For a few decades, ML technology has provided promising results in the detection of contagious diseases. Machine learning technology also faced some challenges, such as high-quality and inclusive data availability for the prediction of diseases. The investigation of infectious diseases frequently gathered information on several parameters such as, the number of infected cases, pandemic location, and infected case demographics. However, this information can be prejudiced, partial and noisy, which can disturb the working of machine learning algorithms. Moreover, various contagious diseases past the pandemic do not reflect the current situation as they have a long incubation time. All these challenges can be overcome with the development of various algorithms in ML, including random forest, decision tree, support vector machine and many others. All these models can be applied to different types of data sets, including health data sets, genetic data and societal information. In general, many studies demonstrated that machine learning models can detect the pandemic and outbreak of infectious diseases accurately and better than as compared to traditional methods. The models of machine learning show optimized results in the prediction of diseases and outbreaks. For instance, some research predicts the number of reported cases of specific contagious diseases in a specific area based on the history and present condition with the help of machine learning models. Some researchers demonstrated the source of the pandemic based on the genetic makeup and pattern of contamination of the pathogens using machine learning models. Yet, others used ML to detect the chances of a person to affect by a contagious disease on account of his behavior and characteristics. Conversely, there is debate on using the deep learning method as researchers are not clear about its working yet. In another current research DL model of machine learning shows promising results in cancer pathology. Deep learning models are established to work on cancer pathology including diagnosis of cancer, cancer subtypes, staging, predictions, pathological characteristics identification and changes in genetic makeup and biomarkers. The use of digital pathology reduces the load of work on them so they will have spare time for health decision-making for the improvement of the health sector. Pathologists can predict pandemics of contagious diseases and provide precise and accurate results. Yet, there are some challenges related to ML such as models' verification, interpretability of computer system and human, and skepticism of pathologist, patient and clinicians. The machine learning method showed promising results in the prediction and analysis of infectious diseases, epidemiology, and public health dealing and clinical applications. Nonetheless, some challenges are also faced by machine learning technology such as it includes high-quality data, complex systems, and risk of bias and not suitable for general data sets but limited to

trained information. The most important challenge experienced by ML is the requirement for high-quality data; that is not possible as some data is not completed, information is missing, and modifications in

the formats of collected data are needed. Moreover, the creation of an accurate machine-learning model is limited by the outbreak of contagious disease.

Descriptive Statistical Analysis

Table 1: Result of Descriptive Statistical Analysis

Name	No.	Mean	Median	Scale Min	Scale Max	Standard Deviation	Excess Kurtosis	Skewness	Cramér-Von Mises P Value
MLM1	1	1.510	1.000	1.000	3.000	0.610	-0.305	0.794	0.000
MLM2	2	1.388	1.000	1.000	3.000	0.565	0.441	1.164	0.000
MLM3	3	1.429	1.000	1.000	3.000	0.535	-0.671	0.709	0.000
MLM4	4	1.531	1.000	1.000	3.000	0.575	-0.634	0.541	0.000
MLM5	5	1.469	1.000	1.000	3.000	0.538	-0.915	0.530	0.000
MLM6	6	1.531	1.000	1.000	3.000	0.575	-0.634	0.541	0.000
MLM7	7	1.653	2.000	1.000	3.000	0.555	-0.706	0.094	0.000
PET1	8	1.571	2.000	1.000	3.000	0.606	-0.545	0.567	0.000
PET2	9	1.531	2.000	1.000	3.000	0.538	-1.113	0.279	0.000
PET3	10	1.531	1.000	1.000	3.000	0.575	-0.634	0.541	0.000
PET4	11	1.653	2.000	1.000	3.000	0.656	-0.659	0.522	0.000
PET5	12	1.408	1.000	1.000	3.000	0.569	0.200	1.063	0.000
PET6	13	1.408	1.000	1.000	3.000	0.569	0.200	1.063	0.000
PD1	14	1.592	2.000	1.000	3.000	0.569	-0.756	0.312	0.000
PD2	15	1.796	2.000	1.000	3.000	0.669	-0.763	0.271	0.000
PD3	16	1.796	2.000	1.000	3.000	0.669	-0.763	0.271	0.000
PD4	17	1.735	2.000	1.000	4.000	0.750	0.238	0.791	0.000

The previous findings shown in table 1 that descriptive statistical analysis results indicate mean values, median values, and standard deviation rates, as well as explain the probability value of each indicator, both dependent and independent. According to the results, the total minimum value is 1.000, while the maximum value is 3.000, indicating that the median rate is 2.000. The mean values for PD1,2,3,

and 4 are 1.592, 1.796, and 1.735, respectively, indicating a positive average rate with standard deviation rates of 56%, 66%, and 75% deviating from the mean. Similarly, the results show that the skewness value is 27%, with 79% skewness rate for each dependent variable. The MLM exhibits a positive average rate and a positive standard deviation rate between the dependent and independent variables.

Correlation Coefficient

Table 2: Result of Correlation Coefficient

	MLM1	MLM2	MLM3	MLM4	MLM5	MLM6	MLM7	PET1	PET2	PET3	PET4	PET5	PET6	PD1	PD2	PD3	PD4
MLM1	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MLM2	0.137	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MLM3	0.206	-0.145	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MLM4	0.043	-0.131	-0.209	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MLM5	0.016	-0.330	0.152	0.382	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MLM6	0.275	0.058	0.256	0.198	0.052	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MLM7	0.161	-0.027	0.364	-0.127	0.067	0.257	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PET1	-0.071	0.247	-0.252	0.067	-0.447	0.067	-0.139	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PET2	-0.389	-0.140	-0.365	-0.052	-0.085	-0.316	-0.408	0.322	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PET3	-0.190	0.309	-0.142	0.198	-0.277	0.136	0.065	0.360	0.013	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PET4	-0.119	-0.188	0.075	-0.161	-0.059	0.055	0.230	-0.323	-0.461	-0.053	1.000	0.000	0.000	0.000	0.000	0.000	0.000
PET5	0.106	-0.048	-0.173	-0.350	-0.226	-0.163	0.125	-0.025	0.093	-0.101	0.051	1.000	0.000	0.000	0.000	0.000	0.000
PET6	-0.071	0.143	0.029	-0.101	-0.026	-0.038	0.319	0.034	-0.041	0.024	0.051	-0.137	1.000	0.000	0.000	0.000	0.000
PD1	0.247	0.239	0.240	0.101	-0.107	0.350	0.263	0.085	-0.293	0.225	0.058	0.010	-0.053	1.000	0.000	0.000	0.000
PD2	0.005	0.155	0.073	-0.143	-0.187	0.175	0.084	0.187	-0.153	0.016	0.303	-0.103	-0.049	0.317	1.000	0.000	0.000
PD3	-0.045	0.371	0.016	-0.090	-0.131	0.016	0.194	0.237	-0.209	0.122	0.025	-0.049	0.219	0.371	0.362	1.000	0.000
PD4	-0.150	-0.335	0.080	-0.005	0.207	-0.147	-0.074	-0.160	-0.157	-0.147	0.435	0.063	0.110	-0.254	0.217	-0.149	1.000

The following result of table 2 describes the correlation coefficient analysis result, which reveals that there are some positive and negative relationships between dependent and independent variables. Non-Hodgkin lymphoma (NHL) is the world's tenth most common cancer in 2018, with diffuse large B-cell lymphomas (DLBCL) accounting for approximately 30% of NHL. Lymphoma can develop in any tissue, and around one-third of patients present with extranodal sites. Between 1975 and 2003, the incidence of NHL in the United States became by more than 80%, with exogenous NHL accounting for the large majority of new cases. The etiological exposure and the pattern of incidence and death can change based on the anatomical place. Existing data on DLBCL incidence and death trends do not effectively depict extranodal DLBCL, do not account for the effects of tumour features or patient

characteristics, and do not allow for systematic comparisons of mortality and incidence trends depending on these parameters. Most DLBCL patients' survival rates have improved considerably since the introduction of R-CHOP (rituximab with cyclophosphamide, doxorubicin, vincristine, and prednisone) as a first-line treatment. However, therapy-related side effects and long-term repercussions have developed, and a considerable minority of patients do not react to early treatment or eventually relapse. The International Prognostic Index (IPI) is the most extensively used prognostic tool for DLBCL patients. The number of extranodal sites is one of the parameters used to compute the IPI score. However, Lu CS et al. observed that some extranodal sites were more informative than the total number of extranodal sites implicated. Furthermore, there is emerging evidence

that initial extranodal placements have distinct clinical characteristics and predictive outcomes, necessitating a particular therapeutic strategy. As a result, a new risk categorization signature that integrates extranodal origin sites is necessary to guide extranodal DLBCL treatment. Nomographs, which are graphical representations of mathematical models, can leverage information from many sources

to predict a given outcome in clinical practice. A nomograph can assess the probability of an event for each patient, such as death or recurrence, by including a number of important information. As a result, the nomograph has become a significant tool for forecasting clinical outcomes in many forms of cancer, perhaps presenting clinicians with appropriate therapeutic methods.

Smart PLS Algorithm Model

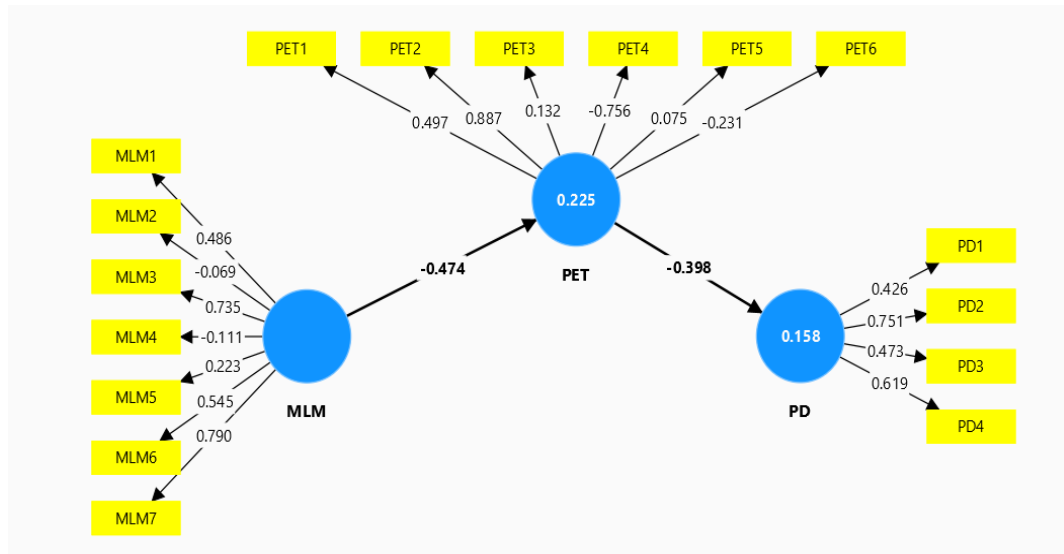


FIGURE 1: Smart PLS Algorithm Model

The above model of figure 1 called smart Algorithm model in between MLM and PD also that PET result shows that MLM is main independent variable result shows that 0.486, -0.069, 0.735, -0.111, 0.223, 0.545 also that 0.790 its shows that some positive and some negative link with MLM. The result shows that MLM present negative but its 47% significant link with PET. According to the result the PET shows that -0.398 shows that negative but its significant level between them. the PD shows that 42%, 75%, 47% and 61% significant level between them.

Conclusion

When evaluating our findings, we should keep in mind the limitations caused by the SEER database's availability of information. For starters, there isn't much information on treatment. Chemotherapy, a significant therapeutic option for DLBCL, is not included in SEER. As a result, we were unable to investigate the effect of therapy on survival, as well as the incidence and death trends. We addressed this issue in part by categorizing diagnostic time periods into 1970s, 1980s, 1990s, 2000s, and 2010s. Second, because the SEER database has collected Ann Arbor Stage data since 1983, incidence-based mortality rates may be underestimated. Third, no information was provided about radiation exposure, environmental exposures, individual lifestyle, or family history, all of which could be linked to incidence and mortality rates. Fourth, there was no record of baseline performance status, B-symptoms, bulky disease, or lactate

dehydrogenase levels. As a result, we were unable to integrate these potential prognostic signals into our prediction model. As far as we know, this is the first large-scale investigation to detect and systematically assess trends in the incidence and mortality of extranodal DLBCL based on clinical parameters, thus assisting scientists in identifying illness-related risk factors and effectively managing the condition. Extranodal DLBCL should be addressed in such a way that patients can be classified into different prognosis groups. As a result, in the rituximab era, we developed a new prediction model that can provide accurate risk classification of individuals in order to select the optimal treatment options for them. We show that extranodal DLBCL incidence and death have been rising for decades, despite a promising recent reduction. Our survival analysis could help us better comprehend the incident. Between the 1970s and the 2010s, 5-year survival rates increased considerably (44.15 vs. 63.7%), with the greatest improvement shown in patients with a primary site in the head/neck. The overall mortality trend for extranodal DLBCL began to decline in 1993, ten years before the incidence, indicating that increased survival is the primary driver of the decrease. We hypothesize that improved patient care and therapy will result in increased survival rates. One of the most significant advancements was the use of the IPI index as the gold standard for distinguishing high-risk and low-risk individuals and directing therapy. Another major influence was the introduction of rituximab in 1988, which had the potential to cure patients. Furthermore, a better understanding of DLBCL genetics and molecular biology, including BCL2 protein expression and genetic complexity, may improve patient management and survival.

References

- [1] B. Beaulieu-Jones *et al.*, "Trends and focus of machine learning applications for health research," *JAMA network open*, vol. 2, no. 10, pp. e1914051-e1914051, 2019.
- [2] N. Bollig, L. Clarke, E. Elsmo, and M. Craven, "Machine learning for syndromic surveillance using veterinary necropsy reports," *PloS one*, vol. 15, no. 2, p. e0228105, 2020.

- [3] S. Eloranta and M. Boman, "Predictive models for clinical decision making: Deep dives in practical machine learning," *Journal of Internal Medicine*, vol. 292, no. 2, pp. 278-295, 2022.
- [4] O. A. Karadaghy, M. Shew, J. New, and A. M. Bur, "Development and assessment of a machine learning model to help predict survival among patients with oral squamous cell carcinoma," *JAMA Otolaryngology–Head & Neck Surgery*, vol. 145, no. 12, pp. 1115-1120, 2019.
- [5] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, "Machine learning applications in cancer prognosis and prediction," *Computational and structural biotechnology journal*, vol. 13, pp. 8-17, 2015.
- [6] N. Meti *et al.*, "Machine learning frameworks to predict neoadjuvant chemotherapy response in breast cancer using clinical and pathological features," *JCO Clinical Cancer Informatics*, vol. 5, pp. 66-80, 2021.
- [7] A. B. Kadir and N. R. Singh, "The Role of Digital Pathology in Enhancing Diagnostic Accuracy in Oncology," *J Bras Patol Med Lab*, 2024.
- [8] S. Chae, S. Kwon, and D. Lee, "Predicting infectious disease using deep learning and big data," *International journal of environmental research and public health*, vol. 15, no. 8, p. 1596, 2018.
- [9] C. Poirier *et al.*, "Real time influenza monitoring using hospital big data in combination with machine learning methods: comparison study," *JMIR public health and surveillance*, vol. 4, no. 4, p. e11361, 2018.
- [10] H. H. Rashidi, N. K. Tran, E. V. Betts, L. P. Howell, and R. Green, "Artificial intelligence and machine learning in pathology: the present landscape of supervised methods," *Academic pathology*, vol. 6, p. 2374289519873088, 2019.
- [11] S. Basu, K. T. Johnson, and S. A. Berkowitz, "Use of machine learning approaches in clinical epidemiological research of diabetes," *Current diabetes reports*, vol. 20, pp. 1-19, 2020.
- [12] S. Basu, K. T. Johnson, and S. A. J. C. d. r. Berkowitz, "Use of machine learning approaches in clinical epidemiological research of diabetes," vol. 20, pp. 1-19, 2020.
- [13] U. Chandran *et al.*, "Machine learning and real-world data to predict lung cancer risk in routine care," vol. 32, no. 3, pp. 337-343, 2023.
- [14] N. Meti *et al.*, "Machine learning frameworks to predict neoadjuvant chemotherapy response in breast cancer using clinical and pathological features," vol. 5, pp. 66-80, 2021.
- [15] E. Mbunge, J. J. T. Batani, and I. Reports, "Application of deep learning and machine learning models to improve healthcare in sub-Saharan Africa: Emerging opportunities, trends and implications," p. 100097, 2023.
- [16] P. Maurovich-Horvat, "Current trends in the use of machine learning for diagnostics and/or risk stratification in cardiovascular disease," ed: Oxford University Press, 2021.
- [17] R. W. Pettit, R. Fullem, C. Cheng, and C. I. J. E. t. i. l. s. Amos, "Artificial intelligence, machine learning, and deep learning for clinical outcome prediction," vol. 5, no. 6, pp. 729-745, 2021.
- [18] S. Eloranta and M. J. J. o. I. M. Boman, "Predictive models for clinical decision making: Deep dives in practical machine learning," vol. 292, no. 2, pp. 278-295, 2022.
- [19] N. Bollig, L. Clarke, E. Elsmo, and M. J. P. o. Craven, "Machine learning for syndromic surveillance using veterinary necropsy reports," vol. 15, no. 2, p. e0228105, 2020.
- [20] U. A. Usmani and J. Jaafar, "Machine learning in healthcare: current trends and the future," in *International Conference on Artificial Intelligence for Smart Community: AISC 2020, 17–18 December, Universiti Teknologi Petronas, Malaysia, 2022*: Springer, pp. 659-675.
- [21] Y. Li, Z. Shi, L. Tang, and Y. Li, "Epidemiological trends and new prognostic assessment methods for patients: a study based on the SEER database," in *Third International Conference on Biomedical and Intelligent Systems (IC-BIS 2024)*, 2024, vol. 13208: SPIE, pp. 480-486.
- [22] J. H. Ahn *et al.*, "Development of a novel prognostic model for predicting lymph node metastasis in early colorectal cancer: analysis based on the surveillance, epidemiology, and end results database," vol. 11, p. 614398, 2021.
- [23] H. Tian *et al.*, "Application of machine learning algorithms to predict lymph node metastasis in early gastric cancer," vol. 8, p. 759013, 2022.
- [24] J. Adeoye, J. Y. Tan, S.-W. Choi, and P. J. I. j. o. m. i. Thomson, "Prediction models applying machine learning to oral cavity cancer outcomes: a systematic review," vol. 154, p. 104557, 2021.
- [25] G. Battineni, G. G. Sagaro, N. Chinatalapudi, and F. J. J. o. p. m. Amenta, "Applications of machine learning predictive models in the chronic disease diagnosis," vol. 10, no. 2, p. 21, 2020.
- [26] A. Zhang, L. Xing, J. Zou, and J. C. J. N. B. E. Wu, "Shifting machine learning for healthcare from development to deployment and from models to data," vol. 6, no. 12, pp. 1330-1345, 2022.
- [27] K. Roster, C. Connaughton, and F. A. J. A. J. o. E. Rodrigues, "Machine-learning–based forecasting of dengue fever in Brazilian cities using epidemiologic and meteorological variables," vol. 191, no. 10, pp. 1803-1812, 2022.
- [28] J. M. Niehues *et al.*, "Generalizable biomarker prediction from cancer pathology slides with self-supervised deep learning: A retrospective multi-centric study," vol. 4, no. 4, 2023.
- [29] W. Uegami *et al.*, "MIXTURE of human expertise and deep learning—developing an explainable model for predicting pathological diagnosis and survival in patients with interstitial lung disease," vol. 35, no. 8, pp. 1083-1091, 2022.
- [30] T. H. Aldhyani, M. Alrasheed, A. A. Alqarni, M. Y. Alzahrani, and A. H. J. M. Alahmadi, "Deep learning and holt-trend algorithms for predicting covid-19 pandemic," p. 2020.06.03.20121590, 2020.
- [31] P. Johannet *et al.*, "Using machine learning algorithms to predict immunotherapy response in patients with advanced melanoma," vol. 27, no. 1, pp. 131-140, 2021.



This is an open-access article distributed under the terms of the Creative Commons Attribution License.