Failure Detection in Continuous Glucose Monitors Using Transformers

Francisco Gonzalez Barriga, Dulce Ibarra Montes, Alma Y. Alanis, and Gabriel Martinez-Soltero

Centro Universitario de Ciencias Exactas e Ingenierías, Universidad de Guadalajara, Guadalajara 44430, Mexico
erasmo.martinez@academicos.udg.mx

Abstract. In this paper, we propose a method for detecting failures in Continuous Glucose Monitors (CGMs) using Transformer-based neural networks. CGMs are critical for diabetes management, providing continuous glucose level readings. However, failures such as spikes, and loss of sensitivity can compromise the accuracy of these devices, posing risks to users' health. The Transformer architecture, known for its ability to process sequential data efficiently, to classify CGM signal segments and identify faults. The proposed approach is compared against classical methods such as Long Short-Term Memory, Multi-Layer Perceptron, and Support Vector Machines. The Transformer model achieved the highest F1-score, outperforming the other approaches. Our results demonstrate the potential of Transformers in improving fault detection in CGMs, ensuring more reliable diabetes management.

Keywords: Transformers · Glucose continuous monitor · Fault detection.

1 Introduction

Continuous glucose monitors (CGM) are at the forefront of the transformation of diabetes management, offering a revolutionary approach to monitoring blood glucose levels. Diabetes, a chronic condition affecting over 463 million adults worldwide (diagnosed) [17], is characterized by the body's inability to produce or effectively use insulin, leading to elevated blood glucose levels [14] [3]. Without proper management, diabetes can result in severe complications, including cardiovascular disease, kidney failure, and neuropathy [13]. Traditional methods of glucose monitoring, such as periodic finger-prick tests, are invasive, inconvenient, and provide only snapshots of glucose levels. In contrast, CGMs provide continuous, real-time readings of glucose levels throughout the day, enabling patients and healthcare providers to make more informed decisions about insulin dosing, diet, lifestyle adjustments, and providing continuous readings of glucose levels in blood throughout the day without relying on periodic finger prick tests [23].

CGMs work by using a small sensor inserted under the skin to measure glucose levels in the interstitial fluid. These sensors transmit data wirelessly to a

https://doi.org/10.61728/AE20255312

receiver or smartphone app, allowing users to track their glucose trends and receive alerts when levels fall outside predefined thresholds [1]. This continuous stream of data not only improves glycemic control but also reduces the risk of hypoglycemic and hyperglycemic episodes, which can be life-threatening. However, like any electronic device, glucose monitors can experience failures that affect the accuracy of measurements and put users' health at risk[11]. Failures such as spikes, loss of sensitivity, false detections can lead to unreliable data and potentially harmful treatment decisions [23]. These failures can arise from various factors, such as sensor degradation, electromagnetic interference, or improper insertion. For example, a sensor spike might result in an erroneously high glucose reading, prompting unnecessary insulin administration and increasing the risk of hypoglycemia. Conversely, a loss of sensitivity might lead to undetected hyperglycemia, leaving the patient vulnerable to long-term complications.

Early detection of glucose monitor failures is crucial to ensure the safety and efficacy of diabetes treatment. Traditional methods of fault detection, such as laboratory testing and visual inspections, can be costly, time-consuming and are not always reliable. In recent years, the use of techniques, such as transformers [18], Kalman filter [10], some nonelinear filters [12], k-medoids clustering algorithm [22], has been explored to improve fault detection in glucose monitors[25, 9].

A Transformer (attention-based encoder–decoder model) is a deep learning neural network architecture designed to process data streams, such as text, efficiently and in parallel [19]. It was introduced in 2017 in the article "Attention is All You Need." [20] and has revolutionized the field of Natural Language Processing (NLP) [8]. Unlike previous models, such as Recurrent Neural Networks (RNN) or Long-Term Memory Networks (LSTM), Transformers do not process sequences sequentially, but analyze all parts of the sequence simultaneously, which allows capturing long-term dependencies and complex relationships between elements [19].Unlike convolutional networks, Transformers require minimal inductive biases in their design and are naturally suited as ensemble functions [6]. Transformers can be seen from threee perspectives: architectural modification, pre-training, and applications [7]. Transformers are widely used in various areas, such as medical image synthesis/reconstruction, registration, segmentation, detection and diagnosis, as well as medical image analysis [4]. They are also used in the field of automatic language processing (ALP) and computer vision, for modeling time series data [21], among others. This article proposes a method for fault detection in continuous glucose monitors (CGMs) using Transformersdue to its performance shown. The article is structured as follows: Section 2 is the methodology , Section 3 presents the experimental results obtained through confusion matrices. Finally, Section 4 concludes the paper and discusses possible future applications.

2 Methodology

Failure detection will be performed over a delayed vector of size $d = 16$ containing the measurements of glucose sensor then input of the approaches is $X = \{x_1, x_2, \ldots, x_{16}\}$ and a binary output where one means without failure and zero means with failure. The transformer is compared against three classical approaches to classify, Long short-Term Memory neural network (LSTM) [5], Multi-Layer Perceptron (MLP)[16, 15], and Support Vector Machine (SVM)[2].

The dataset used is Shanghai_T2DM presented in [24] which contains measurements of 100 patients for a period of 3 to 14 days with a sensing frequency of 15 minutes, measurements of a patient are shown in Fig.1.

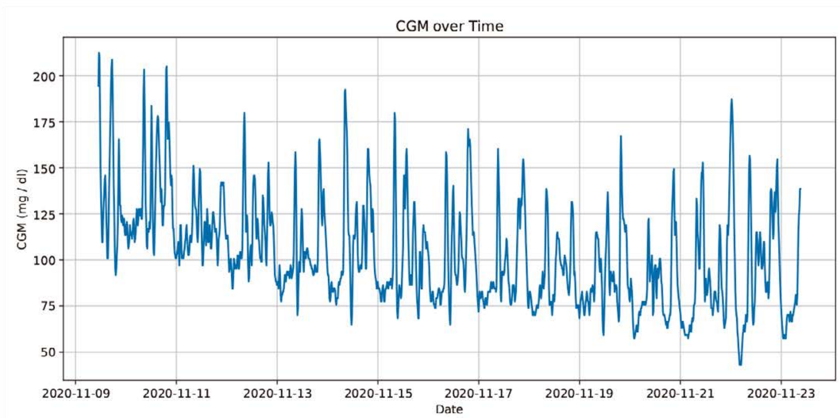


Fig. 1: signal of the CGM

Among the time series of some patients there are some instants that were omitted, to fill the measurements absent, the previous value was preserved. With the times series consistent three types of failures were generated artificially, noise, when it goes to zero and holding the previous value as is shown in Fig.2. Then a new dataset is produced forming vectors of size d taking all the measurements with their previous $d-1$ values. The previous dataset ends with 82,671 samples for training and 28,687 samples for testing. For the transformer the input embedding block was excluded as is presented in Fig. 3. Each one of the techniques were trained for 100 epochs with a binary cross entropy as loss function except for de SVM which uses a Kernel of radial base.

As the test dataset is unblanced having 21487 without failure and 7200 with failure being an unbalanced dataset leading to probles such as the accuracy is not a reliable metric, the models tend to class with more samples more often since it may not learn the characteristics of the minority class properly. After fitting

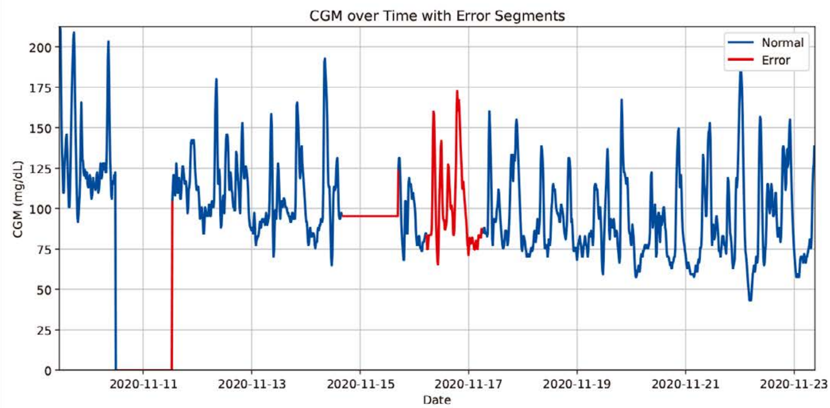


Fig. 2: Time series with artificial errors generated

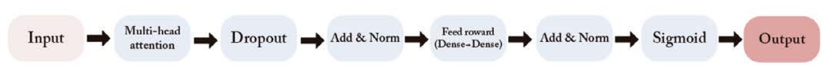


Fig. 3: Blocks of the transformer

the model, the metrics of accuracy, precision also called positive predictive rate, recall named as true positive rate and F1 score were calculated according to

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$precision = \frac{TP}{TP + FP} \tag{2}$$

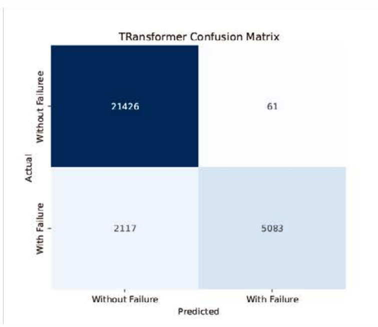
$$recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1 - score = 2 \cdot \frac{precision \cdot recall}{precision + recall} \tag{4}$$

where TP is the number of correct samples classified as the class of interest in this case are the samples correctly identified as faults, FP are the samples of the other class incorrectly identified as the class of interest, FN is the number of samples incorrectly classified as which are of the class of interest but were classified as other class TN are the correct classification of no interest class.

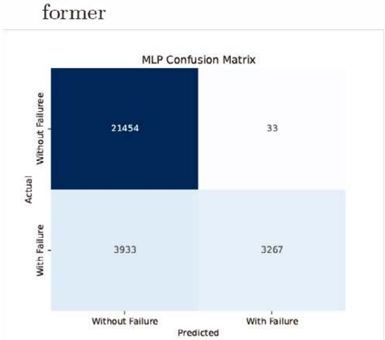
3 Results

To measure the performance of the classifiers, a test dataset consisting of 28,687 samples was used. The results were evaluated using confusion matrices, which provide a detailed breakdown of the model's predictions compared to the actual labels. These confusion matrices, shown in Fig. 4, summarize the performance of the algorithm and allow for further inspection of its effectiveness. To quantify the performance of the classifiers, several key metrics were calculated from the confusion matrices: Accuracy measures the proportion of correctly classified samples out of the total number of samples. It is a general indicator of the model's overall performance. Precision indicates the proportion of true positive predictions out of all positive predictions made by the model. It reflects the model's ability to avoid false positives. Recall represents the proportion of true positives correctly identified out of all actual positives. It measures the model's ability to detect all relevant instances. F1 Score the harmonic mean of precision



(a) Confusion matrix of the transformer

(b) Confusion matrix of the LSTM



(c) Confusion matrix of the MLP

(d) Confusion matrix of the SVM

Fig. 4: Confusion matrix of all approaches to identify failure

and recall, providing a balanced measure of the model's performance, especially useful when dealing with imbalanced datasets.

The metrics calculated from the confusion matrices are presented in Table 1. The results highlight the following key findings: The transformer-based classifier achieved the best overall performance, with an F1 score of 0.8236. This score integrates both precision and recall, showing the model's ability to balance accurate predictions with comprehensive detection of positive cases. The transformer outperformed the other models in most metrics, showing an 8% improvement in F1 score compared to the second-best model. While the Multilayer Perceptron (MLP) achieved the highest precision among the classifiers, its overall performance was lower than the transformer. This suggests that the MLP is effective at minimizing false positives but may struggle with false negatives, as reflected in its lower recall and F1 scores. The remaining classifiers (e.g., LSTM, SVM) showed competitive results but were consistently outperformed by the transformer.

The superior performance of the transformer model can be attributed to its ability to capture complex patterns and dependencies in the data, particularly in sequential or time-series datasets like those generated by continuous glucose monitors (CGMs). Unlike traditional models, transformers take advantage self-attention mechanisms to analyze the entire sequence of data simultaneously, enabling them to detect subtle anomalies and long-term trends that other models might miss.

The MLP's higher precision indicates that it is more conservative in making positive predictions, which can be advantageous in scenarios where false positives are costly. However, its lower recall and F1 score suggest that it may miss some true positive cases, which could be critical in applications like fault detection in CGMs, where missing a failure could have serious health implications.

Table 1: Classification Metrics

Approach	Accuracy	Precision	Recall	F1 Score
Transformer	**0.9241**	0.9881	**0.7060**	**0.8236**
LSTM	0.8946	0.8953	0.6568	0.7577
MLP	0.8617	**0.9900**	0.4537	0.6223
SVM	0.8607	0.9852	0.4518	0.6195

4 Conclusions

The proposed method uses Transformer neural network architecture to effectively detect faults in CGMs. By analyzing segments of CGM time series, the Transformer model showed superior performance compared to traditional methods such as LSTM, MLP, and SVM, achieving an F1-score of 82.36%. This highlights the ability of transformers to capture complex patterns in sequential data, making it a promising tool for improving the reliability of sensors. The results suggest that Transformers can play a significant role in enhancing fault detection systems, ultimately contributing to safer and more accurate diabetes management. Future work could explore the application of this approach to other medical devices and real-time monitoring systems.

References

1. Clarke, W., Kovatchev, B.: Statistical tools to analyze continuous glucose monitor data. Diabetes technology & therapeutics **11**(S1), S–45 (2009)
2. Cortes, C., Vladimir, V.: Support-vector networks. Machine Learning (1995)
3. Egan, A.M., Dinneen, S.F.: What is diabetes? Medicine **47**(1), 1–4 (2019)
4. He, K., Gan, C., Li, Z., Rekik, I., Yin, Z., Ji, W., Gao, Y., Wang, Q., Zhang, J., Shen, D.: Transformers in medical image analysis. Intelligent Medicine **3**(1), 59–78 (2023)
5. Hochreiter, S.: Long short-term memory. Neural Computation MIT-Press (1997)
6. Khan, S., Naseer, M., Hayat, M., Zamir, S.W., Khan, F.S., Shah, M.: Transformers in vision: A survey. ACM computing surveys (CSUR) **54**(10s), 1–41 (2022)
7. Lin, T., Wang, Y., Liu, X., Qiu, X.: A survey of transformers. AI open **3**, 111–132 (2022)
8. Liu, Y., Zhang, Y., Wang, Y., Hou, F., Yuan, J., Tian, J., Zhang, Y., Shi, Z., Fan, J., He, Z.: A survey of visual transformers. IEEE Transactions on Neural Networks and Learning Systems (2023)
9. Lu, Y., Liu, D., Liang, Z., Liu, R., Chen, P., Liu, Y., Li, J., Feng, Z., Li, L.M., Sheng, B., et al.: A pretrained transformer model for decoding individual glucose dynamics from continuous glucose monitoring data. National Science Review p. nwaf039 (2025)
10. Mahmoudi, Z., Boiroux, D., Hagdrup, M., Nørgaard, K., Poulsen, N.K., Madsen, H., Jørgensen, J.B.: Application of the continuous-discrete extended kalman filter for fault detection in continuous glucose monitors for type 1 diabetes. In: 2016 European Control Conference (ECC). pp. 714–719. IEEE (2016)
11. Mahmoudi, Z., Nørgaard, K., Poulsen, N.K., Madsen, H., Jørgensen, J.B.: Fault and meal detection by redundant continuous glucose monitors and the unscented kalman filter. Biomedical Signal Processing and Control **38**, 86–99 (2017)
12. Mahmoudi, Z., Wendt, S.L., Boiroux, D., Hagdrup, M., Nørgaard, K., Poulsen, N.K., Madsen, H., Jørgensen, J.B.: Comparison of three nonlinear filters for fault detection in continuous glucose monitors. In: 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). pp. 3507–3510. IEEE (2016)
13. Molina, R., Rodríguez, C., et al.: Definición, clasificación y diagnóstico de la diabetes mellitus. Revista venezolana de endocrinología y metabolismo **10**, 7–12 (2012)
14. Reyes Sanamé, F.A., Pérez Álvarez, M.L., Alfonso Figueredo, E., Ramírez Estupiñan, M., Jiménez Rizo, Y.: Tratamiento actual de la diabetes mellitus tipo 2. Correo científico médico **20**(1), 98–121 (2016)
15. Rosenblatt, F.: The perceptron: a probabilistic model for information storage and organization in the brain. Psychological review **65**(6), 386 (1958)
16. Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Learning representations by back-propagating errors. nature **323**(6088), 533–536 (1986)

17. Salgado, V.R., Hernández, J.L., López, G.M.: Descontrol glucémico asociado a la falta de bienestar emocional en personas con diabetes mellitus (2025)
18. Thomas, J.B., Chaudhari, S.G., Shihabudheen, K., Verma, N.K.: Cnn-based transformer model for fault detection in power system networks. IEEE Transactions on Instrumentation and Measurement **72**, 1–10 (2023)
19. Tunstall, L., Von Werra, L., Wolf, T.: Natural language processing with transformers. " O'Reilly Media, Inc." (2022)
20. Vaswani, A.: Attention is all you need. Advances in Neural Information Processing Systems (2017)
21. Wen, Q., Zhou, T., Zhang, C., Chen, W., Ma, Z., Yan, J., Sun, L.: Transformers in time series: A survey. arXiv preprint arXiv:2202.07125 (2022)
22. Yu, X., Sun, X., Zhao, Y., Liu, J., Li, H.: Fault detection of continuous glucose measurements based on modified k-medoids clustering algorithm. Neural Computing and Applications pp. 1–10 (2020)
23. Zhao, C., Fu, Y.: Statistical analysis based online sensor failure detection for continuous glucose monitoring in type i diabetes. Chemometrics and Intelligent Laboratory Systems **144**, 128–137 (2015)
24. Zhao, Q., Zhu, J., Shen, X., Lin, C., Zhang, Y., Liang, Y., Cao, B., Li, J., Liu, X., Rao, W., et al.: Chinese diabetes datasets for data-driven machine learning. Scientific Data **10**(1), 35 (2023)
25. Zhu, T., Kuang, L., Piao, C., Zeng, J., Li, K., Georgiou, P.: Population-specific glucose prediction in diabetes care with transformer-based deep learning on the edge. IEEE Transactions on Biomedical Circuits and Systems (2024)