Healthcare Transformation with Informatics and Artificial Intelligence J. Mantas et al. (Eds.) © 2023 The authors and IOS Press. This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0). doi:10.3233/SHTI230493

A Method to Predict Comorbid Conditions Using Risk Factor Profile of Multiple Fluid Biomarkers

PRIYANKA JADHAV^{a,1}, VINOTHINI SELVARAJU^a and RAMAKRISHNAN SWAMINATHAN^a

^aBiomedical Engineering, Department of Applied Mechanics, Indian Institute of Technology Madras, Chennai, India

Abstract. The co-occurrence of diabetes, hypertension, and cardiovascular diseases together can make clinical management and treatment more complex. Early detection of comorbid conditions can help in creating personalized treatment plans. Multiple fluid biomarkers can be used to enhance the diagnostic accuracy of identifying comorbidity. This study aims to distinguish non-comorbid and comorbid conditions using the risk factor profile of multiple fluid biomarkers, such as creatine phosphokinase, platelet count, serum creatinine, and ejection fraction. Area feature is computed by utilizing risk factor profile of the biomarkers, and a random forest classifier is used to distinguish the two conditions. The results indicate that the area of the radar plot is more significant for differentiating comorbid from non-comorbid conditions. RF classifier achieves the highest accuracy of 59.91% to differentiate the two conditions. Thus, multiple fluid biomarkers could be used to accurately detect the comorbid condition and improve the treatment plan individually.

Keywords. Fluid biomarkers, comorbidity, risk factor profile, machine learning

1. Introduction

Noncommunicable diseases such as diabetes and hypertension pose serious health issues globally [1]. Approximately 50-80% of individuals with type 2 diabetes also suffer from hypertension at the same time. The presence of high blood pressure (BP) in individuals with diabetes is associated to a 57% higher likelihood of experiencing cardiovascular diseases (CVD) [2]. The presence of more than one condition simultaneously can lead to further complications and impact the treatment of the diseases [3].

Biomarkers present in different bodily fluids such as blood, saliva, urine and tears can act as diagnostic indicators for evaluating systemic disorders and associated comorbidities [4]. The use of fluid biomarkers helps in the detection and monitoring of an individual's comorbidities, which can lead to the development of personalized medicine [3]. In clinical settings, the examination of fluid biomarkers is considered a standard measure for diagnosing conditions like diabetes and hypertension [5].

Machine learning (ML) techniques can be utilized to create prediction models to identify comorbidity that incorporate fluid biomarkers. Early identification of

¹ Corresponding Author: Priyanka Jadhav, Biomedical Engineering, Department of Applied Mechanics, Indian Institute of Technology Madras, Chennai, India; E-mail: am22s052@smail.iitm.ac.in.

comorbidities can help in preventive therapy and alleviate the burden on the healthcare system.

This study aims to identify multiple fluid biomarkers that can differentiate between comorbid conditions. Risk factor profiles of the obtained fluid biomarkers are generated and analyzed. The extracted features are applied to ML algorithms to discriminate comorbid and non-comorbid conditions, and the performance is evaluated.

2. Methodology

The proposed methodology of this study is illustrated in Figure 1, which involves obtaining fluid biomarker features for comorbid and non-comorbid conditions from an open access dataset. Risk factor profiles are generated using radar plots for both conditions. The area of the radar plot is obtained and then utilized as input for the classifier to distinguish between comorbid and non-comorbid conditions.



Figure 1. Proposed methodology

2.1. Dataset

This study employs a dataset for predicting heart failure that was collected between April and December 2015 from the Faisalabad Institute of Cardiology and the Allied Hospital located in Faisalabad, Punjab-Pakistan. [6]. The dataset consists of 299 participants, comprising of 105 females and 194 males aged between 40 and 95 years. All patients in the dataset have previously been diagnosed with heart failure and have systolic dysfunction in the left ventricle. There are 13 features, namely serum sodium (SS), creatine phosphokinase (CPK), ejection fraction (EF), platelet count (PL), serum creatinine (SC), age, high BP, diabetes, anemia, sex, smoking, death event, and time included in the dataset.

The dataset is categorized into two groups: non-comorbid (participants without hypertension and diabetes) and comorbid (participants with both hypertension and diabetes). Moreover, this study primarily focuses on fluid biomarkers present in the dataset, such as CPK, EF, PL, SC, and SS for further investigation. The radar plots are used to compare the risk factor profile of the non-comorbid and comorbid groups based on different combinations of fluid biomarkers. Groups 1 to 5 correspond to combinations of fluid biomarkers created by excluding SS, EF, PL, SC, and CPK respectively. The area of the radar plot for each group is computed which is given as the input to the classifier.

The number of samples in the comorbid and non-comorbid conditions are not balanced. It is likely to detect majority (non-comorbid) class. Hence, Data balancing technique namely, synthetic minority oversampling technique (SMOTE) is utilized to balance the samples [7]. This study employs 10-fold cross validation technique.

A random forest (RF) classifier is a supervised ML algorithm that divides the training dataset into multiple subsets and builds a decision tree for each subset. To classify a new data point, the RF algorithm uses all the decision trees to make predictions, and then selects the classification with the majority votes from the individual decision trees [8]. In this study, the number of trees is set to 50. The performance of the classifier, such as accuracy, sensitivity, and specificity, is assessed to evaluate the performance of the classifier [8].

3. Results and Discussion

The representative radar plot for group 1 under comorbid and non-comorbid conditions is shown in Figure 2(a). The radar plot displays the value of the features on each axis and the overall risk level based on the area of the plot. It is observed that comorbidity has a higher platelet count than the non-comorbid condition. An elevated platelet count over an extended duration is regarded as a potential risk factor for the development of hypertension and cardiac abnormalities [9]. This might affect the area of the risk factor profile in both conditions. On the other hand, fluid biomarkers such as SC, CPK, and EF, are observed to exhibit no significant difference in distinguishing between comorbid and non-comorbid conditions. Therefore, utilizing a ML approach could aid in distinguishing non-comorbid and comorbid conditions. Figure 2(b) depicts the performances of the RF classifier for group 1.



Figure 2. (a) Radar plot for representative subjects belonging to non-comorbid and comorbid population for the combination of features corresponding to group 1, (b) Performance measures for group 1

Table 1 shows the classifier performance for group 1, 2, 3, 4, and 5. RF classifier yields an accuracy of $59.91\pm0.12\%$, $49.66\pm0.06\%$, $47.66\pm0.11\%$, $56.75\pm0.13\%$, and $56.08\pm0.10\%$ for group 1, 2, 3, 4, and 5, respectively. It shows that group 1 achieves the highest accuracy. It is evident that group 3 has the lowest accuracy due to the exclusion of platelet count. This clearly indicates that platelet count can be used as a biomarker for differentiating these two conditions. Sensitivity and specificity of 40.50% and 67.80% are observed to be higher for group 1. It is evident from these results that group 1 is optimal for differentiating comorbid and non-comorbid conditions. However, the limited

dataset cannot be used effectively to develop ML model and, in the future, more samples will be collected in real-time to develop better models.

Performance	Combination of fluid biomarkers				
metrics (%)	Group 1	Group 2	Group 3	Group 4	Group 5
Accuracy	59.91±0.12	49.66±0.06	47.66±0.11	56.75±0.13	56.08±0.10
Sensitivity	40.50±0.24	32.00±0.19	30.00±0.19	30.00±0.26	34.50±0.15
Specificity	67.80±0.13	56.13±0.12	54.54±0.12	66.96±0.18	64.39±0.11

Table 1. Comparison of classification performances for group 1,2,3,4, and 5

4. Conclusion

In this study, an attempt has been made to differentiate comorbid from non-comorbid conditions using the risk factor profile of the multiple fluid biomarkers. For this purpose, fluid biomarkers namely, SS, EF, PL, SC, and CPK are considered. RF classifier is utilized to categorize comorbid and non-comorbid conditions. The results indicate that the area of the radar plot derived from the selected fluid markers is able to distinguish the comorbid condition. Platelet count is found to be high in comorbid subjects and shows a significant difference. RF classifier shows the highest accuracy with the combination of PL, CPK, EF, and SC. Analyzing risk factor profiles of multiple fluid biomarkers can provide a more comprehensive and accurate assessment of an individual's health status to predict or manage comorbid conditions. However, further investigations on various biomarkers might be required due to co-existing diseases and will be explored in the future.

References

- Li Y, Sun X, Wei J, Wu J, Wang Y. Time trends in comorbidity and management of hypertension and self-reported diabetes: a 15-year nationwide longitudinal study in China. American Journal of Hypertension. 2021 Aug;34(8):810-20. doi: 10.1093/ajh/hpab024.
- [2] Przezak A, Bielka W, Pawlik A. Hypertension and Type 2 Diabetes—The Novel Treatment Possibilities. International Journal of Molecular Sciences. 2022 Jun 10;23(12):6500. doi: 10.3390/ijms23126500.
- [3] Petrie JR, Guzik TJ, Touyz RM. Diabetes, hypertension, and cardiovascular disease: clinical insights and vascular mechanisms. Canadian Journal of Cardiology. 2018 May 1;34(5):575-84. doi: 10.1016/j.cjca.2017.12.005.
- [4] Benincasa G, Mansueto G, Napoli C. Fluid-based assays and precision medicine of cardiovascular diseases: the 'hope'for Pandora's box?. Journal of Clinical Pathology. 2019 Dec 1;72(12):785-99. doi: 10.1136/jclinpath-2019-206178.
- [5] Thomas V, Mithrason AT. A review on biomarkers of hypertension. International journal of clinical biochemistry and research. 2022;9:186-190. doi: 10.18231/j.ijcbr.2022.037.
- [6] Chicco D, Jurman G. Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. BMC medical informatics and decision making. 2020 Dec;20(1):1-6., doi: 10.1186/s12911-020-1023-5.
- Blagus R, Lusa L. SMOTE for high-dimensional class-imbalanced data. BMC bioinformatics. 2013 Dec;14:1-6. doi: 10.1186/1471-2105-14-106.
- [8] Duda RO, Hart PE, Stork DG, Pattern Classification, John Wiley & Sons, New York, 2012.
- [9] Chiu PC, Chattopadhyay A, Wu MC, Hsiao TH, Lin CH, Lu TP. Elucidation of a causal relationship between platelet count and hypertension: A bi-directional Mendelian randomization study. Frontiers in Cardiovascular Medicine. 2021 Nov 26;8:743075. doi: 10.3389/fcvm.2021.743075.