

RESEARCH ARTICLE

Transformer-based DNA methylation detection on ionic signals from Oxford Nanopore sequencing data

Xiuquan Wang¹, Mian Umair Ahsan², Yunyun Zhou^{2,*}, Kai Wang^{2,3,*}

¹ Department of Mathematics and Computer Science, Tougaloo College, Jackson, MS 39174, USA

² Raymond G. Perelman Center for Cellular and Molecular Therapeutics, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA

³ Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

* Correspondence: wangk@chop.edu; zhouy6@chop.edu

Received November 19, 2022; Revised November 27, 2022; Accepted November 27, 2022

Background: Oxford Nanopore long-read sequencing technology addresses current limitations for DNA methylation detection that are inherent in short-read bisulfite sequencing or methylation microarrays. A number of analytical tools, such as Nanopolish, Guppy/Tombo and DeepMod, have been developed to detect DNA methylation on Nanopore data. However, additional improvements can be made in computational efficiency, prediction accuracy, and contextual interpretation on complex genomics regions (such as repetitive regions, low GC density regions).

Method: In the current study, we apply Transformer architecture to detect DNA methylation on ionic signals from Oxford Nanopore sequencing data. Transformer is an algorithm that adopts self-attention architecture in the neural networks and has been widely used in natural language processing.

Results: Compared to traditional deep-learning method such as convolutional neural network (CNN) and recurrent neural network (RNN), Transformer may have specific advantages in DNA methylation detection, because the self-attention mechanism can assist the relationship detection between bases that are far from each other and pay more attention to important bases that carry characteristic methylation-specific signals within a specific sequence context.

Conclusion: We demonstrated the ability of Transformers to detect methylation on ionic signal data.

Keywords: Nanopore; long-read sequencing; deep learning; Transformer model; DNA methylation.

Author summary: Transformer is an algorithm that adopts self-attention architecture in the neural networks and has been widely used in natural language processing. In the current study, we apply Transformer architecture to detect DNA methylation on ionic signals from Oxford Nanopore sequencing data. We evaluated this idea using real data sets (*Escherichia coli* data and the human genome NA12878 sequenced by Simpson *et al.*) and demonstrated the ability of Transformers to detect methylation on ionic signal data.

INTRODUCTION

The complexity of the human genome and transcriptome lies not only in the composition of 3 billion base pairs, but also in the chemical modifications that make it interpretable to enzymes (writers, erasers, readers) through epigenetic regulation [1]. Genome-wide epigenetic change, such as DNA 5-methylcytosine

(5mC), is a hallmark of cancer [2–4]. It is also widely known that 5mC methylation play important roles in brain development and function [5–9]. In addition to being diagnostic biomarkers in many diseases, DNA methylation is now a therapeutic target for cancer with several drugs being tested or approved by the US Food and Drug Administration [10,11]. For example, 5-Aza-2'-deoxycytidine is among the first methylation inhibitor

used in cancer clinical trials [12], and we demonstrated that it leads to isoform switching and exon skipping such as EZH2 in addition to de-methylation [13]. Similarly, anti-psychotic treatments have been linked to the alteration of DNA methylations [14], suggesting the potential of differential DNA methylation profiles as predictors of antipsychotic response.

Existing technologies, such as whole-genome bisulfite sequencing and PacBio long-read sequencing, have some inherent limitations to detect DNA modifications, because of the inability to detect modification in complex repetitive region, biases by incomplete and context-dependent enzyme, and low signal-to-noise ratio [15,16]. Instead, Oxford Nanopore Technology (ONT) long-read sequencing, which measures ionic current signals when DNA molecules translocate pores, may be a better option for the detection of DNA methylation. Recently, several analytic tools have been developed for DNA methylation detection on ONT long-read sequencing data. The analysis methods can be generally classified into two types. One type of method compares raw signals of methylated DNA copies with signals of the un-methylated DNA copies at specific genomic positions, such as Tombo/Nanoraw and NanoMod [17]. However, this method requires both methylated and un-methylated samples that are available at the same time. Another type of method directly calls DNA modifications from ONT raw signals using machine learning approaches, such as Nanopolish [18], Megalodon [19], DeepSignal [20], Guppy [21], METEORE [22], and DeepMod [23]. For example, METEORE used random forest and multiple linear regression models, NanoPolish used hidden Markov model, and DeepSignal used convolutional neural network (CNN).

We previously developed DeepMod [23] which adopted a recurrent neural network (RNN) in a bidirectional LSTM architecture to detect DNA methylation from ionic signal data generated by ONT. We also recently released DeepMod2 that can handle moving tables generated by Guppy-called or Tombo re-squiggled data, since DeepMod requires event table used in older generations of data. In DeepMod, raw signals of each read are first translated into nucleotide sequences (basecalling). Signals are then aligned to corresponding reference nucleotides. After that, the target motif (e.g., CpG) and its signals in a window of a fixed length are transformed into event-based features as the input of methylation callers. Typical event-based features include signal mean, signal standard deviation, event length, and nucleotide information, which tells the reference base in one hot encoding of the ACGT bases. The LSTM model bi-directionally reads given nucleotide events sequentially (left-to-right or right-to-left). In

DeepMod, we have released several pre-trained models from different types of datasets including *Escherichia coli*, *Chlamydomonas reinhardtii* and human samples and demonstrated that it achieves good performance on these datasets. We note that Liu *et al.* comprehensively summarized current methods and made a comparison for human DNA methylation detection when comparing DeepMod with other tools such as Nanopolish and Tombo [24]. DeepMod's DNA methylation detection was based on older basecallers (such as Metrichore), but other tools were based on Guppy basecalled data, therefore they are not comparable; instead, DeepMod2 should be used in this case.

Although DeepMod achieves good performance for different types of datasets according to several studies [22,25–27], we believe that the base modification problem is a language translation problem fundamentally and can be further improved. Since Transformers have proven to perform better than RNN and CNN in language modeling, here we leverage Transformers to identify modified DNA bases from signal data. Unlike LSTM, Transformer algorithms, such as BERT [28], do not necessarily process the input data in sequential order. Indeed, using DeepMod's signal pre-processing features, Zhang *et al.* further proposed MethBERT [25] utilizing a refined BERT method to detect DNA modification on ONT long-read sequencing data.

Transformers adopt the mechanism of self-attention, differentially weighting the significance of each part of the input data. The essence of the self-attention mechanism is to detect the most useful information from many pieces of information. Just like the human brain, it will scan the entire field of view through the eyes, and then quickly locate the area of interest. Then it devotes more attention resources to a specific area to obtain more detailed information about the target that needs attention, while ignoring other less useful information. This is a strategy by which humans use limited attention resources to quickly focus on high-value information from a large amount of information. Multi-head attention allows the model to jointly attend to information from different representation subspaces at different positions. Transformer's self-attention mechanism greatly improves the efficiency and accuracy of information processing.

In this project, we explored the use of a Transformer-based BERT to further improve the methylation detection, and performed several experiments to answer the following questions: (1) how signal distribution differs in windows of varying sizes (such as 21 bp) surrounding methylated versus unmethylated cytosine, or in regions with enriched methylation (such as CpG islands) versus isolated methylation sites; (2) how the contextual signal changes with the length of the event

window; (3) whether the number of heads in Transformer model will influence the accuracy of methylation detection.

RESULTS

Comparison of signal features for methylated and unmethylated sites in human genome

The general framework of our method is shown in Fig. 1. Our method can directly detect DNA modification on ionic signals from ONT long-read sequencing data based on the signal difference on methylated and unmethylated DNA. To give a visual demonstration, Fig. 2 shows an example of how raw contextual signal is observed for 21-bp window size on methylated and unmethylated location for CpG site at chr2:159219855 (GRCh38) on NA12878. This Fig. 2 shows that the mean signal for methylated and un-

methylated location is different, and the CpG signal is influenced by its surrounding nucleotides.

Fine-tuning of model parameters

To search for the best hyperparameters and optimize the model performance, we fine-tuned the number of attention heads, the window size of events, and the feature size of each event. Simulation results of these hyperparameters tested on human genome NA12878 are shown in Fig. 3. Detailed explanations for these experiments are given below.

Performance with different position embedding (PE) strategy: summation vs. concatenation

The first experiment was to check the performance of the model on different position embedding method. We chose two different ways to combine our data feature

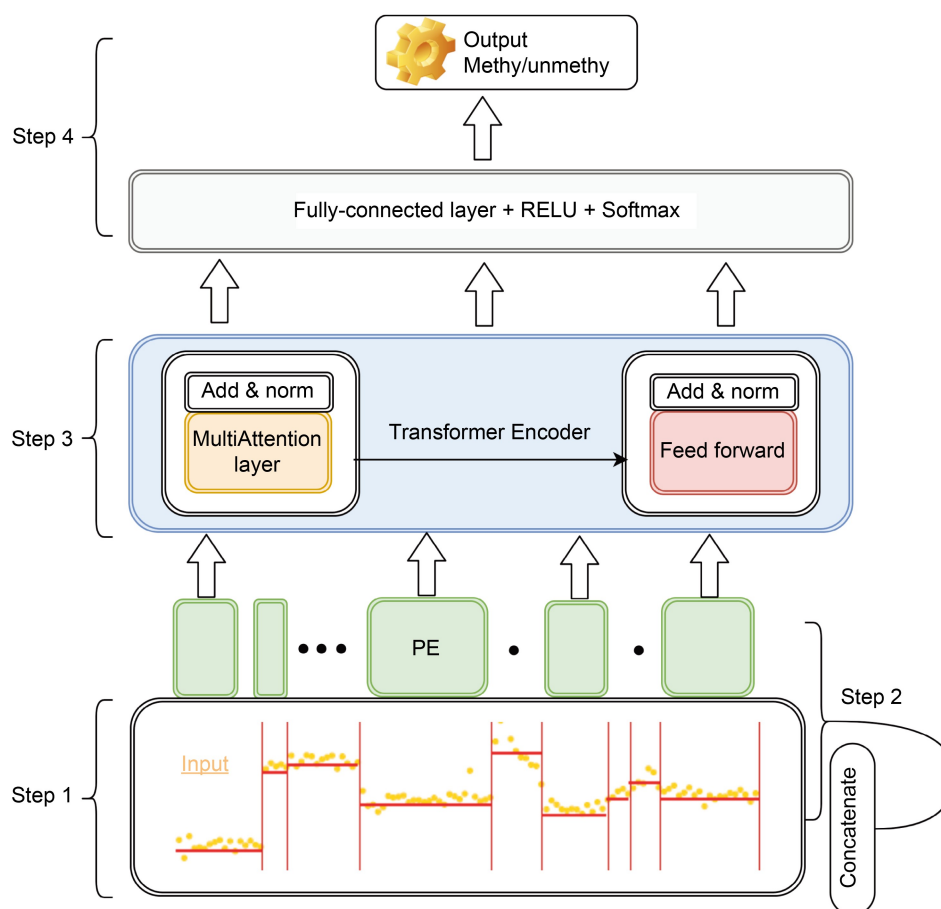


Figure 1. Architecture of the transformer-based DNA methylation detection on ONT long-read sequencing data. Step 1: Data pre-processing: aligned lone read Nanopore signals to reference genome GRCh38 to build a 7-feature vectors for each event. Step 2: Position E Embedding (PE) matrix and then concatenate with 7-feature vectors as input of the model. Step 3: Transformer Encoder and modification prediction. Step 4: Methylation classification.

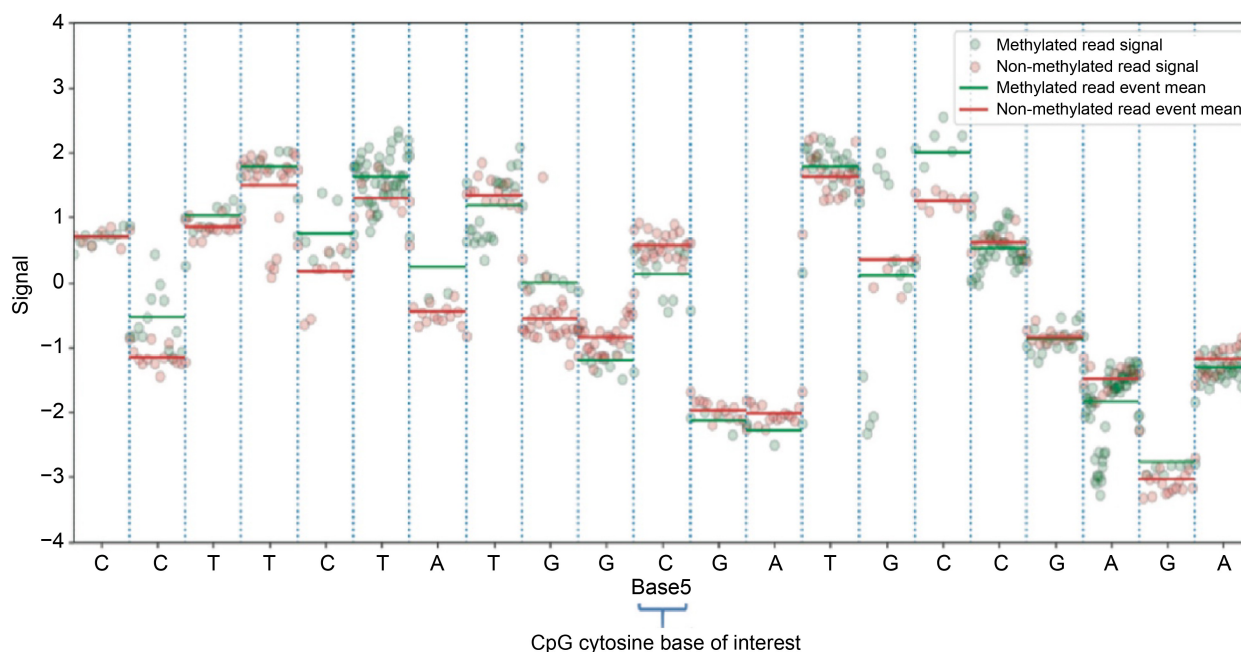


Figure 2. Signal visualization. Nanopore signals for $w=21$ events for CpG site at chr2:159219855 (GRCh38). Green and red dots represent signals of one read each from methylated and unmethylated control samples, respectively. Horizontal lines represent mean signal value for each event. Bases adjacent to the methylation site show varying degrees of differences between methylated and unmethylated read.

vector and position embedding, *i.e.*, summation vs. concatenation, and we then ran 50 epochs on our validation data. We found that concatenation method was much better than summation on performance (Fig. 3A).

Performance with different number of attention head

We choose three attention head size (2, 4, and 8) to evaluate our model's preformation (Fig. 3B). Here, we set our d_{model} size to be equal to 16. We found that with 2 attention head, our model has the best performance. That is probably because 16 features represent a relatively small feature size; if we add more attention head to it, it may result in overfitting.

Performance with different size of d_{model}

As mentioned above, when we implement the PE method, we found that the performance of concatenating PE vector with event vector is better than that of direct summation. However, the difference of these two methods may also be influenced by the size of d_{model} , *i.e.*, feature size of each event vector. Therefore, we checked how the feature size of each event will affect the performance on our model. We start from feature size of 16 (concatenated 9-feature position embedding to the original 7-feature event), and continue to increase 16

more position embedding size each time. We found that the performance significantly improved during the first 15 epochs, and then converged to similar values at 50 epochs (Fig. 3C).

Performance with different window sizes

The last experiment we did was to check the model performance if we add more neighboring events on both sides of the event that we are interested in, *i.e.*, given more contextual content by changing the window size. We found that 21 is the best window size to use in our model (Fig. 3D).

After fine-tuning the parameters to optimize the performance of the model, we suggest the values of the parameter at Table 1.

Comparison of embedding patterns before and after concatenating PE

In order to assess how concatenating PE vectors influence the performance, we showed the embedding patterns before and after PE concatenation. When we choose 8 embedding vector size (Fig. 4A), the pattern of the embedding is not clear: after we concatenate this embedding vector to the event features, the order of the event and the contextual relationship are not clearly seen. When we raise the embedding dimension to 48, we

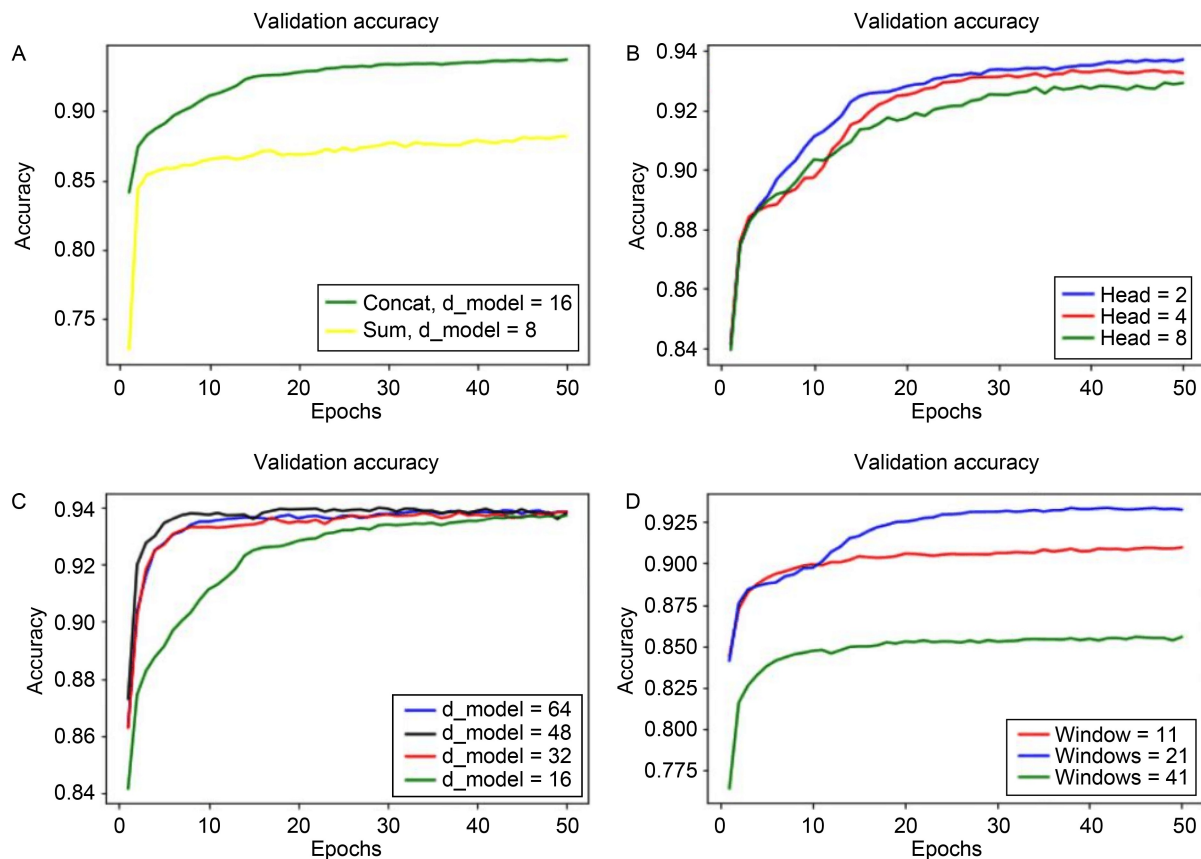


Figure 3. Experiments for model optimization. A demonstration of performance on Simpson's *E. coli* data with position embedding method (A), different attention head number (B), d_model size (C), and window size (D).

Table 1 Configuration of values for hyper-parameter tuning

Parameter	Annotation of parameters	Type	Scope
d_model	Size of feature vector	Even integer	[32, 64]
Att. head	Number of attention heads	Integer	[27]
Att. layer	Number of attention layers	Integer	[27]
Batch size	Batch size	Integer	[256, 512]
Learning rate	Learning rate	Real	[0.001, 0.01]

Att. means attention.

can easily see an increasing order of events based on the pattern on the hidden dimension (Fig. 4B). After the hidden dimension is large enough to show the order of events, increasing dimension further will not improve the model performance.

Performance comparison across human genome and bacteria species

We also performed cross-species evaluation, and tested our model on human genome NA12878 and *E. coli* by training on one genome and testing on another. Figure 5 shows the ROC curves for 5mC detection on both species.

As we can see from the Fig. 5, our method achieves good performance on both *E. coli* (AUC = 0.99) and NA12878 (AUC = 0.95). We further compared the performance with DeepMod and DeepSignal. We found that DeepMod achieves higher F1 scores on both human and bacteria genomes than the other two methods, as shown in Table 2.

CONCLUSION AND DISCUSSION

In the current study, we propose a Transformer-based method for detecting DNA methylations from ionic signal data generated by Nanopore sequencing. We begin with the hypothesis that the performance of 5mC

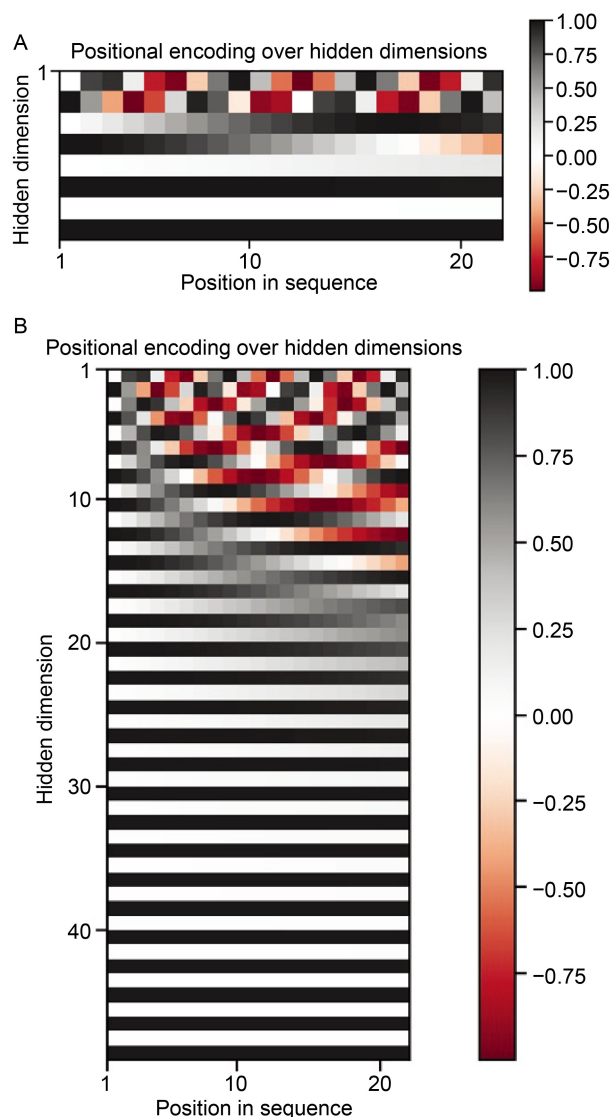


Figure 4. Embedding pattern visualization. Position embedding with different size for each event. (A) 8 embedding vector size. (B) 48 embedding vector size.

prediction may be improved using more sophisticated deep neural networks—Transformer. Then by modifying the encoding part of the Transformer, we can capture the difference of properties between methylated and unmethylated bases. Our results demonstrate preliminary success of Transformers in detecting methylations, but the current model did not perform optimally yet (for example, DeepMod outperformed the current transformer model). We stress that this is an exploratory study to see whether Transformer may work on methylation detection, since the methBERT study did not evaluate different model architectures (such as position embedding strategies) comprehensively.

We believe that DNA methylation detection closely

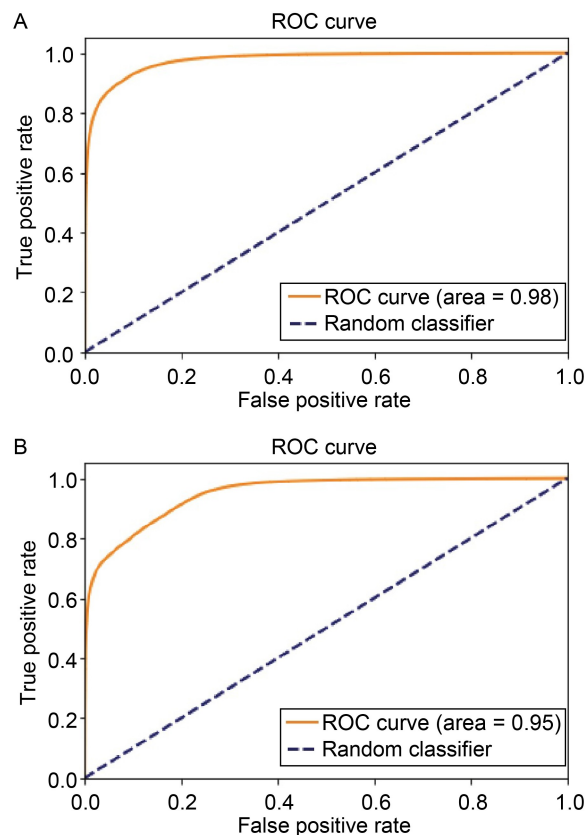


Figure 5. Performance evaluation. ROC curve for the evaluation of the performance on 5mC prediction using *E. coli* (A) and human genome NA12878 (B).

Table 2 Performance of Transformer-based method comparing with DeepMod and Deepsignal on NA12878 and *E. coli*

Dataset	Models	F1 score
NA12878	Transformer method	0.8600
	DeepMod	0.8706
	deepsignal	0.8520
<i>E. coli</i>	Transformer method	0.9410
	DeepMod	0.9501
	deepsignal	0.9330

resembles language translation because of both short-term distance dependency and long-term distance dependency. When DNA strands translocate Nanopore, approximately 7 nucleotides are covered within the pore, so they had the greatest contribution to the signal patterns. However, adjacent nucleotides, before or after the translocation, as well as the sequence contexts (such as the formation of secondary structure or location within a specific sequence motif), also determines the signal patterns in Nanopore sequencing. Therefore, such relationships closely resemble language translation,

which we believe can be addressed by employing methods used in NLP tasks.

In the past few years, the field of NLP is revolutionized by the use of various Transformer models. Transformers, such as BERT, is able to achieve better performance than RNN [29] and enable parallel computing with faster running speed. Parallel processing is particularly useful for processing Nanopore signals, because it can speed up the prediction of modifications from Nanopore signals on large data set from PromethION flowcells (>1 TB/flowcell). Additionally, self-attention in Transformer can efficiently capture long-range dependencies, which is a critical issue that RNN may not address well.

To the best of our knowledge, only one study (MethBERT) used BERT model to detect DNA methylation from ONT data. MethBERT utilized the DeepMod framework to perform the same pre-processing for the electrical raw signal data and then used a refined BERT model as a core (instead of LSTM) to detect DNA methylation. The refined BERT uses learning PE and relative position representation. The learnable PE takes positional embedding vectors as parameters, which are updated during the learning process. Their experiment show that the refined BERT can achieve competitive and even better results than the state-of-the-art bidirectional recurrent neural network (bi-RNN) model on a set of 5mC and 6mA benchmark datasets while the model inference speed is about 6x faster.

Similar to DeepMod's approach using a bi-RNN, Transformers implements the concept of two-way signal processing as well, but are more efficient in parallelization. Based on our experiments, Transformer-based method did not significantly outperform DeepMod's LSTM method for both human and bacteria genome. We should note that the performance of a model is dependent on the task and the data, and these factors should be taken into consideration when choosing a model. Additional improvements in this field requires innovations that directly assay signals (without basecalling), rather than two-step deep learning-models. In the future, we plan to improve the model from the following perspective: (1) computational speed; (2) how we embed the signal; (3) generate more training samples.

METHODS

We developed a Transformer-based method for the detection of DNA modifications from Nanopore sequencing data. The Nanopore sequencing dataset we used for training and testing for 5mC detection include *E. coli* data and the human genome NA12878 sequenced

by Simpson *et al.* [18]. The human genome NA12878 has been well-studied with various sequencing data, including Nanopore, PacBio, Illumina bisulfite sequencing with two replicates, and methylation microarrays.

As the framework of our method shown in Fig. 1, it consists of four steps: anchoring signals to reference positions after read alignment, feature generation, modification prediction via Transformer model, and genome-scale modification summary.

Step 1: The input includes a reference genome and FAST5 files containing raw signals and events (nucleotide, A, C, G or T), which were generated by Nanopore sequencers with base calls. Each event is then encoded in one-hot form in the order of (ACGT) and was represented with a 4-feature vector (*e.g.*, $C := \langle 0, 1, 0, 0 \rangle$). In this 4-feature vector, 1 indicates that the mapped reference base is a specific nucleotide type C, whereas 0 means otherwise. Raw signals for all aligned bases in a long read were normalized and rescaled from -5 to 5. Then the signal mean, standard deviation, and the number of signals associated with an event were extracted. Thus 7-feature vector was obtained $x_i = \langle f_m, f_{sd}, f_l, f_A, f_C, f_G, f_T \rangle$.

The self-attention heads (attention modules) can learn contextual relations between nucleotides. Therefore, for the interested event x_i , we take its $\left\lfloor \frac{w}{2} \right\rfloor$ upstream events and $\left\lfloor \frac{w}{2} \right\rfloor$ downstream events into consideration as context, *i.e.*

$$x = \langle x_{i-\frac{w}{2}}, \dots, x_i, \dots, x_{i+\frac{w}{2}} \rangle$$

By default, $w = 21$, but this is a parameter that can be changed by users.

The problem with the event sequence x is that it only records the base type and some basic statistical information of each event, but does not record the position information of these events in the sequence. We know that when the same event appears in different position in the sequence, its function or signal characteristics may be completely different. So, we should also record the position information of the event in the sequence, which is the goal of position embedding.

Step 2: Position embedding was first proposed by Vaswani *et al.* in 2017 [29]. It was originally used in NLP and added after the word vector layer to supplement position information. In our project, we compared two embedding methods: Adding vs. Concatenating. We then choose the concatenating method based on the performance of our experiments, *i.e.*, we let $x_i = \langle f_m, f_{sd}, f_l, f_A, f_C, f_G, f_T, f_{pe_1}, \dots, f_{pe_n} \rangle$, to capture the position information, where $f_{pe_j}, j = 1, \dots, n$

are position vector created with the Sinusoidal PE (Eq. (1))

$$x_{(i,2j)} = \sin\left(i * w_0^{\frac{2j}{d_{\text{mod}}}}\right), \quad x_{(i,2j+1)} = \cos\left(i * w_0^{\frac{2j}{d_{\text{mod}}}}\right), \quad (1)$$

where i is the position of the event and $w_0 = 1/10,000$ is the minimum frequency of the embedding. $d_{\text{mod}} = n + 7$, the dimension of the x_i , which is also known as the dimension of the model, must be an even integer and a multiple of the number of attention heads.

This feature-based event vector with size $[w \text{ by } (n + 7)]$ was used as input of the following Transformer encoder module to predict whether the center event of the w events is generated from a modified base.

Step 3: The Transformer encoder module follows overall encoder layer of Transformer architecture using stacked self-attention and point-wise, fully connected layers to capture complicated relationship between signals and prediction target of modifications.

Given a long read with sequential events, an interested event x_i with dimension $(n + 7)$ and its neighborhood $\langle x_{i-\frac{w}{2}}, \dots, x_i, \dots, x_{i+\frac{w}{2}} \rangle$ could be used as input of this Transformer encoder. We then pass these vectors to the “self-attention” layer. Self-attention uses the attention mechanism to calculate the association between each event and all other events to get attention score. Using these attention scores, a weighted representation can be obtained, which is an equivalent list of $(n + 7)$ -dimension vectors, and it is then fed into the feedforward neural network to obtain a new representation, which takes contextual information into account. The output of the feedforward neural network is also a list of $(n + 7)$ -dimension vectors, and then the output is passed up to the next encoder layer. This process can be repeated for events of interest in a long read, and then for all long reads that are available for analysis. Each layer has two sub-layers. The first sub-layer represents a multi-head self-attention mechanism, and the second sub-layer is a simple, position-wise fully connected feedforward network.

Step 4: The output of the Transformer encoder is a list of vectors of floats. The modification for the interested event is predicted by the final fully connected network, ReLU activate function, and softmax layer. Suppose our model trains a total of 10,000 interested events from the training dataset in the experiment. The corresponding output vectors from Transformer encoder was sent to the final fully connected layer to generate a vector with 10,000 dimensions, each representing the score of an interested event. After the linear layer there is a softmax layer that converts these scores into probabilities. We then identify the interested event with probability higher than 0.5 as methylation nucleotides and events with

probability equal to or lower than 0.5 as unmethylation nucleotides, and then use these predictions as the output for the downstream analysis.

Training and testing process

We train the model on the *E. coli* datasets, and then test it on human genomes in order to demonstrate the feasibility of using our method on different species as we did for DeepMod. In detail, given a training set: $D = \{(x_i, y_i) | x_i \in \mathbb{R}^{d_{\text{mod}}}, y_i \in \{0, 1\}\}_{i=1}^{N_{\text{total}}}$, we assume that $y_i = 1$ for methylation bases while $y_i = 0$ for un-methylation bases. For each input event x_i , it outputs a predicted value $H(x_i)$. Then the predicted value will be compared with the actual label y_i , we will get the following four situations:

- the prediction is positive, and the actual is also positive, we call it true positive (TP),
- the prediction is positive, the actual is negative, we call it false positive (FP),
- the prediction is negative and the actual is positive, which is called false negative (FN).
- the prediction is negative and the actual is also negative, which is called true negative (TN).

If we define a test set, the number of completely modified bases is P and the number of completely unmodified bases (or motifs of interest) is N , then Accuracy, Precision, Recall, and AUC are used to evaluate the performance.

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (P + N) \quad (2)$$

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \quad (3)$$

$$\text{Recall} = \text{TP} / P \quad (4)$$

$$F1 = (2 * \text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (5)$$

DATA AND CODE AVAILABILITY

All code for data cleaning and analysis associated with the current submission is available upon request.

ACKNOWLEDGEMENTS

We thank Drs. Qian Liu and Jacqueline Peng for technical assistance and thank Dr. Li Fang for insightful discussions on the model. We thank the authors of methBERT for an initial implementation of BERT models on the problem of methylation detection from signal data sets. We thank the Genome In A Bottle Consortium for providing the NA12878 data sets and ground truth methylation calls for evaluating our method.

COMPLIANCE WITH ETHICS GUIDELINES

The authors Xiuquan Wang, Mian Umair Ahsan, Yunyun Zhou and

Kai Wang declare that they have no conflict of interest or financial conflicts to disclose.

This article does not contain any studies with human or animal materials performed by any of the authors.

OPEN ACCESS

This article is licensed by the CC BY under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

REFERENCES

- Villicaña, S. and Bell, J. T. (2021) Genetic impacts on DNA methylation: research findings and future perspectives. *Genome Biol.*, 22, 127
- Kulis, M. and Esteller, M. (2010) DNA methylation and cancer. *Adv. Genet.*, 70, 27–56
- Jin, B. and Robertson, K. D. (2013) DNA methyltransferases, DNA damage repair, and cancer. *Adv. Exp. Med. Biol.*, 754, 3–29
- Bernstein, C., Nfonsam, V., Prasad, A. R. and Bernstein, H. (2013) Epigenetic field defects in progression to cancer. *World J. Gastrointest. Oncol.*, 5, 43–49
- Martínez-Iglesias, O., Carrera, I., Carril, J. C., Fernández-Novoa, L., Cacabelos, N. and Cacabelos, R. (2020) DNA methylation in neurodegenerative and cerebrovascular disorders. *Int. J. Mol. Sci.*, 21, 2220
- Jeong, H., Mendizabal, I., Berto, S., Chatterjee, P., Layman, T., Usui, N., Toriumi, K., Douglas, C., Singh, D., Huh, I., *et al.* (2021) Evolution of DNA methylation in the human brain. *Nat. Commun.*, 12, 2021
- Jobe, E. M. and Zhao, X. (2017) DNA methylation and adult neurogenesis. *Brain Plast.*, 3, 5–26
- Tognini, P., Napoli, D. and Pizzorusso, T. (2015) Dynamic DNA methylation in the brain: a new epigenetic mark for experience-dependent plasticity. *Front. Cell. Neurosci.*, 9, 331
- McCoy, C. R., Glover, M. E., Flynn, L. T., Simmons, R. K., Cohen, J. L., Ptacek, T., Lefkowitz, E. J., Jackson, N. L., Akil, H., Wu, X., *et al.* (2019) Altered DNA methylation in the developing brains of rats genetically prone to high versus low anxiety. *J. Neurosci.*, 39, 3144–3158
- Jones, P. A., Issa, J. P. and Baylin, S. (2016) Targeting the cancer epigenome for therapy. *Nat. Rev. Genet.*, 17, 630–641
- Mani, S. and Herceg, Z. (2010) DNA demethylating agents and epigenetic therapy of cancer. *Adv. Genet.*, 70, 327–340
- Issa, J. P., Garcia-Manero, G., Giles, F. J., Mannari, R., Thomas, D., Faderl, S., Bayar, E., Lyons, J., Rosenfeld, C. S., Cortes, J., *et al.* (2004) Phase 1 study of low-dose prolonged exposure schedules of the hypomethylating agent 5-aza-2'-deoxycytidine (decitabine) in hematopoietic malignancies. *Blood*, 103, 1635–1640
- Ding, X. L., Yang, X., Liang, G. and Wang, K. (2016) Isoform switching and exon skipping induced by the DNA methylation inhibitor 5-Aza-2'-deoxycytidine. *Sci. Rep.*, 6, 24545
- Ovenden, E. S., McGregor, N. W., Emsley, R. A. and Warnich, L. (2018) DNA methylation and antipsychotic treatment mechanisms in schizophrenia: progress and future directions. *Prog. Neuropsychopharmacol. Biol. Psychiatry*, 81, 38–49
- Clark, T. A., Lu, X., Luong, K., Dai, Q., Boitano, M., Turner, S. W., He, C. and Korch, J. (2013) Enhanced 5-methylcytosine detection in single-molecule, real-time sequencing via Tet1 oxidation. *BMC Biol.*, 11, 4
- Beaulaurier, J., Zhang, X.-S., Zhu, S., Sebra, R., Rosenbluh, C., Deikus, G., Shen, N., Munera, D., Waldor, M. K., Chess, A., *et al.* (2015) Single molecule-level detection and long read-based phasing of epigenetic variations in bacterial methylomes. *Nat. Commun.*, 6, 7438
- Liu, Q., Georgieva, D. C., Egli, D. and Wang, K. (2019) NanoMod: a computational tool to detect DNA modifications using Nanopore long-read sequencing data. *BMC Genomics*, 20, 78
- Simpson, J. T., Workman, R. E., Zuzarte, P. C., David, M., Dursi, L. J. and Timp, W. (2017) Detecting DNA cytosine methylation using Nanopore sequencing. *Nat. Methods*, 14, 407–410
- Pimiento, C., Ehret, D. J., Macfadden, B. J. and Hubbell, G. (2010) Ancient nursery area for the extinct giant shark megalodon from the Miocene of Panama. *PLoS One*, 5, e10552
- Ni, P., Huang, N., Zhang, Z., Wang, D.-P., Liang, F., Miao, Y., Xiao, C.-L., Luo, F. and Wang, J. (2019) DeepSignal: detecting DNA methylation state from Nanopore sequencing reads using deep-learning. *Bioinformatics*, 35, 4586–4595
- Weirather, J. L., de Cesare, M., Wang, Y., Piazza, P., Sebastiano, V., Wang, X.-J., Buck, D. and Au, K. F. (2017) Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. *F1000 Res.*, 6, 100
- Yuen, Z. W.-S., Srivastava, A., Daniel, R., McNevin, D., Jack, C. and Eyra, E. (2021) Systematic benchmarking of tools for CpG methylation detection from Nanopore sequencing. *Nat. Commun.*, 12, 3438
- Liu, Q., Fang, L., Yu, G., Wang, D., Xiao, C.-L. and Wang, K. (2019) Detection of DNA base modifications by deep recurrent neural network on Oxford Nanopore sequencing data. *Nat. Commun.*, 10, 2449
- Liu, Y., Rosikiewicz, W., Pan, Z., Jillette, N., Wang, P., Taghbalout, A., Foox, J., Mason, C., Carroll, M., Cheng, A., *et al.* (2021) DNA methylation-calling tools for Oxford Nanopore sequencing: a survey and human epigenome-wide evaluation. *Genome Biol.*, 22, 295
- Zhang, Y.-Z., Yamaguchi, K., Hatakeyama, S., Furukawa, Y.,

- Miyano, S., Yamaguchi, R. and Imoto, S. (2021) On the application of bert models for Nanopore methylation detection. In: 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 320–327
26. Jiao, L., Zhang, F., Liu, F., Yang, S., Li, L., Feng, Z. and Qu, R. (2019) A survey of deep learning-based object detection. *IEEE Access*, 7, 128837–128868
27. Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E. and Gouil, Q. (2020) Opportunities and challenges in long-read sequencing data analysis. *Genome Biol.*, 21, 30
28. Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K. (2018) Bert: pre-training of deep bidirectional transformers for language understanding. *arXiv*, 1810.04805
29. Vaswani, A., Shazeer, N. M., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. and Polosukhin, I. (2017) Attention is all you need. *arXiv*, 1706.03762