Majlesi Journal of Electrical Engineering
Vol. 18, No. 1, March 2024

Support Vector and Linear Regression Machine Learning Model on Amperometric Signals to Predict Glucose Concentration and Hematocrit Volume

Kirti Sharma[1], Pawan K. Tiwari[2], Sanjay Kumar Sinha[3]
Department of Physics, Birla Institute of Technology, Mesra, Ranchi 835215, India.
Email: kirti.sharma151700@gmail.com
Email: pawan.ktiwari@gmail.com (Corresponding author)
Email: sksinha@bitmesra.ac.in

ABSTRACT:
Data represents a compendium of information that perpetually expands with each passing moment, contributed by individuals worldwide. Within the domain of medical science, this reservoir of data accumulates at an almost exponential rate, doubling in volume annually. The emergence of advanced machine learning tools and techniques, subsequent to a substantial evolution in data mining strategies, has bestowed the capacity to glean insights and discern concealed patterns from vast datasets, thus enabling extensive analytical pursuits. This study delves into the application of machine learning algorithms to enhance societal well-being by harnessing the transformative potential of machine learning advancements in the domain of blood glucose concentration estimation through regression analysis. The culmination of this investigation involves establishing a correlation between glucose concentration and hematocrit volume. The dataset employed for this research is sourced from clinically validated electrochemical glucose sensors (commonly referred to as glucose strips). It encompasses diverse levels of both glucose concentration and hematocrit volume, the latter being furnished by an undisclosed source to ensure copyright compliance. This dataset comprises four distinct variables, and the aim of this research involves training the dataset using regression techniques to predict two of these variables. Our results indicate that when utilizing linear regression, the R2 score for GC is approximately 0.916, whereas for HV, it reaches around 0.537. In contrast, employing the support vector regressor yielded R2 scores of about 0.961 for GC and 0.506 for HV.

KEYWORDS: Estimation, Correlation, Analysis, Regression, Healthcare, Enlightenment, Machine learning, Quantum leap, Data Mining, Insights.

1. INTRODUCTION

Machine learning is an advanced technology that is expanding its applications exponentially throughout the world by using futuristic technology to uplift socio-technicality [1]–[3]. The application of machine learning is expanding its horizon in the healthcare sector to analyze complex data sources and find unique patterns to interconnect each data point of big datasets to predict the onset of disease at an early stage [4]. After the incorporation of the ML system, optimal and accurate diagnosis is possible to find out the appropriate treatment of such kinds of ailments such as cancer, diabetes, Alzheimer's, etc. [1]. Great anticipation of machine learning mechanisms to diagnose medical issues and eject enriched intricacies from a big dataset [5]. Enormous informatic embedded data is being generated every day in the medical field, every patient has their own characteristics and valuable medical report [4]. Nowadays every person suffers from different

Paper type: Research paper
https://doi.org/10.30486/mjee.2023.2004331.1339
Received: 16 December 2023; revised: 11 January 2024; accepted: 5 February 2024; published: 1 March 2024
How to cite this paper: K. Sharma, P. K. Tiwari and S. K. Sinha, **"Support Vector and Linear Regression Machine Learning Model on Amperometric Signals to Predict Glucose Concentration and Hematocrit Volume"**, *Majlesi Journal of Electrical Engineering*, Vol. 18, No. 1, pp. 105-115, 2024.

types of health issues and that is accumulated as a medical report for the respective person. Therefore, the ML system emerged to introduce program software that is trained with all the informatic embedded data and learns from big data and intelligently predicts new insights from it.

Machine learning tools and techniques have made significant contributions to the healthcare field. These technologies are being used to improve patient care, diagnostics, treatment plans, drug discovery, and more. Here are some ways in which machine learning is being employed in healthcare such as Disease Prediction and Diagnosis using machine learning algorithms that can analyze large sets of medical data, such as patient records, lab results, and imaging data, to predict diseases and assist in accurate diagnoses. For example, models have been developed to predict conditions like heart disease, diabetes, and certain types of cancer. Medical Imaging [6] using machine learning which has revolutionized medical imaging interpretation. Algorithms can detect abnormalities in X-rays, MRIs, CT scans, and other imaging modalities [7], aiding radiologists in identifying potential issues early and accurately. Drug discovery by machine learning models that can predict how different compounds will interact with biological systems, accelerating drug discovery processes. They can help identify potential drug candidates and optimize molecular structures. Personalized Treatment Plans by analyzing patient data and medical history, machine learning can assist in creating personalized treatment plans. This includes selecting appropriate medications and dosages based on individual patient characteristics. Machine learning is used to analyze genomic data to identify genetic markers related to diseases. This aids in understanding disease risk, prognosis, and potential treatment approaches [8]. Machine learning helps in managing and extracting valuable insights from electronic health records, enhancing patient care coordination and decision-making [9]. Machine learning can help identify suitable candidates for clinical trials based on their medical history, improving trial success rates [10].

It is important to note that while machine learning has immense potential in healthcare, its implementation requires careful consideration of ethical, privacy, and regulatory concerns. Proper data handling, model interpretability, and patient consent are critical aspects of applying machine learning in the healthcare sector.

Besides the healthcare field, machine learning tools and techniques are employed in areas such as business development, market research, academic and clinical research, etc. [11]. A comprehensive survey on one of the areas is explained as follows:

1.1. Academic and Clinical Research

With the tremendous transformation in the domain of academic and clinical research due to machine learning, big data is increasing day by day, and the data embedded with large enlightenment is broadly used to improve research outcomes and build a new strategy. Clinical investigation is a wide range of observational and experimental research that is used to accomplish a new insight to analyze the traditional trial. There is a considerable difference between academic and clinical research. The academic research, which is thoroughly based on theoretical investigation and the ML contribution is to significantly identify patterns using efficient machine learning algorithms from a repository, improving the precision of the classification models. Summarizing, insights from a massive dataset, improving experimental mechanisms, and implementing algorithms are the actions for the development of academic and clinical research. The delineation of clinical research in the context of experimental research, the importance and continued evolution of investigation that is being processed in heuristically solving industrial-based projects problems [12].

2. RESEARCH OBJECTIVE

Our research is centered around the core objective of comprehensively grasping the connection between blood glucose concentration and hematocrit volume, all the while conducting a thorough comparative analysis of the outcomes derived from two distinct algorithms. Within the healthcare domain, our focus lies in unraveling potential insights into the intricate interplay and mutual influence of these two pivotal parameters. We aim to establish a robust and nuanced relationship between blood glucose concentration and hematocrit volume, delving into the complex dynamics governing their interaction within the human body. To achieve this, our approach involves the utilization of two different algorithms for the estimation of blood glucose concentration and hematocrit volume. By meticulously comparing the outcomes generated by these algorithms, we endeavor to discern any disparities, advantages, or variations in their predictive accuracy and efficiency. Our research endeavors to make a substantial contribution to the broader healthcare field by shedding light on the implications stemming from the correlations identified between blood glucose and hematocrit. The potential ramifications of these findings include advancements in diagnostic tools, treatment methodologies, and patient care practices. Through meticulous data collection, rigorous algorithm implementation, and an exhaustive comparative assessment, our aim is to elevate our understanding of the intricate relationship between blood glucose concentration and hematocrit volume.

3. LITERATURE REVIEW

Numerous research studies have been conducted in the realm of diabetes mellitus prediction utilizing machine learning techniques. These investigations have explored a diverse array of methods and datasets to enhance the accuracy of detection. Various innovative approaches have been employed, each offering unique insights into this critical area of healthcare. The following literature review provides an overview of some of these approaches:

3.1. Continuous Glucose Monitoring Systems

A substantial body of research has centered around the utilization of continuous glucose monitoring systems for diabetes mellitus prediction. These systems leverage real-time glucose level data to facilitate early detection and personalized management of the disease. Machine learning algorithms have been harnessed to analyze the intricate patterns and fluctuations in glucose levels, aiding in accurate prediction [13].

3.2. Plethysmography Signals

Another avenue explored is the use of plethysmography signals for diabetes prediction. By analyzing variations in blood volume within tissues, these signals offer valuable insights into vascular function and blood flow dynamics. Machine learning algorithms have been applied to decipher complex patterns within plethysmography data, contributing to the development of predictive models for diabetes [13], [14].

3.3. Calorimetry Methods

Calorimetry techniques, which involve measuring heat release or absorption during metabolic processes, have also been integrated with machine learning for diabetes detection. These methods provide an indirect indicator of metabolic activity and energy expenditure. By harnessing machine learning algorithms, researchers have endeavored to derive correlations between calorimetry data and diabetes risk, enhancing prediction accuracy [15].

3.4. Colorimetry Techniques

Colorimetry, a method that quantifies color changes in response to chemical reactions, has been harnessed for diabetes prediction. The interaction between blood components and reagents can yield insights into glucose concentrations. Machine learning algorithms applied to colorimetric data have contributed to the development of non-invasive and cost-effective prediction models [15].

3.5. Demographic and Anthropometric Measurements

Beyond physiological signals, demographic and anthropometric measurements have also been integrated into diabetes prediction models. Factors such as age, gender, body mass index (BMI), and waist circumference have been shown to have significant correlations with diabetes risk. Machine learning algorithms have been leveraged to create comprehensive predictive frameworks that incorporate these variables [16].

Overall, the literature demonstrates a wide range of innovative approaches for diabetes mellitus prediction using machine learning algorithms. These methodologies, including continuous glucose monitoring systems, plethysmography signals, calorimetry methods, colorimetry techniques, and demographic and anthropometric measurements, collectively contribute to advancing our understanding of diabetes risk factors and enhancing the accuracy of predictive models. The synthesis of these diverse approaches holds promise for improving early detection and intervention strategies in the management of diabetes mellitus.

The methodologies mentioned have several limitations that need to be considered. Continuous glucose monitoring systems, while providing real-time glucose level data, can be affected by sensor inaccuracies, signal lag, and skin irritation, potentially compromising their reliability. Plethysmography signals, commonly used in cardiovascular assessment, may suffer from motion artifacts and provide limited information on specific physiological parameters. Calorimetry methods, although valuable for assessing energy expenditure, can be influenced by variations in metabolic rate and may not capture all relevant factors affecting metabolism. Colorimetry techniques, while useful for various biochemical analyses, might lack the precision and sensitivity required for detecting subtle changes. Demographic and anthropometric measurements are prone to self-reporting errors and may not fully capture individual variations in body composition or other relevant factors. It is crucial to acknowledge and address these limitations when interpreting results obtained from these methodologies, ensuring accurate and meaningful insights into the physiological phenomena under investigation.

4. NOVELTY

The innovation of using amperometric signals for estimating blood glucose and hematocrit levels introduces a groundbreaking approach to the simultaneous measurement of two crucial health indicators. Our study acknowledges

the growing availability of data integrated into medical practices, specifically in the realm of diabetes mellitus. It recognizes the swift accumulation of data and emphasizes the necessity of harnessing this information to enhance the management and predictive understanding of diabetes-related elements. The primary focus of this investigation revolves around establishing a link between blood glucose levels and hematocrit volume – the proportion of red blood cells within the overall blood volume.

By delving into the interplay between these two factors, the research aims to contribute valuable insights into the realm of diabetes and its effective management. The study employs amperometric signals obtained via an enzyme-based electrochemical glucose sensor (commonly known as a glucose strip) through a redox reaction. The purpose of analyzing these glucose signals is to formulate a methodology for forecasting both blood glucose levels and hematocrit volume. This innovative approach adds to the existing body of knowledge by introducing a novel prediction technique rooted in amperometric signals.

However, the novelty of this study is embedded in various aspects. These include the exploration of the relationship between blood glucose concentration and hematocrit volume, the utilization of amperometric signals for predictive analysis, the integration of a Linear Regression and Support Vector Regressor, the implementation of evaluation metrics to gauge performance, and the aspiration to craft effective software for a biochemical glucose estimation analyzer. These elements collectively advance knowledge and hold promise for refining diabetes management, thereby mitigating the limitations inherent in current glucose measurement techniques.

5. MATERIAL AND METHODS

The dataset under consideration has been meticulously compiled through a comprehensive analysis of an electrochemical glucose sensor. This sensor operates based on the principle of amperometric transient current, which is generated as a result of the intricate interplay between micromolecules and macromolecules present within the bloodstream. At its core, the electrochemical glucose sensor functions as a sophisticated analytical tool aimed at quantifying the concentration of glucose within a biological sample, often blood. This process hinges on the phenomenon of amperometry, a technique employed in electroanalytical chemistry to measure the electric current generated during a redox reaction. In this particular application, the redox reaction is mediated by glucose molecules, which are subject to oxidation or reduction at the sensor's electrode surface[18]. The dataset's inception involves the careful selection of experimental parameters, such as the specific electrode material[19], the geometry of the sensor, and the electrochemical conditions. These factors collectively influence the sensor's performance characteristics, including sensitivity, selectivity, and response time. Subsequently, an array of blood samples, each with varying glucose concentrations, is subjected to the sensor's measurement process. The amperometric transient current, an integral aspect of this dataset, manifests as a time-dependent electric current response generated by the sensor [20]–[22]. It is a direct consequence of the interaction between the glucose molecules and the electrode surface. As glucose molecules come into contact with the sensor's electrode, they undergo a redox reaction. This interaction leads to the transient flow of electric charge, which is precisely quantified and recorded by the sensor.

The process of data collection is characterized by precision and systematic execution. Each blood sample is introduced to the sensor under controlled conditions, ensuring consistency and repeatability. The resulting amperometric transient current responses are captured over a defined time period, generating a wealth of temporal data points. These data points collectively form the foundation of the dataset, which encapsulates a diverse range of glucose concentrations and associated electric current profiles. Furthermore, the dataset's uniqueness arises from its ability to encapsulate not only the role of micromolecules like glucose in this electrochemical interaction but also the influence of macromolecules present within the blood matrix. These macromolecules, including proteins and other biomolecules, can impact the sensor's response due to their potential interference with the glucose-electrode interface. Thus, the dataset embodies a comprehensive representation of the complex interplay between various blood components and the sensor's analytical performance.

In conclusion, the dataset obtained through the analysis of the electrochemical glucose sensor, operating via the generation of amperometric transient currents, stands as a testament to meticulous experimentation and data collection[23], [24]. It embodies the intricate interactions between micromolecules such as glucose and the broader context of macromolecules within the blood. The dataset's integrity and value lie in its potential to drive advancements in glucose sensing technology, contribute to medical research, and enhance our understanding of electrochemical phenomena in biological systems.

5.1. Data Collection

Electrochemical Glucose Sensor (also known as glucose strip) is the glucose sensing phenomenon that senses glucose by the reaction of bioassays (enzyme, mediator, buffer salts, etc.) of glucose strips with the blood samples through a redox reaction [22], [24], [25]. The mechanism of the redox reaction is explained via three equations.

$Glucose + GOX_{(OX)} \rightarrow Gluconic\ acid + GOX_{(red)}$…..(1)
$GOX_{(red)} + 2M_{(OX)} \rightarrow GOX_{(OX)} + 2M_{(red)} + 2H^{+}$…....(2)
$2M_{(red)} \rightarrow 2M_{(OX)} + 2e^{-}$……..……..(3)

Where $GOX_{(OX)}$ and $GOX_{(red)}$ are oxidized and reduced glucose oxidase respectively, and $M_{(OX)}$ and $M_{(red)}$ are oxidized and reduced mediators respectively [22]. The above redox reaction shows the reaction between glucose strip bioassays and glucose in the blood sample. The redox reaction produces a transient current, and the transient current is considered a glucose signal [19], [26]. Glucose signals are produced by the variation of micromolecule (glucose concentration of the blood) and macromolecule (hematocrit volume of the blood) of the blood sample [22]. Redox reaction takes some time to complete the entire reaction because bioassays take time in dilution [21], [24].

5.2. ML Model Implementation

Machine Learning algorithms are applied to resolve the problem of the accuracy of a glucometer. We worked on the regression analysis to build the ML model for the prediction of blood glucose concentration and hematocrit volume [22], [27], [28]. There are four variables; peak current (IP), the time corresponding to peak current (TP), blood glucose concentration (GC), and hematocrit volume (HV). Two of them will be predicted and the rest two will be input.

5.2.1. Linear Regression

We employed the mechanism of linear regression in building a machine-learning model because of the continuous form of our dataset. The dataset is split into 70% & 30% proportions, the proportion of the training dataset is 70% and the remaining for the test dataset. On drawing and adjusting a linear straight line between GC and IP which corresponds x and y axes respectively. We obtained a coefficient of ~ 0.828 and an intercept of ~ 0.369 for glucose concentration which is shown in Fig. 1. Similarly, coefficient ~ -0.0122 and intercept ~ 0.369 for a straight line between HV and IP on the axis of x and y respectively, which is shown in Fig.2.

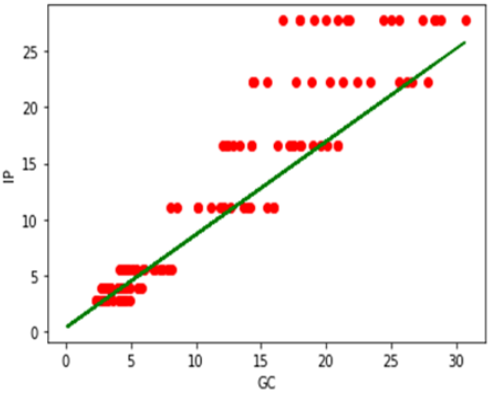

Fig. 1. Glucose Concentration.

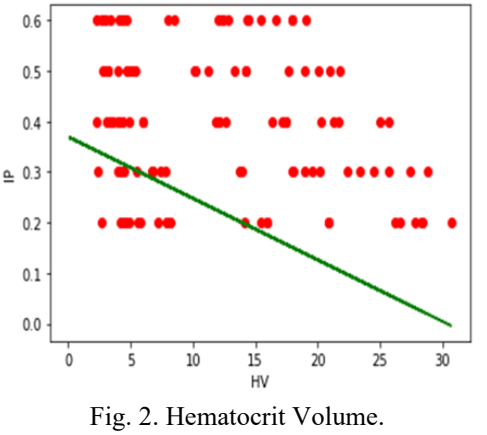

Fig. 2. Hematocrit Volume.

5.2.2. Support Vector Regressor

We observed the regression analysis of the support GC vector machine is suitable for better prediction than linear regression and applied kernel trick for the non-linear dataset. There are four types of kernel functions namely 'linear', 'sigmoid', 'RBF', and 'poly'. The polynomial function is fitted best for the non-linear dataset between GC and IP and kernel = 'RBF' (radial basis function) is fitted best for the HV and IP dataset.

6. RESULTS AND DISCUSSION

6.1. Linear Regression and its Outcome

In our modeling, 70% of the dataset is used for training the model with variables such as peak current (IP), peak time (TP), and glucose concentration (GC). After training the model, we tested the trained model on the remaining 30% of the dataset. We found an R^2 *score* between the test data of GC (y) and predicted values of GC (y_hat) from the sci-kit learn library.

A few code statements can be referred below:

≫from sklearn. metrics import r2_score
print ("R2-score: %.3f" % r2_score (y, y_hat))
R^2-score: 0.916

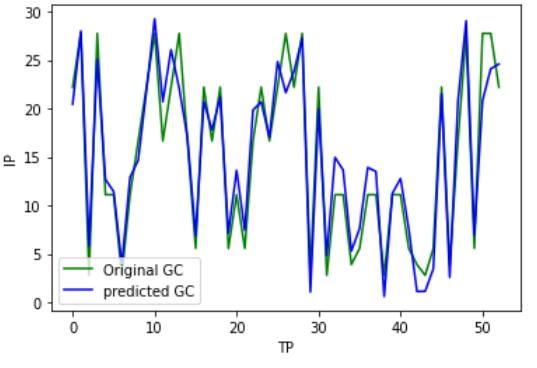

Fig. 3. Original v/s Predicted Glucose Concentration.

Table 1. Comparison of Original Glucose Concentration with Predicted Glucose Concentration on Transient Current and Current Corresponding Time.

Current, Time	Original Glucose Concentration	Predicted Glucose Concentration
[19.5182, 4.32]	[400]	[368.30057307]
[29.187, 4.01]	[500]	[504.41921745]
[23.5459, 5.24]	[500]	[452.35467053]
[11.004, 3.82]	[200]	[228.28444682]
[8.539, 4.37]	[200]	[205.86025839]
[3.527, 2.04]	[50]	[70.36674891]
[12.291, 3.24]	[200]	[232.35781899]
[11.64, 4.84]	[300]	[264.36177387]
[20.6, 4.67]	[400]	[393.55925649]

Similarly, in the context of hematocrit volume (HV) prediction, our study demonstrated the precision of the original values versus the predicted values. Despite the challenges posed by the intricate relationships within the dataset, our study lays the foundation for further research into refining predictive models for hematocrit volume. The nuanced insights gained from this analysis are valuable for medical professionals and researchers striving to better understand hematocrit volume dynamics and their implications for various health conditions. A few lines of code for finding the evaluation metric for hematocrit volume are referred below:

≫from sklearn. metrics import r2_score
print ("R2-score: %.3f" % r2_score (y_2, y_hat_2))
R2-score: 0.537
Where y_2 is the test data of HV and y_hat_2 is the predicted values of HV.

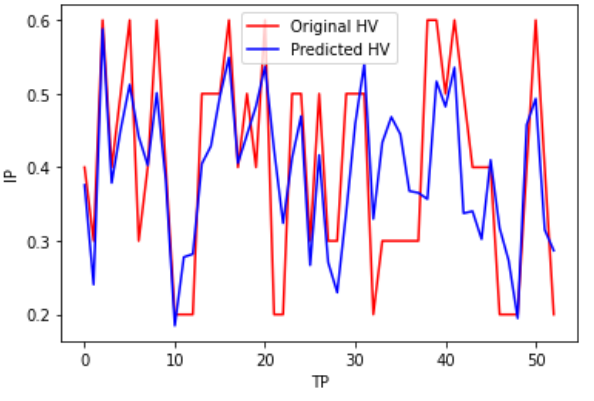

Fig. 4. Original v/s Predicted Hematocrit Volume.

Table 2. Comparison of Original Hematocrit Volume with Predicted Hematocrit Volume on Transient Current and Current Corresponding Time.

Current, Time	Original Hematocrit Volume	Predicted Hematocrit Volume
[19.5182, 4.32]	[40]	[37.58306666]
[29.187, 4.01]	[30]	[24.06292658]
[1.9101, 4.28]	[60]	[58.79058921]
[23.5459, 5.24]	[40]	[37.87630072]
[11.004, 3.82]	[50]	[45.12308021]
[8.539, 4.37]	[60]	[51.22986971]
[3.527, 2.04]	[30]	[44.17142161]
[12.291, 3.24]	[40]	[40.28077642]
[11.64, 4.84]	[60]	[50.1096327]

Our observation as displayed in Fig. 3 shows the difference between the actual values of glucose concentration and predicted values of glucose concentration at the rate of change of the current value. Fig. 4. shows the plotted value of original values and predicted values of hematocrit volume at the rate of change of the current value. And, Table 1. and

Table 2. show the comparison between original values and predicted values of glucose concentration and hematocrit volume, respectively.

6.2. Support Vector Regressor and its Outcome

Utilizing the sci-kit learn library, the dataset is split into trainsets and test sets in the same proportion as in the linear regression model. Firstly, the considered variables are IP, TP, and GC for training the model, and glucose concentration (GC) must be predicted. The dataset can be linear or non-linear, therefore we used the kernel trick for the non-linear dataset. The dataset is fitted best for the polynomial kernel function and the remarkable prediction is observed via the following code lines:

≫svr_poly. score (test_set_x, test_set_y)
0.96066
≫from sklearn. metrics import r2_score
print ("R2-score: %.3f" % r2_score (test_set_y, y_hat))
R^2-score: 0.961

Where test_set_y is the original value of GC and y_hat is the predicted value of glucose concentration (GC).

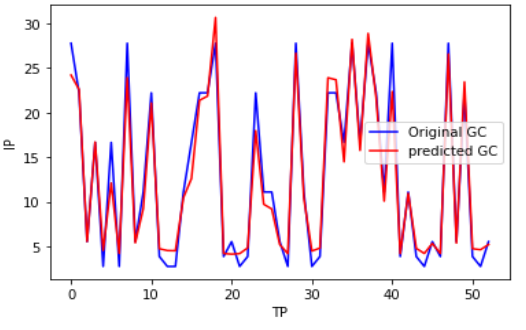

Fig. 5. Original v/s Predicted Glucose Concentration.

Similarly, this process was executed for the prediction of hematocrit volume (HV), but there is a bit slight difference between them in that the 'RBF' radial basis kernel function is used for HV prediction instead of the polynomial kernel function to fit the model for glucose concentration.

≫from sklearn. metrics import r2_score
print ("R2-score: %.3f" % r2_score (test_set_y_2, y_hat_2))
R2-score: 0.506

Where test_set_y_2 is the original value of HV and y_hat_2 is the predicted value of hematocrit volume (HV).

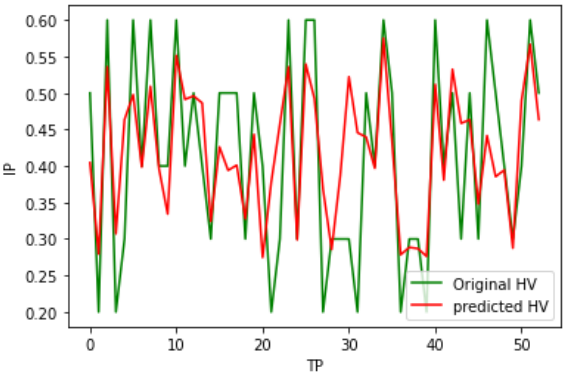

Fig. 6. Original v/s Predicted Hematocrit Volume

The difference between the actual values of glucose concentration and predicted values of glucose concentration at the rate of change of the current value is shown in Fig. 5, moreover, Fig. 6 displays the plotted value of original values and predicted values of hematocrit volume at the rate of change of the current value.

For glucose concentration prediction, using linear regression yielded an R2 score of 0.916, while utilizing a polynomial kernel function with a Support Vector Regressor (SVR) improved the prediction to an R2 score of 0.961. Similarly, for hematocrit volume prediction, we achieved an R2 score of 0.537 with linear regression and an R2 score of 0.506 with SVR using an RBF radial basis kernel function. These results demonstrate the effectiveness of the SVR approach, especially with the appropriate kernel function, in enhancing the predictive accuracy for both glucose concentration and hematocrit volume. The findings underscore the importance of choosing an appropriate modeling technique and kernel function for different types of datasets to achieve accurate predictions.

7. BIOLOGICAL ASPECTS

The outcomes of our modeling and analysis hold significant implications in the realm of biology and healthcare. The connection existing between blood glucose and hematocrit enables the concurrent assessment of glucose concentration and hematocrit volume. This capability proves advantageous in the identification and diagnosis of diabetes and polycythemia. By leveraging a comprehensive dataset comprising peak current (IP), peak time (TP), glucose concentration (GC), and hematocrit volume (HV), we effectively trained and tested predictive models to discern crucial biological insights. The precision of our predictions, as measured by R-squared scores, highlights the potential of these models in estimating key biological parameters. Notably, the utilization of a polynomial kernel function with a Support Vector Regressor (SVR) significantly enhanced the accuracy of predicting both glucose concentration (GC) and hematocrit volume (HV). These findings provide valuable tools for clinicians and researchers to gain deeper insights into the relationships between these parameters, thereby aiding in the diagnosis and management of various medical conditions. Moreover, the study underscores the pivotal role of selecting appropriate modeling techniques and kernel functions tailored to specific dataset characteristics, emphasizing the importance of precision in biological predictions for advancing medical understanding and patient care.

8. CONCLUSION

The evolution of machine learning tools and techniques within the realm of data mining has emerged as a potent mechanism for extracting valuable insights within the healthcare domain. This article has traversed the landscape of machine learning's evolving landscape, particularly in the context of prediction, setting a benchmark for the advancement of artificial intelligence. Our exploration has been driven by a profound interest in leveraging learning algorithms to enhance the healthcare landscape, employing the transformative potential of cutting-edge machine learning technologies in the estimation of blood glucose concentration through meticulous regression analysis. Additionally, we have delved into the intricate correlation that exists between glucose concentration and hematocrit volume, unraveling new avenues of understanding.

Our research foundation was built upon a clinically validated dataset procured from an electrochemical glucose sensor, commonly known as a glucose strip. This dataset encompassed varying levels of glucose concentration and hematocrit volume. Leveraging this dataset, we developed a software program for a glucometer, a pivotal tool in our experimental approach. The crux of our methodology rested on the utilization of both linear regression and support vector regressor techniques, which were instrumental in predicting the values of glucose concentration (GC) and hematocrit volume (HV).

Our findings reveal that, through linear regression, the R2 score for GC approximates 0.916, whereas for HV, it stands at approximately 0.537. On the other hand, employing the support vector regressor, we attained an R2 score of roughly 0.961 for GC and 0.506 for HV. Upon a meticulous comparative analysis of these two-machine learning regressor models, we gleaned that the support vector regressor exhibited higher accuracy for the test dataset of glucose concentration. This marked our empirical validation of the potential efficacy of this approach in glucose concentration prediction.

In culmination, our methodological journey has significantly contributed to the domain of healthcare prediction by unearthing new horizons through the amalgamation of machine learning techniques and domain-specific insights. The newfound correlation between glucose concentration and hematocrit volume, coupled with the enhanced accuracy achieved through the support vector regressor, underscores the transformative capabilities of machine learning in revolutionizing healthcare analytics and fostering improved patient outcomes.

Data Availability. Data underlying the results presented in this paper are available from the corresponding author upon reasonable request.

Funding. There is no funding for this work.

Conflicts of interest. The authors declare no conflict of interest.

Ethics. The authors declare that the present research work has fulfilled all relevant ethical guidelines required by COPE.

REFERENCES

[1] D. S. (Dhaya) Battina, "**The Role of Machine Learning in Clinical Research: Transforming the Future of Evidence Generation,**" *International Journal of Innovations in Engineering Research and Technology*, vol. 4, no. 12, pp. 1–10, 2017, Accessed: Aug. 15, 2023. [Online]. Available: https://www.neliti.com/publications/429484/

[2] S. Chaudhuri, D. Oudejans, H. J. Thompson, and G. Demiris, **"Real World Accuracy and Use of a Wearable Fall Detection Device by Older Adults,"** *J Am Geriatr Soc*, vol. 63, no. 11, p. 2415, Nov. 2015, doi: 10.1111/JGS.13804.

[3] T. G. Dietterich, **"Editorial Exploratory research in machine learning,"** *Machine Learning 1990 5:1*, vol. 5, no. 1, pp. 5–9, Mar. 1990, doi: 10.1007/BF00115892.

[4] P. Chowriappa, S. Dua, and Y. Todorov**, "Introduction to machine learning in healthcare informatics,"** *Intelligent Systems Reference Library*, vol. 56, pp. 1–23, 2014, doi: 10.1007/978-3-642-40017-9_1.

[5] K. Kalaiselvi and M. Deepika, **"Machine Learning for Healthcare Diagnostics,"** pp. 91–105, 2020, doi: 10.1007/978-3-030-40850-3_5.

[6] "(PDF) Analysis of Medical Images Using Machine Learning." https://www.researchgate.net/publication/370444441_Analysis_of_Medical_Images_Using_Machine_Learning (accessed Aug. 15, 2023).

[7] M. Wernick, Y. Yang, J. Brankov, G. Yourganov, and S. Strother, "Machine learning in medical imaging," *IEEE Signal Process Mag*, vol. 27, no. 4, pp. 25–38, 2010, doi: 10.1109/MSP.2010.936730.

[8] M. W. Libbrecht and W. S. Noble, "**Machine learning in genetics and genomics,"** *Nat Rev Genet*, vol. 16, no. 6, p. 321, May 2015, doi: 10.1038/NRG3920.

[9] I. Landi *et al.*, **"Deep representation learning of electronic health records to unlock patient stratification at scale,"** *npj Digital Medicine 2020 3:1*, vol. 3, no. 1, pp. 1–11, Jul. 2020, doi: 10.1038/s41746-020-0301-z.

[10] N. J. Gogtay, R. Ravi, and U. M. Thatte**, "Regulatory requirements for clinical trials in India: What academicians need to know,"** *Indian J Anaesth*, vol. 61, no. 3, p. 192, Mar. 2017, doi: 10.4103/IJA.IJA_143_17.

[11] A. Qayyum, J. Qadir, M. Bilal, and A. Al-Fuqaha, **"Secure and Robust Machine Learning for Healthcare: A Survey,"** *IEEE Rev Biomed Eng*, vol. 14, pp. 156–180, Jan. 2020, doi: 10.1109/RBME.2020.3013489.

[12] P. Shah *et al.*, **"Artificial intelligence and machine learning in clinical development: a translational perspective,"** *NPJ Digit Med*, vol. 2, no. 1, Dec. 2019, doi: 10.1038/S41746-019-0148-3.

[13] M. Gusev *et al.*, **"Review Article Noninvasive Glucose Measurement Using Machine Learning and Neural Network Methods and Correlation with Heart Rate Variability,"** 2020, doi: 10.1155/2020/9628281.

[14] "(PDF) BGEM^{TM}: Assessing Elevated Blood Glucose Levels Using Machine Learning and Wearable Photo plethysmography Sensors." https://www.researchgate.net/publication/366443831_BGEM_Assessing_Elevated_Blood_Glucose_Levels_Using_Machine_Learning_and_Wearable_Photo_plethysmography_Sensors (accessed Aug. 15, 2023).

[15] "Bio-Resource: How does glucometer or Glucose monitoring device work?" http://technologyinscience.blogspot.com/2015/01/how-does-glucometer-or-glucose.html#.Y_2WdHZBy5d (accessed Feb. 28, 2023).

[16] X. Wang *et al.*, "Exploratory study on classification of diabetes mellitus through a combined Random Forest Classifier," *BMC Med Inform Decis Mak*, vol. 21, no. 1, Dec. 2021, doi: 10.1186/S12911-021-01471-4.

[17] Z. Tang, J. H. Lee, R. F. Louie, and G. J. Kost, **"Effects of Different Hematocrit Levels on Glucose Measurements With Handheld Meters for Point-of-Care Testing,"** *Arch Pathol Lab Med*, vol. 124, no. 8, pp. 1135–1140, Aug. 2000, doi: 10.5858/2000-124-1135-EODHLO.

[18] G. Sittampalam and G. S. Wilson, **"Surface-modified electrochemical detector for liquid chromatography,"** 2002, doi: 10.1021/AC00260A039.

[19] M. N. Szentirmay and C. R. Martin, **"Ion-Exchange Selectivity of Nafion Films on Electrode Surfaces,"** *Anal Chem*, vol. 56, no. 11, pp. 1898–1902, 1984, doi: 10.1021/AC00275A031/ASSET/AC00275A031.FP.PNG_V03.

[20] "Biochemistry Analyzer: Applications, Facts, and Benefits – Blog | Trivitron Healthcare Solutions | Medical Device Company." https://www.trivitron.com/blog/biochemistry-analyzer-applications-facts-and-benefits/ (accessed Feb. 28, 2023).

[21] A. AHeller and B. Feldman, **"Electrochemical glucose sensors and their applications in diabetes management,"** *Chem Rev*, vol. 108, no. 7, pp. 2482–2505, Jul. 2008, doi: 10.1021/CR068069Y/ASSET/IMAGES/CR-2006-08069Y_M011.GIF.

[22] P. K. Tiwari, H. Nam, and Y. S. Sohn**, "Correlation between blood glucose and hematocrit: A new estimation methodology,"** *Biochip J*, vol. 6, no. 3, pp. 206–212, Sep. 2012, doi: 10.1007/S13206-012-6302-5/METRICS.

[23] J. Hönes, P. Müller, and N. Surridge, **"The Technology Behind Glucose Meters: Test Strips,"** *https://home.liebertpub.com/dia*, vol. 10, no. SUPPL. 1, May 2008, doi: 10.1089/DIA.2008.0005.

[24] J. Wang, "Electrochemical glucose biosensors," *Chem Rev*, vol. 108, no. 2, pp. 814–825, Feb. 2008, doi: 10.1021/CR068123A.

[25] A. Heller and B. Feldman, **"Electrochemical glucose sensors and their applications in diabetes management,"** *Chem Rev*, vol. 108, no. 7, pp. 2482–2505, Jul. 2008, doi: 10.1021/CR068069Y/ASSET/IMAGES/LARGE/CR-2006-08069Y_0001.JPEG.

[26] A. Amine, J. M. Kauffmann, G. J. Patriarche, and G. G. Guilbault, **"Electrochemical Behaviour of a Lipid Modified Enzyme Electrode,"** *http://dx.doi.org/10.1080/00032718908052362*, vol. 22, no. 11–12, pp. 2403–2411, 2006, doi: 10.1080/00032718908052362.

[27] J. J. Khanam and S. Y. Foo, **"A comparison of machine learning algorithms for diabetes prediction,"** *ICT Express*, vol. 7, no. 4, pp. 432–439, Dec. 2021, doi: 10.1016/J.ICTE.2021.02.004.

[28] K. Plis, R. Bunescu, C. Marling, J. Shubrook, and F. Schwartz, **"A Machine Learning Approach to Predicting Blood Glucose Levels for Diabetes Management"**, Accessed: Aug. 15, 2023. [Online]. Available: www.aaai.org