Establishing Cause-Effect Relationships from Medical Treatment Data in Intensive Care Unit Settings

Mohammed Abebe Yimer
Faculty of Computing and Software Engineering, Arba Minch University, Ethiopia
mohammed.abebe@amu.edu.et

Özlem Aktaş
Department of Computer Engineering, Dokuz Eylül University, Turkey
ozlem@cs.deu.edu.tr

Süleyman Sevinç
Department of Computer Engineering, Dokuz Eylül University, Turkey
suleysevinc@gmail.com

Ali Rıza Şişman
Department of Medical Biochemistry, Dokuz Eylül University, Turkey
aliriza.sisman@deu.edu.tr

Abstract: *Various studies use numerous probabilistic methods to establish a cause-effect relationship between a drug and a disease. However, only a limited number of machine learning studies on establishing cause-effect relationships can be found on the internet. In this study, we explore machine learning approaches for interpreting large quantities of multivariate patient-based laboratory data for establishing cause-effect relationships for critically ill patients. We adopt principal component analysis as a primary method to capture daily patient changes after a medical intervention so that the causal relationship between the medical treatments and the outcomes can be established. Model validity and stability are evaluated using bootstrap testing. The model exhibits an acceptable significance level with a two-tailed test. Moreover, results show that the approach provides promising results in interpreting large quantities of patient data and establishing cause-effect relationships for making informed decisions for critically ill patients. If fused with other machine learning and probabilistic models, the proposed approach can provide the healthcare industry with an added tool for daily routine clinical practices. Furthermore, the approach will be able to support clinical decision-making and enable effective patient-tailored care for better health outcomes.*

Keywords: *Capture variances, cause-effect relationships, causal inference, decision making, principal component analysis.*

Received January 25, 2021; accepted October 11, 2022
https://doi.org/10.34028/iajit/20/5/1

1. Introduction

Researches show that Randomized Controlled Trials (RCT) have long been accepted as the gold standard for medical research in establishing causal relationships. This process simply amasses a large number of subjects and randomly divides them into two groups, one exposed to the treatment under the study and the other one used as a control group. This process is performed repeatedly to prove exposing the patients to a certain treatment result in a better or alternate outcome. However, as Ketchersid [7] in his work, big data in nephrology: friend or foe? points out, there is an abundant number of demerits to this approach; the process is expensive, takes a long time to complete, and frequent exclusion of patients with some attributes, which in turn makes it difficult to generalize the results of the study. Moreover, the study also reveals that there is a strong urge for personalized healthcare these days, rather than applying clinical guidelines to every unique patient. In addition to the aforementioned drawbacks, the study [1] mentions that the randomization process makes patients and physicians uncomfortable. And this is where the possibility of using big data analytics and machine learning models becomes vital. The healthcare industry is starting to adopt machine learning and data analysis tools to push the boundaries. This effort primarily involves data analysis using a vast amount of healthcare data such as x-ray results, vaccinations, blood samples, vital signs, DNA sequences, current medications, past medical history, and much more. Artificial Intelligence (AI) and Machine Learning (ML) tools are expected to help in this regard and enhance healthcare quality and are believed to add further value to the healthcare industry. Furthermore, more researchers agree that machine learning is making healthcare smarter.

Numerous challenges remain in the advancement and improvement of machine learning tools in healthcare. Moreover, little or no observational study has been conducted to design ML tools to establish causal inferences from clinical laboratory data. The study by Stern and Price [14], states that this is due to the generalizability of ML model results remain questionable mainly in situations where ML fails to demonstrate causality due to the nature of the algorithms, and discover predictive patterns rather than causal relationships. Moreover, an ML model developed in one hospital setting might not be appropriate in different hospital settings unless causal inference tools were used in the development. The study [13] points out that multiple disciplines have

benefitted from the advancement of machine learning, however, the application of these tools has not been widespread in some areas such as causal inference. This is because sample sizes in RCT pose major limitations. However, if proper control of confounding and other matters is handled, observational data may reveal hidden insights. Moreover, the study also advocates using the mixture of both RCT and observational data for establishing causal inference.

Causality plays an important role in monitoring adverse drug events as well as risk factors for diseases with the help of Electronic Medical Records (EMR). A review study by Kleinberg and Hripcsak [8], presents graphical models and Granger causality as convenient frameworks for causal inference. The study also points out that more recent approaches such as temporal logic methods address some of the limitations of the above models. Also, the study denotes that we cannot fully automate causal inference from observational data without human involvement in the process. Numerous factors affect the cure or improvement of disease [11]. This is why multiple experiments need to be conducted to make meaningful statements regarding cause-effect relationships. Little or no exhaustive study has been conducted using ML models for causal inference. For instance, the study [9] extends Optimal Discriminant Analysis (ODA) for causal inference by framing the treatment-outcome relationship as a classification problem. The study concludes that ODA offers several benefits. Furthermore, the work [12] presents a patient-level ML method for causal inference for decision support for critically ill patients. Advances in data technology open new opportunities for more targeted queries regarding patient treatment for better patient-centered outcomes. In this paper, we explore the use of Principal Component Analysis (PCA) as a principal method for establishing causal relationships from longitudinal clinical laboratory data in Intensive Care Unit (ICU) settings. PCA is applied to capture daily changes from medical laboratory data and present results along with daily prescriptions so that cause-effect relationships can be established by healthcare practitioners.

2. Materials and Methods

The study used a subset of data extracted from the publicly available Medical Information Mart for Intensive Care III (MIMIC-III) dataset [5, 6]. Some studies conducted using the same database include [2, 4, 12, 15]. The top ten frequent diseases diagnosed in the hospital are selected for the study. Table 1 presents the selected diagnoses for the study. A total of 50 patients diagnosed with these diseases were selected for analysis proportionally. For instance, 9 and 8 patients diagnosed with sepsis and pneumonia respectively, are selected. Besides, their corresponding laboratory test results and available medical treatment data such as the prescriptions provided are extracted from the original database. We hypothesize that PCA can be able to capture the changes after medical intervention and a cause-effect relationship can be established between the interventions and their outcomes by healthcare practitioners. For this particular study Principal Components (PCs) that explain 99% of the variance in the data are retained.

The dataset contains laboratory tests conducted over a specified period. Numerous laboratory tests were conducted during the ICU stay, and in some cases, a test is conducted multiple times per day as presented in Table 2. Laboratory test dates are de-identified (from the source) according to the Health Insurance Privacy and Accountability Act (HIPAA) privacy rule i.e., 9/9/2120 for instance is not an error. It indicates a de-identified date according to the HIPAA rule not to disclose the actual patient laboratory test dates. However, date sequences are properly maintained. At the end of their hospital stay, the subjects were discharged from the hospital alive or dead to home or another healthcare unit.

Table 1. Results of the subject selection process.

Diagnosis	Count	Total Percentage (X)	Selected (X% of 50)
Sepsis	63	19	9
Pneumonia	55	16	8
Gastrointestinal Bleed	50	15	7
Fever	39	12	6
Congestive Heart Failure	26	8	4
Respiratory Failure	26	8	4
Coronary Artery Disease	25	7	4
Abdominal Pain	21	6	3
Chest Pain	16	5	2
Pancreatitis	15	4	2
Total	336	100	50

Non-numeric and variables with a single measured value (no variance) were excluded first. Scikit-Learn most_frequent strategy is employed for imputing missing values. The strategy replaces the missing values with the most frequent (mode) value of each variable. This process is applied to each selected patient separately. Then the data is standardized for analysis. This is followed by grouping daily laboratory tests as per the day they are collected and fed to the model for analysis. Finally, results were collected and model validity and stability are conducted using bootstrap testing.

Fifty percent (50%) of the total data is sampled for every bootstrap iteration. For model evaluation Total Variance Accounted For (TVAF) is computed as a statistic of interest. TVAF is equal to the sum of the Eigenvalues of the first *n* principal components. In our case *n* represents the number of principal components that make up 99% of the total variance in the data. This is followed by *r* number of iterations (r=1000) and statistical estimates (estimated TVAF) for each

iteration are computed. The statistical significance between the observed TVAF and the estimated TVAF is determined by comparing the p-value to a significance level. A significance level of α=0.05 is adopted as a rejection rule for this study. The alternative hypothesis that there is a statistical difference between the observed explained variance vs the estimated explained variance is tested against the null hypothesis that there is not a statistical difference between the observed explained variance and the estimated explained variance. A p-value is consulted to either accept or reject the null hypothesis. Since we used a two-tailed test the significance level α=0.05 will be split in half and put on both sides of the distribution and we are looking for a p-value of 0.025.

3. Results

A detailed retrospective analysis of the selected ICU patients is conducted. The selected target patients were diagnosed with different diseases as indicated in Table 1 and at the end of their stay, the subjects were discharged alive or dead. In the PCA analysis, principal components that make up to 99% of the overall variance in the data are retained in each analysis. For example, analysis results for a sample patient (diagnosed with coronary artery disease) shows the first eight PCs for the first day (9/9/2120-Figure 1-(a)) and the first seven PCs for the second day (10/9/2120-Figure 1-(b)) and the first five PCs for the third day (11/9/2120-Figure 1-(c)) that amount 99% of the variances explained in the data as depicted in Figure 1. Moreover, in most cases, the first principal components contain most of the variance. For instance, the first PC amounts to approximately 80% of the variance in the data on both day two and day three.

Besides, the results of a visualization and analysis environment to understand patient progression over time are presented in Figure 2. Through the use of two-dimensional plots, we allow users to explore how patients and their progression change over time (daily). Compared to existing techniques, our work provides additional flexibility in analyzing patient data and has the potential to be used in a real-time hospital setting.

Figure 2 shows the daily changes after a certain medical intervention. To interpret each principal component, we can examine the magnitude and direction of the coefficients for the original variables. Moreover, Table 4 presents the daily medical prescriptions given to the patient on the specified dates. This will allow the user to be able to establish cause-effect relationships.

In these results, on the first day (9/9/2120) (Figure 2-(a)), the first principal component has large positive associations with Lactate whereas the second principal component has large positive associations with Oxygen saturation. Also, we see that there is a change on the next day (10/9/2120) (Figure 2-(b)) in which the first principal component has large positive associations with Atypical Lymphocytes, Lymphocytes, and

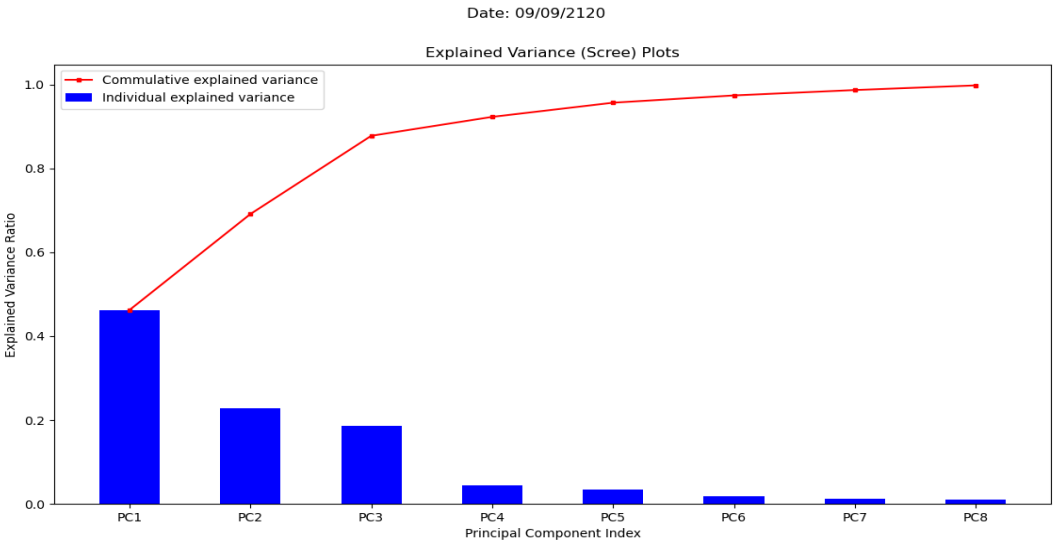
Table 2. Sample dataset content.

Date	A-aO_2	Calc. Total CO_2	Required O_2	Temperature	pCO_2	pH	pO_2
8/12/2166 23:45	540	26	88	38.8	47	7.34	137
8/13/2166 8:47		23		37.5	39	7.37	71
8/13/2166 17:19	590	26	96	38.3	57	7.25	75
8/13/2166 22:00	599	24	97	39.2	40	7.36	83
8/14/2166 0:19	537	25	88	39.2	40	7.38	72
8/14/2166 2:52	526	24	87	37.8	39	7.38	84
8/14/2166 4:01	472	23	79	37.4	37	7.38	68
8/15/2166 0:45	466	23	78	38.1	40	7.35	74

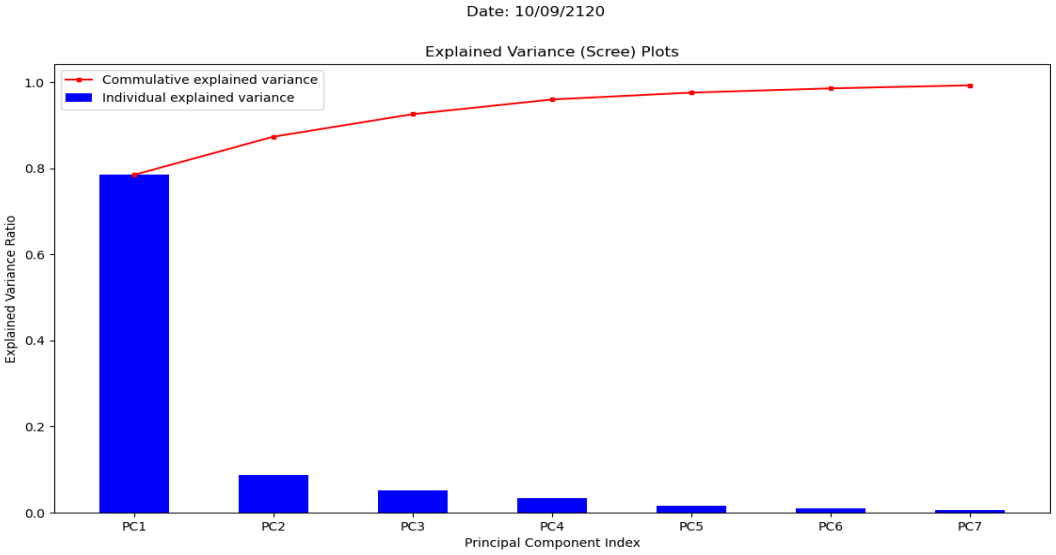
Neutrophils whereas the second principal component has large positive associations with Lactate. Moreover, on the third day (11/9/2120) (Figure 2-(c)) also shows there are changes. The first principal component has large positive associations with fibrinogen functional whereas the second principal component has large positive associations with calculated total CO2 and large negative associations with Oxygen saturation. The loading plot visually shows the results for the first two principal components.

We can observe that there are changes in magnitude and direction both on the second and third days of the ICU stay. Similarly, some laboratory tests such as Alkaline Phosphatase (ALP), Amylase, and Bands have a magnitude/contribution of zero throughout all the PCs, which indicates that they are insignificant on those dates, with no contribution/changes at all. Additional tabular information that describes the labels and associated values for Figure 2 is presented under Table 3.

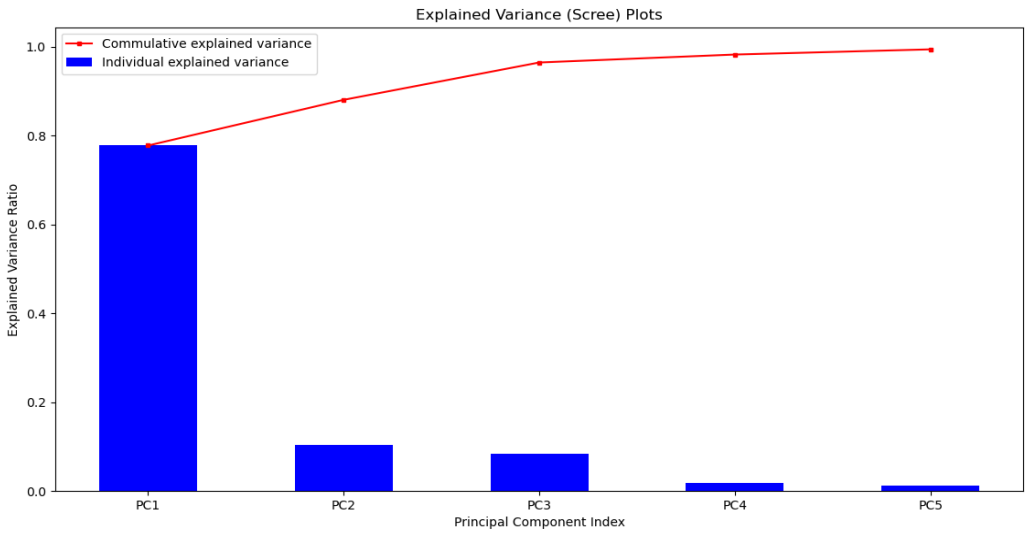
At this point, it is worth mentioning that, the patient has been given medical prescriptions on the specified dates (see Table 4). This may mean that the condition of the patient is either improving or worsening. Or it may also show if a medical treatment is working or not. Based upon this, a trained physician can be able to easily infer the implication and make a causal inference. The plots show the comparison and progression of successive daily-based contributions (negative or positive) of the laboratory tests under different PCs for a patient i.e., daily changes. They depict what changes happened on a specific day based on a treatment applied on the previous day. This can be used to decide on further steps that need to be carried out.



a) Day 1 results.

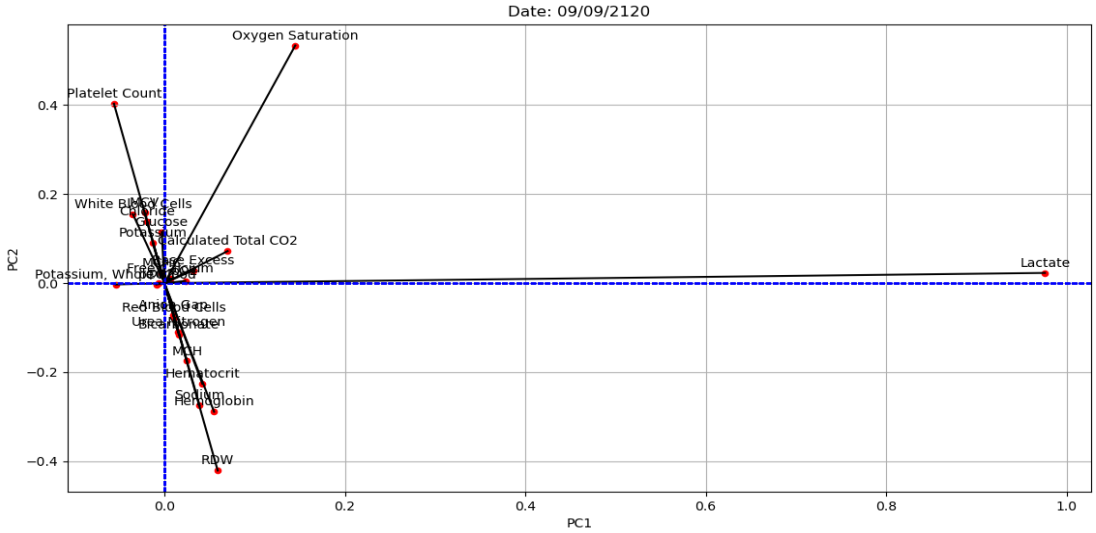


b) Day 2 results.

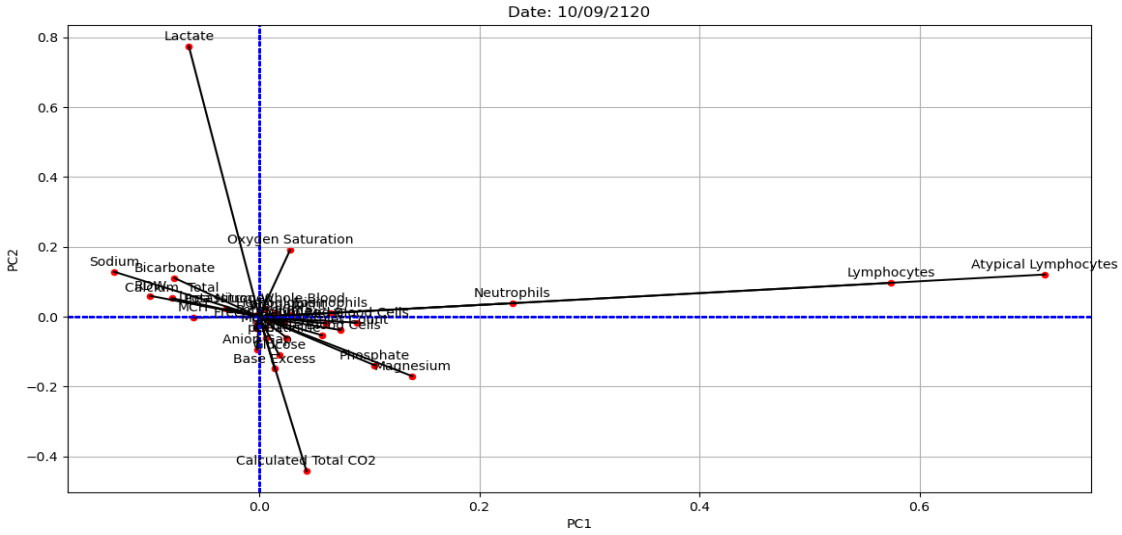


c) Day 3 results.

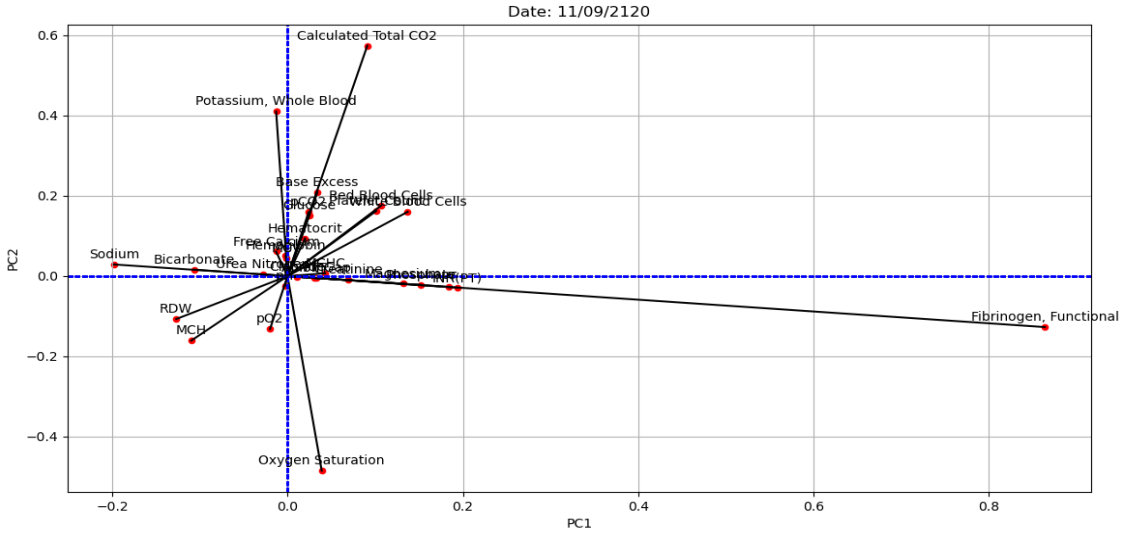
Figure 1. Principal Component Scree Plot ((a) Day 1 results, (b) Day 2 results and (c) Day 3 results).



a) Day 1 results.



b) Day 2 results.



c) Day 3 results.

Figure 2. Magnitude and direction of the original variables ((a) Day 1 results, (b) Day 2 results and (c) Day 3 results).

Table 3. Three-day principal component analysis results.

	9/9/2120		10/9/2120		11/9/2120	
	PC1	PC2	PC1	PC2	PC1	PC2
Anion Gap	0.01	-0.07	0.00	-0.09	0.03	-0.01
Atypical Lymphocytes	0.00	0.00	0.71	0.12	0.00	0.00
Base Excess	0.03	0.03	0.01	-0.15	0.03	0.21
Bicarbonate	0.02	-0.12	-0.08	0.11	-0.11	0.02
Calcium, Total	0.00	0.00	-0.08	0.05	0.00	0.00
Calculated Total CO2	0.07	0.07	0.04	-0.44	0.09	0.57
Chloride	-0.02	0.14	0.02	-0.02	0.01	0.00
Creatinine	0.00	0.00	0.03	-0.06	0.07	-0.01
Eosinophils	0.00	0.00	0.07	0.01	0.00	0.00
Fibrinogen, Functional	0.00	0.00	0.00	0.00	0.86	-0.13
Free Calcium	0.01	0.01	0.00	-0.02	-0.01	0.06
Glucose	0.00	0.12	0.02	-0.11	0.03	0.15
Hematocrit	0.04	-0.23	0.03	0.01	0.02	0.09
Hemoglobin	0.06	-0.29	0.02	0.01	0.00	0.05
INR(PT)	0.00	0.00	0.00	0.00	0.19	-0.03
Lactate	0.98	0.02	-0.06	0.78	0.00	0.00
Lymphocytes	0.00	0.00	0.57	0.10	0.00	0.00
MCH	0.03	-0.18	-0.06	0.00	-0.11	-0.16
MCHC	0.00	0.02	0.06	-0.02	0.04	0.01
MCV	-0.02	0.16	0.00	-0.03	0.00	0.04
Magnesium	0.00	0.00	0.14	-0.17	0.13	-0.02
Neutrophils	0.00	0.00	0.23	0.04	0.00	0.00
Oxygen Saturation	0.15	0.53	0.03	0.19	0.04	-0.49
PT	0.00	0.00	0.00	0.00	0.18	-0.03
PTT	0.00	0.00	0.00	0.00	0.03	-0.01
Phosphate	0.00	0.00	0.11	-0.14	0.15	-0.02
Platelet Count	-0.06	0.40	0.07	-0.04	0.10	0.16
Potassium	-0.01	0.09	0.01	-0.01	0.00	0.00
Potassium, Whole Blood	-0.05	0.00	0.01	0.02	-0.01	0.41
RDW	0.06	-0.42	-0.10	0.06	-0.13	-0.11
Red Blood Cells	0.01	-0.08	0.09	-0.02	0.11	0.18
Sodium	0.04	-0.28	-0.13	0.13	-0.20	0.03
Urea Nitrogen	0.02	-0.11	-0.03	0.02	-0.03	0.00
White Blood Cells	-0.04	0.16	0.06	-0.05	0.14	0.16
pCO2	-0.01	0.00	0.01	-0.06	0.02	0.16
pH	-0.01	0.00	0.00	0.01	0.00	-0.03
pO2	0.02	0.01	0.01	-0.03	-0.02	-0.13

It can be seen from Table 4 that similar medicines have been used on different dates. The usage of some of the medicines listed also extends beyond the dates presented as a reference in the plots. The drugs were used as a base or main drug type as indicated in the table. Finally, the dimensionality suggested by the scree plots (Figure 1) of the analysis is variant depending on the number of variables and samples in the dataset, corresponding to 99% of the explained variance in the data. Furthermore, bootstrapped p-values are reported as a measure of statistical significance. Since we used a two-tailed test, if our p-value is less than or equal to our anticipated significance level, our Null hypothesis will be rejected. The logic is that the p-value is the likelihood of us getting a random result outside of our 95% confidence interval. If the p-value is smaller than our alpha, which means it is unlikely that the result outside of our 95% was random, meaning that it was significant and shouldn't be ignored as an error. Whereas if the p-value is higher than our alpha, it means it is likely that the result outside of our 95% interval is random, so we shouldn't freak out and will fail to reject our null hypothesis. In this study since our mean p-value is 0.27 that means we are going to accept the null hypothesis and conclude that there is no statistical difference between the observed explained variance vs the estimated explained variance. This posits the stability and validity of the proposed model.

4. Discussion

Many factors must be considered when interpreting the results of any clinical laboratory test. Healthcare practitioners use normal reference ranges as guidelines for what is normal or abnormal. However, we believe that even if minor changes in successive measurements of patient vital signs are within normal reference ranges, fused with other vital signs, they may provide significant information for critically ill patients. The study presented a novel non-disease-specific model that can observe daily clinical changes and provide non-disease-specific analysis of patient progress over time. Moreover, by providing the daily prescriptions along with the analysis results a professional can be able to establish cause-effect relationships. Given these facts, the results can be used to decide what treatment or therapy to prescribe or which diagnosis to perform further. From the results, the ranked direction of changes (principal components) along with the contribution of each original variable can be observed. Fusing and presenting this with the daily prescriptions the patient has been given, a healthcare professional can be able to establish causal inference. In addition, using this information, a professional can be able to decide which laboratory test to perform further and/or exclude. It can also help decide which prescriptions to avoid or prescribe additional medicines. Moreover, each principal component is a linear combination of the original individual variables. With a closer look at this, a physician can be able to understand the combined effects of the original variables to make medical inferences.

Table 4. Associated daily medical prescriptions.

9/9/2120 - 18/10/2120		**10/9/2120 - 18/9/2120**		**11/9/2120 - 14/9/2120**	
Drug	Drug Type	Drug	Drug Type	Drug	Drug Type
5% Dextrose	BASE	5% Dextrose	BASE	Amiodarone	MAIN
Albumin 5%	MAIN	Amiodarone	MAIN	Furosemide	MAIN
Albuterol 0.083% Neb Soln	MAIN	D5W	BASE	5% Dextrose (EXCEL BAG)	BASE
Albuterol Inhaler	MAIN	Furosemide	MAIN	Norepinephrine	MAIN
Amiodarone	MAIN	Haloperidol	MAIN	0.9% Sodium Chloride	BASE
Ciprofloxacin IV	MAIN	Ipratropium Bromide Neb	MAIN		
D5W	BASE	Milrinone	MAIN		
Fluticasone Propionate 110mcg	MAIN	Norepinephrine	MAIN		
Furosemide	MAIN	Vasopressin	MAIN		
Heparin	MAIN	Xopenex	MAIN		
Hydrocortisone Na Succ.	MAIN				
Ipratropium Bromide Neb	MAIN				
Magnesium Sulfate	MAIN				
Milrinone	MAIN				
Oxycodone-Acetaminophen	MAIN				
Racepinephrine	MAIN				
SW	BASE				
Vasopressin	MAIN				
Xopenex Neb	MAIN				

The effectiveness of a treatment and the improvement or cure of a disease can be influenced by multiple factors. The observation in a single patient may suggest the likelihood of a new property of a drug, or a contrary effect on the patient. This may not insure it happened due to causality; however, it can help rule out the possibility of coincidence between the clinical interventions and the outcomes. It is through causality we can be able to infer the behavior of a medical treatment. Hence the vitality of causality in medicine. In this study causal inference refers to the process of uncovering causal relationships from medical treatment data. This does not mean that we remove the need for human judgment, but rather help healthcare professionals validate the results and make informed decisions. It is worth mentioning that no matter how detailed or clean the data is, machine learning models cannot eliminate unmeasured or unprecedented factors coincident with a particular intervention that may explain an apparent outcome change. Given all the daily vital sign measurements for ICU patients, a medical professional has to make the most out of it for saving the patient. It is because of these facts; we propose a tool to assist the users in daily clinical routine practices. Results showed that the approach if fused with other machine learning models, it presents a bright future for real-time patient monitoring in ICU settings. It can also help anticipate and avoid life-threatening conditions from happening proactively.

In any statistical model, where PCA is not an exception, model validation is imperative to generalize the results of a proposed model. Studies [10] advocate nonparametric methods such as permutation and bootstrap tests as theoretically better matches for the nonparametric nature of PCA. By applying these nonparametric methods different matrices can be generated by permutation or resampling of the data, and their Eigenvalues and Eigenvectors will no longer be the same. Given these facts, the study applied bootstrap testing for model validation. In experimental studies with treatment and control groups, results may be analyzed by simply regressing the outcome on a treatment group indicator variable to estimate treatment effects [9]. However, this is impossible to achieve in observational studies. Our results demonstrate that the proposed method can be combined with other strategies to improve causal inference for critically ill patients. However, it is worth mentioning that we did not address the issue of whether the change happened due to the introduction of treatment or by chance. Besides, it is worth stating that, the medical protocols used and input events recommended to the patients during the ICU stay were not considered in this study. The proposed approach may help physicians feel confident about their decisions. However, we would like to emphasize that any tool developed out of this approach is not meant to replace or undermine the skills and instincts of medical practitioners. It is only meant to provide an alternative or a second eye for the users in presenting hidden insights. When fully realized, machine learning models could analyze longitudinal electronic health records to provide a second eye to the healthcare professionals. Principal component analysis is an interesting approach for patient monitoring because it holds several advantages in observational and exploratory studies. These advantages include it being easy to compute. It speeds up other machine-learning algorithms. It can also counteract the issues of high-dimensional data. In addition, PCA improves the performance of the ML algorithm as it eliminates correlated variables that don't contribute to any

decision-making. PCA results in high variance and thus improves visualization.

5. Conclusions

Machine Learning (ML) technology is a technique for data analysis utilized in the medical field for disease diagnosis, therapy, and treatment [3]. Machine learning models can detect hidden patterns and relationships in certain diseases from electronic health records. In this aspect, these models can be regarded as the second pair of eyes in monitoring patient health. Moreover, they can also be able to help healthcare professionals make informed, timely, lifesaving, and effective decisions. They can also help save and avoid unnecessary wastage of both monetary and material resources. Also, the decision-making process in clinical medicine can be supported and facilitated with the appropriate selection and application of relevant machine learning models. This study showed that PCA can be used as part of a tool for establishing cause-effect relationships from medical treatment data. Nevertheless, to assist researchers and stakeholders in the field, it will be of great paramount if the proposed method is fused with other machine learning frameworks and models for a better and full-fledged application. Furthermore, additional investigation will be conducted to fuse this approach with other probabilistic and machine-learning approaches to provide a better and more robust tool for medical practitioners. Finally, the study took a great deal of time to come up with the selected machine learning models for establishing cause-effect relationships from medical treatment data. However, we believe that there are areas that still need a great deal of work and improved upon and that we consider as future works.

Conflict of Interest

The authors declare no conflicts of interest.

Ethics Approval

The study used a publicly available, deidentified dataset, hence no ethics approval was required.

References

[1] Angus D., "Fusing Randomized Trials with Big Data: The Key to Self-Learning Health Care Systems*?*," *Journal of American Medical. Association*, vol. 314, no. 8, pp. 767-768, 2015. doi: 10.1001/jama.2015.7762.

[2] Dai Z., Liu S., Wu J., Li M., Liu J., and Li K., "Analysis of Adult Disease Characteristics and Mortality on MIMIC-III," *Journal of PLoS One*, vol. 15, no. 4, 2020. https://doi.org/10.1371/journal.pone.0232176

[3] Elbasi E. and Zreikat A., "Heart Disease Classification for Early Diagnosis based on Adaptive Hoeffding Tree Algorithm in IoMT Data," *The International Arab Journal of Information Technology*, vol. 20, no. 1, pp. 38-48, 2023. DOI:10.34028/iajit/20/1/5

[4] Feng M., Mcsparron J., Kien D., Stone D., Roberts D., Schwartzstien R., Baron A., and Celi L., "Transthoracic Echocardiography and Mortality in Sepsis: Analysis of the MIMIC-III Database," *Intensive Care Medicine*, vol. 44, no. 6, pp. 884-892, 2018. doi: 10.1007/s00134-018-5208-7

[5] Goldberger A., Amaral L., Glass L., Hausdorff J., Ivanov P., Mark R., Mietus J., Moody G., Peng C., and Stanley E., "PhysioBank, PhysioToolkit, and PhysioNet: Components of A New Research Resource for Complex Physiologic Signals," *Journal of the American Heart Association*, vol. 101, no. 23, 2000. doi: 10.1161/01.cir.101.23.e215.

[6] Johnson E., Pollard T., Shen L., Lehman L., Feng M., Ghassemi M., Moody B., Szolovits P., Celi L., and Mark R., "MIMIC-III, a Freely Accessible Critical Care Database," *Journal of Nature Scientific Data*, vol. 3, pp. 1-9, 2016. doi: 10.1038/sdata.2016.35.

[7] Ketchersid T., "Big Data in Nephrology: Friend or Foe*?*," *Blood Purification*, vol. 36, no. 3-4, pp. 160-164, 2014. doi: 10.1159/000356751.

[8] Kleinberg S. and Hripcsak G., "A Review of Causal Inference for Biomedical Informatics," *Journal Biomedical Informatics*, vol. 44, no. 6, pp. 1102-1112, 2011. doi: 10.1016/j.jbi.2011.07.001.

[9] Linden A. and Yarnold P., "Combining Machine Learning and Matching Techniques To Improve Causal Inference in Program Evaluation," *Journal of Evaluation in Clinical Practice*, vol. 22, no. 6, pp. 864-870, 2016. DOI:10.1111/jep.12592

[10] Linting M., Os B., and Meulman J., "Statistical Significance of the Contribution of Variables to the PCA Solution: An Alternative Permutation Strategy," *Psychometrika*, vol. 76, no. 3, pp. 440-460, 2011. DOI: 10.1007/S11336-011-9216-6

[11] Monleon-Getino A. and Canela-Soler J., "Causality in Medicine and its Relationship with the Role of Statistics," *Biomedical Statistics and Informatics*, vol. 2, no. 2, pp. 61-68, 2017. DOI:10.11648/j.bsi.20170202.14)

[12] Pirracchio R. Cohen M., Malenica I., Cohen J., Chambaz A., Cannesson M., Lee C., Resche-Rigon M., and Hubbard A., "Big Data and Targeted Machine Learning in Action To Assist Medical Decision in the ICU," *Anaesthesia Critical Care and Pain Medicine*, vol. 38, no. 4,

pp. 377-384, 2019. doi: 10.1016/j.accpm.2018.09.008.
[13] Rose S. and Rizopoulos D., "Machine Learning for Causal Inference in Biostatistics," *Biostatistics*, vol. 21, no. 2, pp. 336-338, 2020. https://doi.org/10.1093/biostatistics/kxz045
[14] Stern A. and Price W., "Regulatory Oversight, Causal Inference, and Safe and Effective Health Care Machine Learning," *Biostatistics*, vol. 21, no. 2, pp. 363-367, 2020.
[15] Vincent J., Nielsen N., Shapiro N., Gerbasi M., Grossman A., Doroff R., Zeng F., Young P., and Russell J., "Mean Arterial Pressure and Mortality in Patients with Distributive Shock: A Retrospective Analysis of The MIMIC-III Database," *Ann. Intensive Care*, vol. 8, no. 1, 2018. doi: 10.1186/s13613-018-0448-9.

Mohammed Abebe Yimer graduated with B.Sc. Degree in Computer Science and Information Technology from Arba Minch University, an MSc Degree in Computer Science from Addis Ababa University, and a Ph.D. degree in 2021 from the Computer Engineering Department of Dokuz Eylul University (DEU), Izmir, Turkey. He has been employed and working as a lecturer at Arba Minch University, Ethiopia since 2007. Currently serving as an Assistant Professor at the Faculty of Computing and Software Engineering of Arba Minch University, Ethiopia. His research interest includes natural language processing, artificial intelligence, machine learning, deep learning, and data analytics.

Özlem Aktaş received a B.Sc. degree in 2003, an M.Sc. degree in 2005, and a Ph.D. degree in 2010 from the Computer Engineering Department of Dokuz Eylul University (DEU), Izmir, Turkey. She worked as a Computer Engineer in Gonen Information Technologies in 2003, worked as a Research Assistant from 2004-2009, and has been working as Assistant Professor since 2010 at Dokuz Eylul University Department of Computer Engineering. Her research interests include natural language processing, artificial intelligence, machine learning, and software engineering. She has many papers in journals and conferences in these interest areas since 2003.

Süleyman Sevinç obtained his BSc Degree from Istanbul Technical University in 1983. He completed his MSc and a Ph.D. degree from the University of Arizona, the USA between 1986-1988. He has been serving in multiple universities as an academician both in Turkey and Australia on different levels and capacities. He has been working at Sydney University, Celal Bayar University, Halic University, and Dokuz Eylul University as a professor. Moreover, he has been appointed and has been serving as Industry 4.0 coordinator of Dokuz Eylul University between the years 2017-2019. Currently, he is serving at Labenko Bilişim A.Ş. Izmir-Turkey both as a co-founder and CEO. Prof. Sevinç has numerous articles published in different international and reputable journals and has multiple citations. His main focus and interest are the automation of medical laboratories, hospital laboratory information systems, robotics, and the application of artificial intelligence in medical laboratories.

Ali Rıza Şişman is a Clinical Biochemistry specialist with 20 years of experience; He is a lecturer in the department of medical biochemistry of Dokuz Eylul University Faculty of Medicine. Dr. Şişman has 69 articles and 860 citations in WEB of Science, with an H index of 18. His main focus is quality management in medical laboratories, training of medical laboratory residents, laboratory automation, laboratory information systems, and artificial intelligence in medical laboratories. His publications, projects, and other academic activities can be accessed from the official website of Dokuz Eylül University: http://debis.deu.edu.tr/akademik/index.php?cat=3&akod=19990661