An Improved Big Data Analysis of Diabetic Condition Based on Hemoglobin Protein

Mohammed Tanzim Shoaib¹, Manisha Joshi²

¹M.Tech. Student, Department of Medical Electronics, BMS College of Engineering, Bangalore, India
²Asst. Professor, Department of Medical Electronics, BMS College of Engineering, Bangalore, India

Abstract: Machine learning has undergone significant development over the past decade and is being used successfully in many intelligent applications covering a wide array of data related problems. One of the most intriguing questions is whether machine learning can be successfully applied to the field of medical diagnostics. Moreover, there is a question as to what kind of data are needed. Several examples of successful applications of machine learning methods in specialized medical fields exist. Recently, a model capable of classifying skin cancers based on images of the skin was presented that achieves a level of competence comparable to that of a dermatologist. There are however, no successful applications of machine learning that tackle broader and more complex fields in medical diagnosis, such as HbA1c level.

Keywords: Big data, Hemoglobin Protein.

I. INTRODUCTION

It is increasingly recognized that the management of hyperglycaemia in the hospitalized patient has a significant bearing on outcome, in terms of both morbidity and mortality. This recognition has led to the development of formalized protocols in the intensive care unit (ICU) setting with rigorous glucose targets in many institutions. However, the same cannot be said for most non-ICU inpatient admissions. Rather, anecdotal evidence suggests that inpatient management is arbitrary and often leads to either no treatment at all or wide fluctuations in glucose when traditional management strategies are employed. Although data are few, recent controlled trials have demonstrated that protocol driven inpatient strategies can be both effective and safe. As such, implementation of protocols in the hospital setting is now recommended. However, there are few national assessments of diabetes care in the hospitalized patient which could serve as a baseline for change. The present analysis of a large clinical database was undertaken to examine historical patterns of diabetes care in patients with diabetes admitted to a US hospital and to inform future directions which might lead to improvements in patient safety. In particular, we examined the use of HbA1c as a marker of attention to diabetes care in a large number of individuals identified as having a diagnosis of diabetes mellitus.

II. AIM AND SCOPE

A. Methodology

This study used the health Facts database (Cerner Corporation, Kansas City, MO), a national data warehouse that collects comprehensive clinical records across hospitals throughout the United States.
mortality, and hospital characteristics. All data were identified in compliance with the Health Insurance Portability and Accountability Act of 1996 before being provided to the investigators. Continuity of patient encounters within the same health system (HER system) is preserved.

B. Algorithm process

- Data acquisition
- Data filtering
- Data pre-processing
- Data modelling
- Evaluation

III. RESULTS AND DISCUSSION

In this study, we showed that a machine learning approach, using a random forest algorithm trained on large amounts of multianalyte sets of HbA1c level laboratory blood test results, is able to interpret the results and predict diseases with an accuracy on par with experienced diabetic specialists, while outperforming internal medicine specialists by a margin of more than two.

A. Random forest

- HbA1C levels consist of none, norm, >7, >8
- Random forests: None

Fig. 1. Algorithmic process

Fig. 4. Random forests: None

Fig. 5. Normal

Fig. 6. >7

Fig. 7. >8
B. Logistic regression

<table>
<thead>
<tr>
<th></th>
<th>Predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt;8</td>
</tr>
<tr>
<td>Actual</td>
<td></td>
</tr>
<tr>
<td>&lt;8</td>
<td>84%</td>
</tr>
<tr>
<td>&gt;7</td>
<td>7%</td>
</tr>
<tr>
<td>&gt;8</td>
<td>4%</td>
</tr>
<tr>
<td>Normal</td>
<td>5%</td>
</tr>
</tbody>
</table>

100% 100% 100% 100%

Fig. 8. Confusion matrix

Fig. 9. Random forests: None

Fig. 10. Normal

Fig. 11. >7

Fig. 12. >8

Fig. 12. Predictive analysis of A1C vs. HbA1C levels

IV. CONCLUSION

Machine learning models can recognize Hb1AC levels laboratory patterns that are beyond current medical knowledge, resulting in higher diagnostic accuracy compared to traditional quantitative interpretations based on reference ranges. These changes can be large, and physicians can observe them by checking for A1C level parameter values outside of normal ranges. Predictive models show great promise in medical laboratory diagnoses and could not only be of considerable value to both physicians and patients but also have widespread beneficial impacts on healthcare costs.

This study evaluated HbA1c by the of column chromatography with exchange resins in which patients with hemoglobin heterozygotes variants did not present a difference significant difference in relation to the control group.

REFERENCES


