BALKAN JOURNAL OF ELECTRICAL & COMPUTER ENGINEERING, Vol. 9, No. 2, April 2021

Monitoring Medical Interventions for Multidimensional Evaluation of Changes in Patient Test Results with Principal Component Analysis

Mohammed Abebe Yimer*, Özlem Aktaş, Süleyman Sevinç, Ali Rıza Şişman, Oktay Yildirim and Eminullah Yaşar

***Abstract*—Medical doctors of today are challenged with increasingly large volumes of high-dimensional, heterogeneous, and unstructured data from various sources that pose significant challenges for manual analysis. However, this unstructured data is mainly vital for decision making but there exists a shortage of intelligent tools to extract the hidden knowledge. Given these facts, the application of machine learning methods in healthcare is a growing phenomenon. This paper explores machine learning approaches for interpreting large quantities of continuously acquired, multivariate patient-based medical laboratory data, in intensive care unit (ICU) settings. The research hypothesizes that principal component analysis (PCA) can be able to capture the changes in the outcomes after a medical intervention. We adopted PCA as the main method, to observe and capture the daily changes for intensive care unit patients. The approach will be able to inform the physicians, which laboratory tests are exhibiting variances after an intervention, and their associated epiphenomenon. This can be used as a clue to make decisions on which treatment or diagnosis to apply further. Experimental analysis results indicate that PCA was able to capture patient progression in terms of variances. Permutation tests for the validity and stability of the model exhibit an acceptable significance level with a p-value of 0.001. Results showed that the approach provides promising results for interpreting large quantities of patient data for establishing a cause-effect relationship from medical interventions and be used as an early warning system. The study retrospectively demonstrated the capability of PCA to monitor and provide an alert to the clinicians about the patient's changing conditions, thereby providing opportunities for timely interventions. If coupled with other machine learning models, the approach can also be able to support clinical decision making and enable effective patient-tailored care for better health outcomes.**

Index Terms—**Capture variances, Decision making, Intensive care unit, Medical intervention, Patient Care, Principal component analysis.**

MOHAMMED ABEBE YIMER, is with Department of Computer Engineering University of Dokuz Eylül University, Izmir, Turkey, (e-mail: moshehio@gmail.com).
https://orcid.org/0000-0003-0622-4841

ÖZLEM AKTAŞ, is with Department of Computer Engineering University of Dokuz Eylül University, Izmir, Turkey, (e-mail: aktas.ozlem@deu.edu.tr).
https://orcid.org/0000-0001-6415-0698

SÜLEYMAN SEVINÇ, is with Department of Computer Engineering University of Dokuz Eylül University and Labenko Bilişim A.Ş, Izmir, Turkey, (e-mail: suleysevinc@gmail.com).
https://orcid.org/0000-0001-9052-5836

ALI RIZA ŞIŞMAN, is with Department of Medical Biochemistry University of Dokuz Eylül University, Izmir, Turkey, (e-mail: aliriza.sisman@deu.edu.tr).
https://orcid.org/0000-0002-9266-0844

OKTAY YILDIRIM, is with Department of Computer Engineering University of Dokuz Eylül University, Izmir, Turkey, (e-mail: oktay.yildirim@deu.edu.tr).
https://orcid.org/0000-0001-7281-3623

EMINULLAH YAŞAR, is with Department of Computer Engineering University of Dokuz Eylül University, Izmir, Turkey, (e-mail: eminullah.yasar@gmail.com).
https://orcid.org/0000-0003-2298-4808

Manuscript received August 12, 2020; accepted February 6, 2021.
DOI: 10.17694/bajece.782510

I. INTRODUCTION

Several research studies postulate that the development of machine learning systems in healthcare is paramount both for the healthcare industry as well as the patients. The majority of healthcare systems nowadays use a basic visual representation of historical data for analysis by the doctors. This type of viewing is important however, this becomes overwhelming with the increasing number of patients and a vast amount of collected data. This necessitates the development of intelligent tools or systems that can extract insight and semantics from this bulk of data for decision-making processes.

Widanagamaachchi *et al.* [1], presents a system for interactive visualization and exploration of patient progression overtime for decision making using hierarchical clustering and tracking graph. Wang *et al.* [2], also developed a web-based visual mining system that supports explorative analysis of

high dimensional categorical electronic medical records (EMR) for chronic kidney disease (CKD). The proposed method uses Ochiai coefficients to compute patient similarity-based on seventeen CKD factors. This type of exploring and viewing is important, however, it won't be enough to uncover latent information. In addition, early studies such as [3 - 5] use generic patient vital signs to compute early warning scores to predict patient mortality for deteriorating patients. These systems are solely based on physiological parameters of the patient's vital signs such as heart rate, breathing rate, and systolic blood pressure. Moreover, these systems use the sum of assigned points of these vital signs from a subjectively fixed normal range to identify patients that are deteriorating but also can have their outcome changed by timely intervention. The aforementioned studies use ordinary methods and/or patient vital signs to monitor patient progressions. However, recent advancements in machine learning and EMR paves the way for the development of intelligent tools for medical data analysis to help in decision making. In line with this, numerous predictive and analytic models have been presented.

Ye *et al.* [6] applied a tree-based random forest algorithm on data collected from EMRs to predict patients at high risk of intra-hospital mortality and achieved a c-statistics of 0.884. On the other hand, the study by Cai *et al.* [7], presents a predictive model for real-time predictions of the length of stay (LOS), mortality, and readmission for inpatients from electronic health records (EHR). The study employed Bayesian network model to estimate the likelihood of a patient being in one of the following states; at home, in the hospital, or dead, and achieved an average daily accuracy of 80% and an AUROC of 0.82. Furthermore, other highly investigated categories of machine learning models in healthcare include disease-specific predictive models. For example, the work [8], performed experimental studies on data collected on mild cognitive impairments (MCI). The work conducted a comparative study on methods such as decision trees, different statistical t-tests to predict the chance of a patient being positive or negative. The study [9] also demonstrates the use of machine learning methods for predicting acute kidney injuries (AKI) from EHR for better assessment of existing and novel interventions to provide vital treatment. Also, the studies [10, 11] explored the use of PCA as a dimensionality reduction technique and proved that it helps improve both classification accuracy and model training time for classification tasks. Moreover, the authors [11] suggest the use of this approach as part of medical devices for non-invasive, inexpensive decision-making tool. There are a tremendous amount of studies and investigations conducted on predictive models for patient mortality and disease-specific predictions. However, there is little or no investigation conducted on machine learning models to extract hidden knowledge or semantics from electronic health records for decision making. Because of these facts, this research focuses on a non-disease specific model for observing outcome changes after a medical intervention in ICU settings. The proposed approach can serve as an EMR-based early warning system that can be used as part of a daily routine clinical practice.

II. Materials and Methods

The study used a subset of data extracted from the publicly available Medical Information Mart for Intensive Care-III (MIMIC-III) v1.4 database [12, 13]. Some studies conducted using this database include [14 - 16]. The MIMIC-III dataset contains comprehensive, granular, deidentified ICU EHRs collected from hospital medical information systems (both patient bedside workstations and hospital archives). The data is collected from a single tertiary teaching hospital between the years 2001 and 2012. The dataset includes patient information that falls into several categories such as general, physiological, medications, fluid balance, notes, and reports. For our intended research, a total of 1,410 patients with a hospital LOS greater than or equal to 30 days is selected. Next, for each of these patients corresponding demographic and clinical data (laboratory data) is extracted from the original database. After careful inspection of the extracted dataset 1,306 (out of 1410) patients having at least 10 observations were selected for further analysis. This threshold is chosen to have extended observations and to ascertain that the approach captures the intended changes. Table I describes general information about the dataset used.

TABLE I
USED DATASET GENERAL INFORMATION

Gender	Count	Average LOS (in days)
Male	732	55.2
Female	574	57.78
Total	1306	56.33

The dataset used for this work contains laboratory tests conducted over the specified period for multiple laboratory tests for each patient separately. The study conducted a detailed retrospective analysis on 1,306 ICU patients with a minimum hospital LOS of 30 days. The selected target patients were diagnosed with different diseases such as sepsis and pneumonia, with a mean hospital LOS of 56.33 days. At the end of their stay, the subjects were discharged alive or dead to home or another healthcare unit. Table II presents sample data used for a sample patient. Laboratory test dates are de-identified (from the source) according to the Health Insurance Privacy and Accountability Act (HIPAA) privacy rule i.e., 2/26/2191 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.

The study hypothesizes that PCA can be able to capture the changes that may happen due to medical intervention. PCA is commonly used for linear dimensionality reduction and exploratory purposes through variance maximization. The intuition behind PCA is to use a special coordinate system that depends on the cloud of points. The axes are placed in the direction of the highest variance of the points to maximize the

variance along that direction [15]. This intuition can help us determine and show the patient's laboratory result exhibiting the highest variances after applying a certain medical intervention. Moreover, if applied to longitudinal data, it may help us explore patient progression from time T1 to T2. So that appropriate treatment or therapy can be prescribed or further diagnosis can be advised. In addition, each principal component (PC) is a linear combination of the original individual variables. This can be used to see the effect and contribution of each laboratory test in that direction. The succeeding PCs try to capture the next highest variances left out by the preceding PCs. This may show epiphenomenon or parallel medical events or conditions happening. Most studies use PCA as preprocessing for classification tasks [10-11]. However, based on the intuition of how PCA works, we believe that it can be used as part of a tool for early warning of medical conditions by showing laboratory tests with changes. To evaluate our primary hypothesis, a retrospective analysis was performed using PCA to monitor patient progress.

For this analysis, first non-numeric variables and variables with a single measured value (having no variance) throughout the ICU stay are discarded. Since the analysis is patient-based missing values are imputed using the most-frequent strategy along each column. This is followed by data standardization for PCA processing. Then the records are grouped based on the date the laboratory test is taken. A baseline of at least four observations should be available on a specific day to start or apply the analysis. Otherwise, those observations are merged with the next day's test results. This process is repeated iteratively until we achieve the minimum (4 observations) amount of observations for analysis. Finally, the relevant patient data were fed into the model and principal components were computed and visualized. Once this is complete, model validation and stability test are conducted.

To evaluate the validity and stability of the proposed approach, a permutation test sometimes known as randomization test is employed without relying on a specific probability model. Total variance accounted for (TVAF) is computed as the statistics 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 permutation (r=1000) and statistical estimates (estimated TVAF) for each permutation is 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 α=0.05 is adopted as a rejection rule for this study. The alternative hypothesis that the estimated TVAF values do not deviate significantly from the observed TVAF value is tested against the null hypothesis that it does. If $p<\alpha$, the result is marked significant, and H0 is rejected.

TABLE II
SAMPLE DATASET CONTENT

Date	Anion Gap	Bicarbonate	Bilirubin, Direct	Bilirubin, Indirect	Bilirubin, Total	Chloride
2/26/2191 16:20	18	16	0.2	4.2	4.4	114
2/28/2191 0:30			0.3	4	4.3	115
3/1/2191 4:30	19	15	0.3	3.8	4.1	106
3/1/2191 13:55	19	13				110
3/2/2191 5:20	23	14	0.7	2.7	3.4	106
3/2/2191 12:45	23	13				105
3/3/2191 4:30	17	21	0.3	4.6	4.9	99
3/3/2191 23:35	21	20	0.3	6	6.3	97

III. RESULTS

In the PCA analysis, principal components that make up 99% of the overall variance in the data are retained for each analysis. For example analysis results for a sample patient (frequently diagnosed for respiratory failure) shows the first two PCs for the first day and the first three PCs for the next day that amount to approximately 99% of the variances explained in the data (Fig. 1).

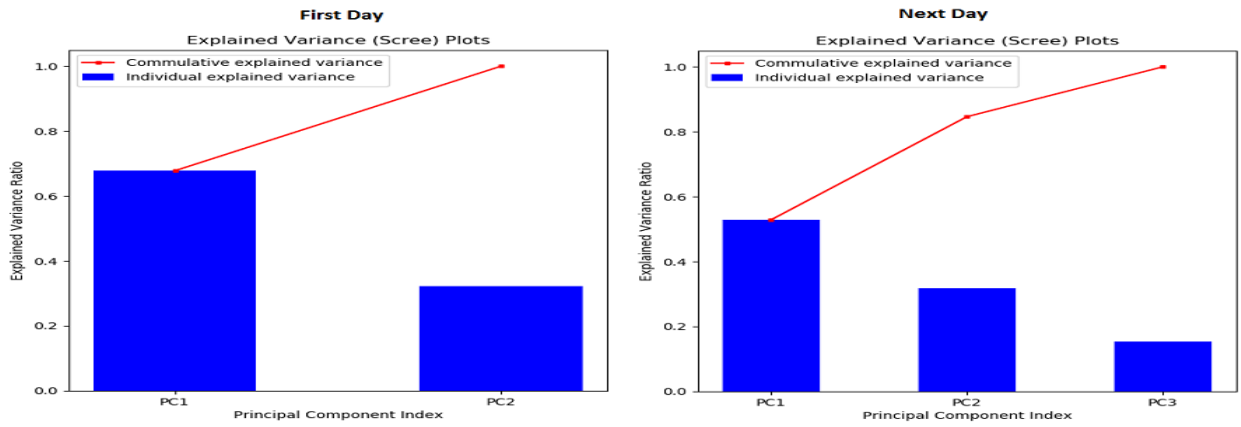

Fig. 1. Sample Principal Component Scree Plot

In addition, the contribution of each variable to a PC can be tabulated (see Table III) or presented using different formats for ease of use. For instance, Fig. 2 represents a 2-dimensional plot representation of variable loadings/contributions of the original variables under their respective principal components.

TABLE III
SAMPLE VARIABLE CONTRIBUTIONS/LOADINGS

PC.No	Chloride	Glucose	Oxygen	PT	Phosphate	Sodium	pCO2	pH
PC1	0.394	0.313	0.028	0	0.394	0	-0.193	0.167
PC2	0.114	0.138	0.078	0	0.114	0	0.357	-0.397
PC3	0	0.017	-0.106	0	0	0	0.096	-0.03
PC4	-0.041	0.007	-0.439	0	-0.041	0	0.26	0.029
PC5	0.023	-0.021	0.605	0	0.023	0	0.285	-0.346

Table III shows sample variable loadings per principal components for a specific laboratory test date. The larger the absolute value of the coefficient is, the more important the corresponding variable is in calculating the principal component. Generally, how large the absolute value of a coefficient has to be to consider it significant is subjective. For instance, we can see from Table III that PT and Sodium do not contribute to any of the PCs, and this indicates the insignificance of these tests on that specific day. On the other hand, Phosphate and Chloride contribute the highest under PC1. Moreover, this tabular information can be presented using other representations for ease of use as depicted in Fig. 2 and show the daily changes after a certain treatment or intervention. To interpret each principal components, we can examine the magnitude and direction of the coefficients for the original variables. The plot on Fig. 2(a), (b), and (c) visually shows the component loading for day one, two, and three of the analysis, respectively. Note that laboratory tests with zero loadings are removed from the plots.

In these results, for day one of the ICU stay (Fig. 2(a)), the first principal component has large positive associations with MCHC, Potassium, White Blood Cells, and PTT. The second component has large positive associations with MCH, MCV, Red Blood Cells, Hematocrit, and Hemoglobin. The loading plot visually shows the results for the first two components.

The third component has large negative associations with Base Excess & pH and large positive associations with Hematocrit (Calculated), Potassium (Whole Blood), and Hemoglobin (see Table IV). Moreover, we can observe that there are changes in magnitude and direction on the second and third days of the ICU stay (Fig. 2(b) and (c)). This may mean that the condition of the patient is either getting better or worse. Or it may also show if medical treatment is working or not. The plots show 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 or in relation to a treatment applied on the previous day. This can be used to decide on further steps that need to be carried out. Based upon this, a trained physician can be able to easily infer the implication and make an informed decision.

The corresponding coefficient values for Fig. 2 are presented in Table IV below only for the top five variables for the first three PC along with changes for three consecutive days. As it can be seen from the table, MCHC, Potassium, White Blood Cells and PTT equally contribute the highest under PC1. On the other hand, MCH, MCV, and Red Blood

Cells equally contribute the highest under PC2 for day 1. Besides, Hematocrit (Calculated), Base Excess, and pH contribute the highest under PC3 for the same day.

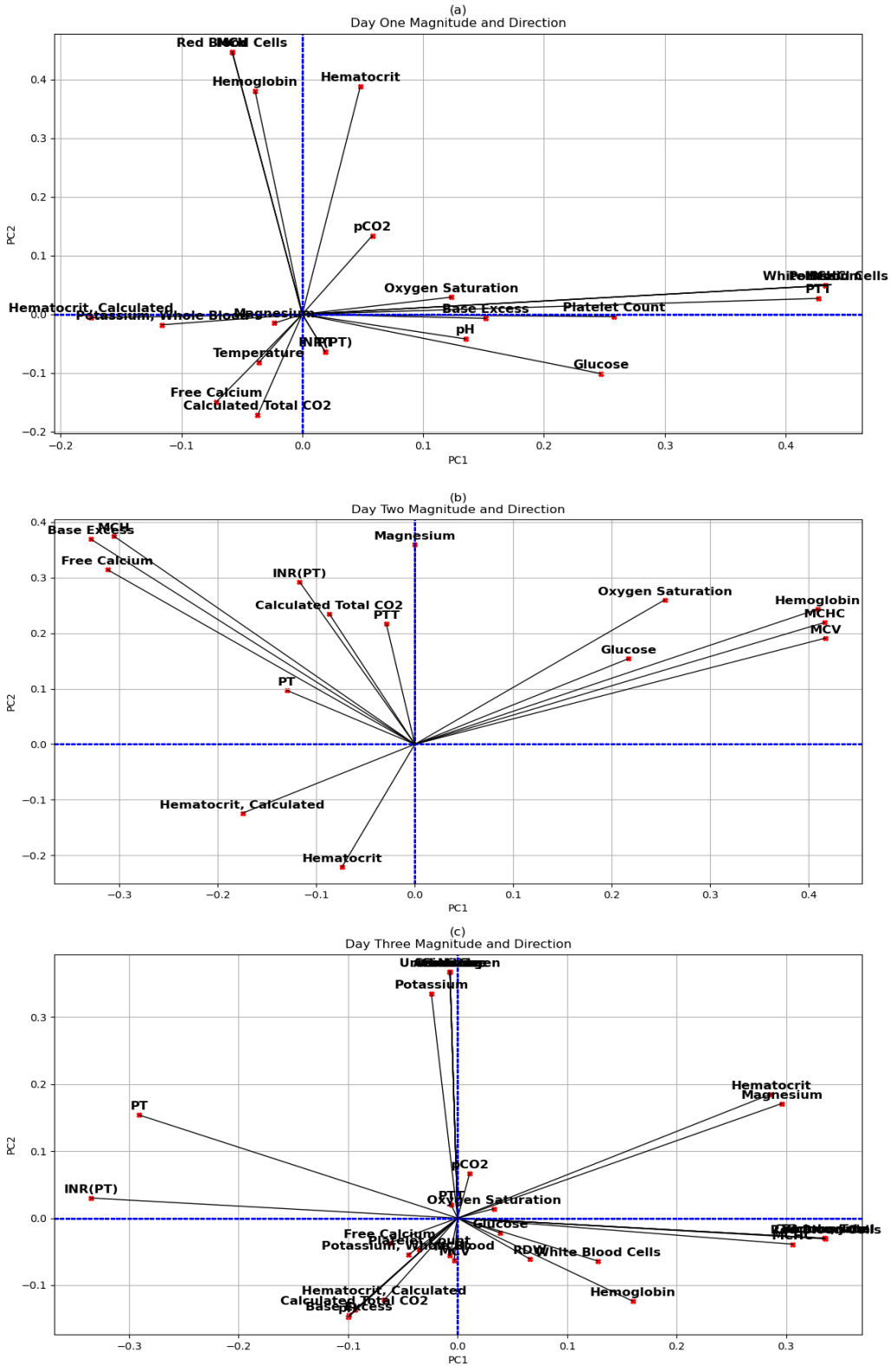

Fig. 2. Magnitude and direction of the coefficients for the original variables

TABLE IV
TOP THREE PC DAILY CHANGES

Day	PC1		PC2		PC3	
1	MCHC	0.43	MCH	0.45	Hematocrit, Calculated	0.46
	Potassium	0.43	MCV	0.45	Base Excess	-0.45
	White Blood Cells	0.43	Red Blood Cells	0.45	pH	-0.40
	PTT	0.43	Hematocrit	0.39	Potassium, Whole Blood	0.28
	Platelet Count	0.26	Hemoglobin	0.38	Hemoglobin	0.26
2	PTT	0.42	Oxygen Saturation	0.38	Hemoglobin	0.56
	PT	0.42	Calculated Total CO2	0.37	Free Calcium	0.52
	INR(PT)	0.41	Platelet Count	0.36	Hematocrit, Calculated	0.49
	Calculated Total CO2	-0.33	Glucose	0.31	Magnesium	-0.21
	Glucose	-0.31	Magnesium	0.29	pH	0.20
3	Red Blood Cells	0.34	Anion Gap	0.37	PTT	0.34
	Bicarbonate	0.34	Chloride	0.37	MCV	0.33
	Calcium, Total	0.34	Creatinine	0.37	RDW	0.30
	INR(PT)	-0.34	Sodium	0.37	pH	-0.29
	Vancomycin	0.34	Urea Nitrogen	0.37	MCH	0.29

Finally, the dimensionality suggested by the scree plots of the analysis is variant depending on the number of variables and samples we have in the dataset, corresponding to 99% of the explained variance in the data. In the end, the stability and validity of the model for the overall analysis were confirmed by the p-values of estimated TVAFs. The estimated statistics achieved a significance level of 0.001 which is much less than the predetermined α. This proves that the alternative hypothesis is accepted which posits the approach's stability and validity.

IV. Discussion

Many variables must be considered when interpreting the results of any medical laboratory test. In clinical medicine, physicians use normal reference ranges as guidelines of what is normal or abnormal. However, using normal reference ranges for ICU patients may not be the most sought out strategy [16]. Furthermore, we believe that even if minor changes of successive measurements of patient vital signs are within normal reference ranges, in conjunction with other laboratory test results, those changes may reveal significant information. This in turn can be used to make informed decisions on a drug, therapy, or further diagnosis. In this study, we presented a model for observing changes in outcome after clinical interventions in ICU settings. To the best of our knowledge, this is a novel non-disease specific model that can observe and capture clinical changes and is built to provide non-disease specific analysis of patient progress over time. The results can be used to decide what treatment or therapy to prescribe or which diagnosis to perform further. From the results, we can be able to see which laboratory tests are contributing more and their combined directions. This insight can be used to decide; 1) on which laboratory tests to perform further analysis and/or omit from further analysis so that redundant and anomalous tests will be avoided and 2) it can also be used to decide what prescription to provide or avoid, for the benefit of the patient. 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 know or judge the combined effects of those variables and make medical inferences. Sometimes it may also be necessary to see which laboratory test is having a change more often than the others so that precautions can be taken, and this can also be achieved with the proposed approach. However, it is worth noting that we did not and cannot address the issue of whether the change happened due to the introduction of a treatment or by chance in this setting.

In any statistical model, where PCA is not an exception, model validation is imperative to generalize the results of a proposed model. However, Lebart [17] points out that conventional analytical approaches are both unrealistic and analytically intricate for computing precision estimates of models such as PCA. Parametric methods have been tried for PCA model validation [18], alternatively, the study [19] advocates nonparametric methods as theoretically better matches for the nonparametric nature of PCA. By applying nonparametric methods such as permutation tests or bootstrap tests, different matrices can be generated by permutation or resampling of the data, and their Eigen values and Eigen vectors will no longer be the same. In view of these facts, the study used permutation testing for model validation. Given all the bulk of data, the physician has to make the most out of it and try saving the lives of the patients. With this in mind, we proposed a model that can help the physicians as part of their daily routine clinical practice. The approach can be used to

look at the results of an ICU patients' treatment from different perspectives. Furthermore, results showed that the approach, if coupled with other machine learning models, can be able to provide a promising future for real-time, non-invasive patient monitoring and early warning system in ICU settings. However, it is worth mentioning that, the medical protocols used, prescriptions provided and input events recommended to the patients during the ICU stay were not taken into consideration in this study. We believe that if included, any one of these inputs can cause certain variables to change.

Based on the research results, we can point out that with the appropriate selection of relevant algorithms, applying machine learning models on EMRs can provide key insights for medical practitioners to facilitate the decision making processes. With the aid of the proposed model, inexpert physicians may feel better confidence about their decisions. Nevertheless, this approach cannot replace or undermine the diagnostic skills and professional instincts of medical practitioners. The results of this research can provide an alternative means for the medical practitioners to consider only the significant factors instead of going through the whole patient data. Moreover, since the study used ICU admissions data from a single tertiary teaching hospital, there is no guarantee that similar results would be obtained in other locations, clinical settings, or specific patient groups. Although the study has done retrospective analysis, we are confident that the system can be implemented as a real-time early warning system. Finally, the proposed model should not be regarded as the sole solution for detecting patient progression over time. Rather, it should be used as part of the daily routine clinical practice.

V. Conclusion

Given longitudinal data, numerous machine learning approaches can be applied to extract different semantic and latent knowledge, so that informed decisions can be made. The advantages of machine learning approaches in healthcare are multifold. It can be used to make timely, lifesaving, and effective decisions. On the other hand, it can save unnecessary wastage of resources both for the patient, healthcare industry, and medical practitioners.

Early studies such as [3 - 5] use the sum of assigned points of generic vital signs from a subjectively fixed normal range to compute early warning scores to identify patients that are deteriorating but also can have their outcome changed by timely intervention. These types of systems are solely based on physiological parameters of the patient's vital signs such as heart rate, breathing rate, and systolic blood pressure. In this study, we showed that PCA can be used as part of a tool for early warning of medical outcome changes and epiphenomenon in ICU settings. It can be used as a novel non-disease specific tool that can observe patient daily clinical changes and provide non-disease specific analysis of patient progress over time for proactive actions. Given the facts, the user can be able to decide what treatment or therapy to prescribe or which diagnosis to perform further. However, additional investigation is required with other machine learning models to provide a full-fledged and more robust support for the medical practitioners.

VI. Author Contribution Statement

The authors confirm contribution to the paper as follows: M. Abebe YIMER and S. SEVINÇ conceived and presented the idea. M. Abebe YIMER took the responsibility of data collection and preparation. M. Abebe YIMER carried out the model implementation with the help of O. YILDIRIM and E. YAŞAR. Analysis and interpretation of results are carried out by M. Abebe YIMER, S. SEVINÇ, and A. Rıza ŞIŞMAN. The draft manuscript is prepared by M. Abebe YIMER. Also, A. Rıza ŞIŞMAN contributed to the interpretation of the results from a medical perspective and S. SEVINÇ, & Ö. AKTAŞ helped supervise the project. A final and critical review of the article before submission not only for spelling and grammar but also for its intellectual content is conducted by M. Abebe YIMER, S. SEVINÇ, A. Rıza ŞIŞMAN, and Ö. AKTAŞ. Finally, all authors reviewed the results, provided critical feedback, and approved the final version of the manuscript.

VII. Ethics Approval

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

VIII. Conflict of Interest

The authors declare no conflicts of interest.

References

[1] W. Widanagamaachchi, Y. Livnat, P.-T. Bremer, S. Duvall and V. Pascucci, "Interactive Visualization and Exploration of Patient Progression in a Hospital Setting," AMIA Annual Symposium Proceeding, vol. v2017, pp. 1773-1782, 2017.

[2] C.-F. Wang, J. Li, K.-L. Ma, C.-W. Huang and Y.-C. Li, "A Visual Analysis Approach to Cohort Study of Electronic Patient Records," in IEEE International Conference on Bioinformatics and Biomedicine, Belfast, UK, 2014.

[3] C. Mathukia, W. Fan, K. Vadyak, C. Biege and M. Krishnamurthy, "Modified Early Warning System improves patient safety and clinical outcomes in an academic community hospital," Journal of Community Hospital Internal Medicine Perspectives, vol. 5, no. 2, 2015.

[4] G. B. Smith, D. R. Prytherch, P. Meredith, P. E. Schmidt and P. I. Featherstoned, "The ability of the National Early Warning Score (NEWS) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission and death," Elsevier Resuscitation, vol. 84, no. 4, pp. 465-470, 2013.

[5] D. R. Prythercha, G. B. Smith, P. E. Schmidt and P. I. Featherstone, "ViEWS—Towards a national early warning score for detecting adult inpatient deterioration," Elsevier Resuscitation, vol. 81, no. 8, p. 932–937, 2010.

[6] C. Ye, O. Wang, M. Liu, L. Zheng, M. Xia, S. Hao, B. Jin, H. Jin, C. Zhu, C. J. Huang, P. Gao, G. Ellrodt, D. Brennan, F. Stearns, K. G. Sylvester, E. Widen, D. B. McElhinney and X. Ling, "A Real-Time Early Warning System for Monitoring Inpatient Mortality Risk: Prospective Study Using Electronic Medical Record Data," Journal of Medical Internet Research, vol. 21, no. 7, 2019.

[7] X. Cai, O. Perez-Concha, E. Coiera, F. Martin-Sanchez, R. Day, D. Roffe and B. Gallego, "Real-time prediction of mortality, readmission, and length of stay using electronic health record data," Journal of Amarican Medical Information Association, vol. 23, no. 3, pp. 553-561, 2016.

[8] F. Babić, M. Vadovský, M. Muchová, J. Paralić and L. Majnarić, "Simple Understandable Analysis of Medical Data to Support the Diagnostic Process," in IEEE 15th International Symposium on Applied Machine Intelligence and Informatics, Slovakia, 2017.

[9] H. Mohamadlou, A. Lynn-Palevsky, C. Barton, U. Chettipally, L. Shieh, J. Calvert, N. R. Saber and R. Das, "Prediction of Acute Kidney Injury With a Machine Learning Algorithm Using Electronic Health Record Data," Canadian Journal of Kidney Health, vol. 5, pp. 1-9, 2018.

10] I. Babaoğlu, O. Fındık and M. Bayrak, "Effects of principal component analysis on assessment of coronary artery diseases using support vector machine," Elsevier Expert Systems with Applications, vol. 37, pp. 2182-2185, 2010.

[11] S. Kara and F. Dirgenali, "A system to diagnose atherosclerosis via wavelet transforms, principal component analysis and artificial neural networks," Elsevier Expert Systems with Applications, vol. 32, pp. 632-640, 2007.

[12] A. E. Johnson, J. T. Pollard, L. Shen, H. L.-W. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. A. Celi and R. G. Mark, "MIMIC-III, a freely accessible critical care database," Scientific Data, 2016.

[13] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. Ivanov, R. Mark, J. Mietus, G. Moody, C. Peng and H. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals," Circulation 101(23):e215-e220, 2000.

[14] J.-L. Vincent, N. D. Nielsen, N. I. Shapiro, M. E. Gerbasi, A. Grossman, R. Doroff, F. Zeng, P. J. Young and J. A. Russell, "Mean arterial pressure and mortality in patients with distributive shock: a retrospective analysis of the MIMIC-III database," Annals of Intensive Care, vol. 8, no. 107, 2018.

[15] I. H. Witten and E. Frank, Data Mining: Practical Machine Learning Tools and Techniques, 2005: Elsevier Science.

[16] P. D. Tyler, H. Du, M. Feng, R. Bai, Z. Xu, G. L. Horowitz, D. J. Stone and L. A. Celi, "Assessment of Intensive Care Unit Laboratory Values That Differ From Reference Ranges and Association With Patient Mortality and Length of Stay," JAMA Network Open, vol. 1, no. 7, 2018.

[17] L. Lebart, "Which Bootstrap for Principal Axes Methods?," in Selected Contributions in Data Analysis and Classification. Studies in Classification, Data Analysis, and Knowledge Organization, Springer, Berlin, Heidelberg, 2007, pp. 581-588.

[18] H. Ogasawara, "Asymptotic biases of the unrotated/rotated solutions in principal component analysis," British Journal of Mathematical and Statistical Psychology, vol. 57, p. 353–376, 2004.

[19] M. Linting, B. van Os and J. J. Meulman, "Statistical Significance of the Contribution of Variables to the PCA Solution: An Alternative Permutation Strategy," Psychometrika, vol. 76, no. 3, pp. 440-460, 2011.

BIOGRAPHIES

Mohammed Abebe YIMER, Born on June 1, 1985 at a city called Ukie estate farm, East Wollega, Oromia, Ethiopia. He completed his elementary school at Ukie elementary school, high school at Nekemte Derge Senior Secondary school and his preparatory education at Nekemte Comprehensive high school. Graduated with B.Sc. Degree in Computer Science and Information Technology from Arba Minch University and with MSc Degree in Computer Science from Addis Ababa University. He has been employed and working as a lectures at Arba Minch University, Ethiopia since 2007. Currently he is pursuing his PhD in Computer Engineering at Dokuz Eylül University, Izmir-Turkey. His research interest include natural language processing, artificial intelligence, machine learning, deep learning, and data analytics.

Özlem Aktaş received the B.S. degree in 2003, M.Sc. degree in 2005 and Ph.D degree in 2010 from 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 she has been working as Assistant Professor since 2010 in 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 PhD degree from University of Arizona, USA between 1986 -1988. He has been serving in multiple universities as an academician both in Turkey and Australia on different levels and capacity. He has been working in 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 is automation of medical laboratories, hospital laboratory information systems, robotics, and the application of artificial intelligence in medical laboratories.

Oktay 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

Oktay YILDIRIM, completed his high school study at Adnan Menderes Anadolu High School. Graduated with B.Sc. Degree in Computer Engineering from Dokuz Eylul University and with MSc Degree in Computer Science from Ege University. He is a Computer Engineer with vast experience. He has developed and contributed many projects as developer, team leader and project manager. In addition, He is also the Co-founder and Project Technical Manager at Labenko Bilisim Inc., Izmir-Turkey. Currently he is pursuing his PhD in Computer Engineering at Dokuz Eylül University.

Eminullah YAŞAR, completed his high school at Anadolu High School in 2008. Graduated both with B.Sc. Degree and with MSc Degree in Computer Engineering from Dokuz Eylul University. He is a Computer Engineer with vast entrepreneur experience. Skills and interests include artificial intelligence, analytical thinking and creative problem solving. He has developed and contributed many projects as developer, team leader and project manager. In addition, He is also the Co-founder, Artificial Intelligence-Robotic System Developer and Robotic Team Leader at Labenko Bilisim Inc., Izmir-Turkey. Currently pursuing his PhD study since 2020 at Izmir Institute of Technology with the motto that "Brains that don't lose their eagerness to learn can achieve great things".