Prediction of Ion Channels with Pre-trained BERT Language Models

Anuj Singh[1,*], Kaushlendra Yadav[1], and Arvind Kumar Tiwari[1]

ARTICLE INFO

Article history:
Received: 06 April 2023
Revised: 30 June 2023
Accepted: 21 July 2023

Keywords:
Ion channels
RoBERTa
Voltage-gated ion channels
1GRAM
BERT

ABSTRACT

The area of natural language processing has now paid close interest to language representation models due to their impressive outcomes. Ion channels are important for physiological processes, like the transmission of nerve impulses, muscle contraction, and the regulation of hormone secretion. Disruptions in ion channel working can lead to a variety of disorders. The study of ion channels has directed significant progress in cellular physiology and the creation of novel therapies for a variety of diseases. Bidirectional Encoder Representations from Transformers (BERT) has recognized itself as a new and more effective language model that is simple and effective. To grab the semantics and context of words, BERT accepted the idea of contextualized word embedding. Our proposed approach showed 95% accuracy, 96% precision, 99% recall, 96% AUC, 98% F1, and 14% loss for the prediction of Voltage-gated ion channels (VGICs), 100% accuracy, 100% precision, 100% recall, 100% AUC, 100% F1 and 00% loss for the prediction of Ligand-gated ion channels (LGICs) and 95% accuracy, 96% precision, 98% recall, 95% AUC, 97% F1 and 13% loss for overall. Comparative analysis shows that the transformer-based method performed better than the other approaches.

1. INTRODUCTION

The Transformer is a framework for natural language processing (NLP) that was introduced in 2017. It is a neural network that is capable of modeling sequential data, such as text, and is used for a large number of NLP tasks, like language translation and text summarization. It used the key concept of self-attention, the Transformer architecture enables the model to make predictions by focusing on various segments of the input sequence. Encoders as well as a decoder are primary elements of the transformer-based model. The encoder accepts input from a sequence of tokens, such as words or sub-words, and processes them in parallel through a series of self-attention layers. Every self-attention layer enables the model to concentrate on various segments of the input sequence, learning which tokens are most useful to the present task [1]. Each self-attention layer is composed of three sub-layers: Multi-Head Attention: This sub-layer enables the model to undertake multiple parts of the input sequence parallelly. It does this by computing multiple attention scores in parallel, each with a different set of learnable weights. These attention scores are then combined to give a final set of context vectors that are used as input to the next sub-layer. Layer Normalization: This sub-layer normalizes the context vectors output by the multi-head attention layer Feed-Forward Network: This applies fully connected neural networks to the context vectors, which enables the transformer model to learn difficult relationships between the input tokens. The decoder takes as input a sequence of tokens, just like the encoder, and uses self-attention and encoder-decoder attention to generate an output sequence. The encoder-decoder attention allows the decoder to attend to various parts of the input sequence when generating every token in the output sequence. It consists of three sub-layers: Masked Multi-Head Attention: This enables the decoder to attend to various parts of the input sequence, but only to tokens that come before the current token being generated. This is done to prevent the model from cheating by using future information when generating the output sequence. Multi-Head Attention: This is similar to the multi-head attention layer in the encoder, but it attends to the output of the encoder instead of the input sequence. Feed-Forward Network: This is the same as in the encoder and used a fully connected neural network to process the context vectors. The encoder and decoder are trained together using a variant of the seq2seq learning framework; here input sequence is mapped to an output sequence. At training time, the model is improved to reduce a loss function that measures the difference among the predicted as well as target output sequences. In summary, the Transformer consists of an encoder as well as a decoder that are trained together using a seq2seq learning framework. The self-attention has been shown to outperform traditional

[1]*KNIT Sultanpur, Uttar Pradesh, India.*
**Corresponding author:* Anuj Singh; E-mail: anuj.2295@knit.ac.in.

recurrent neural networks RNNs) and convolutional neural networks (CNNs) in a large area of NLP processes [2]-[4]. In the past, homology-based approaches were used to predict the ion channels, but they failed when a new ion channel was different from the previous one. Therefore, to remove the problems associated with homology-based traditional approaches it is necessary to design efficient machine learning-based approaches for the prediction of neurotransmitter and its related disease. Here the problem for the prediction of neurotransmitters may be considered a pattern classification problem. The proposed work estimates the VGIC class and LGIC class. Section 1 explains the transformer framework. Section 2 first describes ion channels and then illustrates VGICs and LGISs. Section 3 presents the literature review based on various neurotransmitter prediction techniques. Section 4 presents the material and methods. Section 5 illustrates the result and discussion. Section 6 presents a comparative analysis and Section 6 conclude the work and provide brief detail about future work.

2. ION CHANNELS

Ion channels are specialized proteins found in cell membranes that enable ions to pass through the membrane. Ions are electrically charged particles, like sodium, chloride, potassium, and calcium that are essential for large physiological processes. They are typically gated and enable ions to pass through in response to specific signals. VGICs are those that open in reaction to alterations in the voltage across the cell membrane. One more type is LGIC which activates in response to the binding of a particular molecule, such as a neurotransmitter [5]. Ion channels are important for the transmission of brain impulses, muscle contraction, and the regulation of hormone secretion. Disruptions in its function can cause several diseases, like epilepsy, cardiac arrhythmias, and cystic fibrosis [6]. There are many classes of ion channels, and every ion channel has its unique properties. Fast depolarization of nerve cells during an action potential depends on voltage-gated sodium ion channels, whereas LGICs like nicotinic acetylcholine receptors are important for transmitting signals between nerve cells and muscle cells. The study of ion channels has led to significant improvement in our knowledge of cellular physiology as well as the creation of new therapies for a large number of illnesses [7]. Ion channels operate at multiple scales, from the atomic level to cellular and tissue levels. Integrating different modeling approaches, such as molecular dynamics, electrophysiological modeling, and systems biology, can provide a comprehensive understanding of ion channel function in complex biological systems. Multi-scale models can help predict how ion channel activity influences cellular excitability, organ function, and overall physiological responses [8].

2.1 Voltage-gated Ion Channels

VGICs are important for muscle contraction, neurotransmitter release, and the propagation of electrical signals in the brain. When VGICs are closed, it is said to be resting. As the membrane potential changes, the channel can transition to an open state, allowing ions to pass through. The specific ion that can pass through the channel is based on the type of channel, with some channels allowing only positively charged ions, such as sodium or calcium, to pass through, while others allow only negatively charged ions, like chloride to pass through [9]. The opening, as well as the closing of VGICs, is controlled by variations in the membrane potential. When potential becomes positive, the channel can transition from a closed state to an open state, allowing ions to flow through. Conversely, when potential becomes negative, the channel can transition from an open to a closed state, preventing ions from passing through. The gating process of voltage-gated ion channels is complex and can involve multiple steps. The activity of VGICs is critical for the proper functioning of many different types of cells in the body. Mutations in these channels can lead to many diseases, like epilepsy, cardiac arrhythmias, as well as muscle diseases. As a result, VGICs are a key target for drug development, with many drugs developed to modulate the activity of these channels in specific ways [10, 11].

2.2 Ligand-gated ion channels

Transmembrane proteins called LGICs allow ions to pass through cell membranes in response to the binding of specific molecules called ligands. These channels are essential for the communication between cells, and their dysfunction can lead to a variety of diseases. The structure of LGICs typically consists of five subunits, each with four transmembrane domains that form a central pore. The subunits come together to form a hollow, barrel-shaped structure. The extracellular region contains binding sites for specific ligands, such as neurotransmitters or hormones. The selectivity is determined by amino acid residues lining the pore. Once the ligand dissociates from the channel, the pore closes and the channel returns to its original conformation. LGICs are found in different types of tissues throughout the body, including the nervous system, muscles, and immune system. They play important roles in synaptic transmission, muscle contraction, and the immune response. Dysfunctions in these channels can lead to diseases such as epilepsy, myasthenia gravis, and autoimmune disorders. Some well-known examples of LGICs are the nicotinic acetylcholine receptor, the GABA, and the NMDA receptor. The development of drugs that target these channels has led to the treatment of a variety of diseases, including anxiety, depression, and schizophrenia [12, 13].

3. RELATED WORKS

In the paper [14], the author classified classes of glucose transporters with bidirectional encoder representations from transformers language models. They converted protein sequences into finite-length relevant vectors by treating them as sentences. They found that the output of the BERT-Base and BERT Large models was enhanced by more than 4% in contexts of MCC and sensitivity. In the paper [15], the author investigated correlations between protein sequences and natural language by evaluating contextualized word embedding from BERT models. To estimate FAD-binding sites from the transport proteins that have recently been discovered in nature, they introduced a novel method based on pre-training of BERT, PSSM, and AAIndex. They achieved an accuracy of 85.14%. In the paper [16], the authors proposed techniques that successfully classify TRP channels from non-TRP channels. This method uses contextualized word embeddings from BERT and SVM classifiers to examine protein sequences. This method achieved a specificity of 96.03%, a sensitivity of 80%, an accuracy of 95.47%, as well as an MCC of 0.56.

In the paper [17], the authors used deep learning and the BERT method to identify Drug-Target Interactions. This method achieved an accuracy of 94.7%, 90.1%, 89%, and 94.9% for ion channels, G Protein-coupled receptors, nuclear receptors, and enzymes respectively. In the paper [18], the authors used per-trained transformers for the classification of glutarylation sites. This method achieved overall specificity of 0.6286, recall of 0.7864, and AUC score of 0.7075. In the paper [19], the author developed a machine learning-centered classifier to classify the class of bacterial sortases from sequence-derived data. Sortase enzymes seem to be cysteine transpeptidases that add different proteins to the surface of Gram-positive bacteria. There are six known sortases classes. This method achieved overall specificity of .97, an accuracy of .95, and an MCC of .80 for the independent test with SMOTE. In the paper [20], the authors developed bidirectional encoder representations from a transformer-based method to recognize RNA N7-Methylguanosine sites. This method achieved an accuracy of 95.48% as well as an MCC of 0.91. In the paper [21], the author developed BERT with SVM based approach using the UniProt dataset and achieved an accuracy of 87.13% and 94.15% for the transport and membrane datasets respectively. The authors of [22] introduced S-Pred, a new tool that predicts SS8, ASAs, and IDRs from a sequence. S-Pred uses the MSA of a query sequence as input for feature prediction.

The MSA Transformer is an attention-based protein language model that transforms the MSA input into features. To create the final prediction, an LSTM was used. Several test sets were used to evaluate S-Pred's performance, and each time, the program produced reliable predictions. The SS8 prediction's accuracy was around 76%, and there was a 0.84 Pearson's correlation between the experimental and predicted ASAs. The authors of [23] proposed A-Prot, a newly created MSA Transformer-based protein 3D structure modeling technique. They showed that A-Prot predicts long-distance contacts more accurately than the current approaches. They also created 3D models of the CASP14 targets for free modeling as well as hard template-based modeling.

Table 1: A list of current related works

Approach/Model	Dataset	Accuracy	Ref
BERT-Large (Uncased)	GLUT	92.43%	[14]
BERT-Based (Uncased)	SGLT	97.30%	
BERT-Large (Uncased)	SWEET	92.97%	
FAD-BERT	UniProt	85.14%	[15]
TRP-BERT	UniProt and Transporters Classification Database (TCDB)	95.47%	[16]
DTI-BERT	DrugBanks, KEGG BRITE, and BRENDA, SuperTarget Database	94.7%, 90.1%, 89%, and 94.9% for ion channels, G Protein-coupled receptors, nuclear receptors, and enzymes respectively	[17]
ProtTrans-Glutar	PLMD	Specificity of 62.86%, recall of 78.64%, and AUC score of 0.7075	[18]
SortPred	NCBI and CDD	95%	[19]
BERT-m7G	Dai et al. [8]	95.48%	[20]
BERT with SVM	UniProt	87.13% for the transport dataset and 94.15% for the membrane dataset	[21]
S-Pred	CASP13	76%	[22]

In [24], the authors presented the PA_SPP approach, which predicts the structure profile with no alignment for an individual protein sequence and depends on a pre-trained ProtAlbert transformer. Transformers perform admirably when dealing with natural languages. We can take advantage of these models by considering protein sequences as a language. The Transformer language model, a cutting-

edge language model, was used by the author in [25] to carry out contextual embedding for phage contigs. They may input the protein composition as well as the positions of the proteins in every contig into the Transformer by creating a protein-cluster vocabulary. Employing the self-attention mechanism, the Transformer can learn protein organization as well as associations and forecast the label for test contigs. In paper [26], the authors employed SVM, Logistic Regression, Decision Tree, Random Forest, Adaboost, Naive Bayes, and KNN on 12285 protein data and classified into 27 classes. Some other important papers related to this work are [27, 28] (See Table 1).

4. MATERIAL AND METHODS

4.1 Data Set Description

UniProt (http://www.uniprot.org), and the NCBI (http://www.ncbi.nlm.nih.gov/protein) were used to obtain the ion channel sequences. Non-ion channel sequences were obtained from the UniProt database. The CD-HIT server is used to remove protein sequences with greater than 40% identity. The protein sequence is converted to 1-gram.

4.2 N-gram

In NLP, a contiguous run of n tokens from a given text represents an n-gram. The token can be letters, words, or some other units depending on the application. The concept of n-grams is widely employed in many NLP works, like text classification as well as machine translation. Here's an example to illustrate the concept of n-grams [29]. Suppose we have the sentence: "Mary had a little lamb, its fleece was white as snow." If we want to create 2-grams (also called bigrams) from this sentence, we would take every two adjacent words and create a sequence of pairs, like this: "Mary had", "had a", "a little", "little lamb", "lamb its", "its fleece", "fleece was", "was white", "white as", "as snow". In general, we can create n-grams for any positive integer value of n. The value of n determines the size of the sliding window that we use to create the n-grams. The larger the value of n, the more context we capture in each n-gram, but also the fewer n-grams we get from the same sample of text. N-grams are commonly used in language modeling. N-grams can also be used in text classification and machine translation, where they can capture the local context and dependencies between adjacent words or units [30, 31]. In this paper we used 1-gram.

4.3 Methodology

This paper used RoBERTa-Large, BERT-Base (Uncased), BERT-Large (Uncased), BERT-Base (Cased), BERT-Large (Cased), RoBERTa-Base for the prediction of VGICs and LGICs.

4.3.1 BERT Model

BERT is a type of pre-trained neural network model for NLP tasks. BERT used transformer architecture to process sequences of data, like sentences. The transformer layout is centered on the concept of self-attention, which permits the model to determine the importance of a very large input sequence when making predictions. BERT undergoes training on a significant text corpus using a technique called masked language modeling. In this method, some of the words in a sentence are independently masked out, and the model is trained to foresee the missing words according to the meaning of the sentence. Additionally, BERT is trained to determine whether two sentences are contiguous or not by performing a task known as a next-sentence prediction. Fine-tuning can greatly enhance the capability of BERT on a particular work, as the model has already learned a great deal about the structure and meaning of natural language. Overall, BERT is an important model for a large number of NLP tasks [32]. BERT was only pre-trained on the raw texts, using a fully automated procedure to create inputs without any assistance from humans [33]. BERT was initially made available in base and large variations for both case-sensitive and case-insensitive input text (see Table 2).

Table 2: BERT Model with Parameters

S. No.	Model	Size	Parameters
1	BERT	Base (Uncased)	110M
2	BERT	Base (Cased)	110M
3	BERT	Large (Uncased)	340M
4	BERT	Large (Cased)	340M

4.3.1.1 BERT-Base (Uncased)

A trained learning model in English that uses an MLM objective is called BERT-Base (Uncased). "Base" refers to the smaller version with 12 layers and 110 million parameters, while "uncased" means that the model is trained on lowercase text without distinguishing between uppercase and lowercase letters. This BERT model does not distinguish between English and english [34].

4.3.1.2 BERT-Base (Cased)

BERT base cased is a pre-trained language model generated by Google that employs a transformer to understand the meaning of the text. "Base" refers to the smaller version with 12 layers and 110 million parameters, while "cased" means that the model is trained on the text that preserves the original case of letters, such as "Anuj" and "anuj" being treated as different tokens. The BERT-based cased is trained on a huge collection of text, including Wikipedia and the BookCorpus dataset. Compared to the uncased model, the cased model may be more suitable for tasks that require the distinction of capitalization, such as entity recognition in the biomedical or legal text [35].

4.3.1.3 BERT-Large (Uncased)

BERT Large uncased is a pre-trained language model generated by Google that employs a transformer to understand the meaning of the text. "Large" refers to the larger version with 24 layers and 340 million parameters, and that makes it more powerful than the base model. "Uncased" means that the model is trained on lowercase text without distinguishing between uppercase and lowercase letters. The BERT Large uncased is trained on a huge collection of text, including Wikipedia and the BookCorpus dataset. Compared to the base model, it may be more suitable for tasks that require more complex reasoning and a deeper understanding of the text. However, it also requires more computational resources and time to train and use [36].

4.3.1.4 BERT-Large (Cased)

BERT Large-cased is a pre-trained language model generated by Google that employs a transformer to understand the meaning of the text. "Large" refers to the larger version with 24 layers and 340 million parameters, which makes it more powerful than the base model. "Cased" means that the model is trained on the text that preserves the original case of letters, such as "Anuj" and "anuj" being treated as different tokens. The BERT Large cased is trained on a huge collection of text, including Wikipedia and the BookCorpus dataset. Compared to the uncased model, the cased model may be more suitable for tasks that require the distinction of capitalization, such as entity recognition in biomedical or legal text. The large model may be more suitable for tasks that require more complex reasoning and a deeper understanding of the text. However, it also requires more computational resources and time to train and use [37].

4.3.2 RoBERTa Model

RoBERTa (Robustly Optimized BERT approach) is a pre-trained language model developed by Facebook AI Research (FAIR) that builds upon the BERT architecture with several modifications and training techniques to improve its performance on natural language processing (NLP) tasks. Some key features of RoBERTa include Dynamic Masking: RoBERTa applies a different masking pattern than BERT during pre-training, which helps prevent the model from memorizing the positions of masked tokens. Larger Batch Size: RoBERTa uses a larger batch size during training, which allows the model to see more examples per iteration and improve its understanding of language. More Data and Longer Training: RoBERTa is trained on a huge collection of text, including Wikipedia and BookCorpus, and for longer periods than BERT, which helps improve the model's generalization ability. Byte Pair Encoding (BPE): RoBERTa uses BPE to segment words into smaller subwords, which helps the model better handle out-of-vocabulary words and reduces the vocabulary size. RoBERTa has achieved the best accuracy on a large number of NLP tasks and has also been used as a base model for fine-tuning specific tasks. RoBERTa is available in several different sizes, including RoBERTa-base (125 million parameters) as well as RoBERTa-large (355 million parameters), which can be fine-tuned on specific downstream tasks or used to extract contextualized word representations for other NLP applications [38].

4.3.2.1 RoBERTa-Base Model

RoBERTa-base is a variant of the RoBERTa (Robustly Optimized BERT approach) model. It has 125 million parameters, which is the same as the BERT base model, but it includes several modifications and training techniques to improve its performance on natural language processing (NLP) tasks. RoBERTa-base uses a transformer-based architecture, similar to BERT, and includes several modifications such as dynamic masking, which applies a different masking pattern during pre-training than BERT, allowing the model to prevent memorization of masked tokens. RoBERTa-base also uses byte pair encoding (BPE) to segment words into smaller subwords, which helps the model handle out-of-vocabulary words and reduces the

vocabulary size. RoBERTa-base is trained on a huge collection of text, including Wikipedia and BookCorpus, for longer periods than BERT, which helps improve the model's generalization ability. Additionally, RoBERTa-base uses a larger batch size during training, which allows the model to see more examples per iteration and improve its understanding of language. It can be fine-tuned on specific downstream tasks or used to extract contextualized word representations for other NLP applications [39].

4.3.2.2 RoBERTa-Large Model

RoBERTa Large uses a transformer-based architecture and includes several modifications and training techniques to improve its performance on natural language processing (NLP) tasks. These modifications include dynamic masking, which applies a different masking pattern during pre-training than BERT, and byte pair encoding (BPE) to segment words into smaller subwords, which helps the model, handle out-of-vocabulary words and reduces the vocabulary size. RoBERTa Large is trained on a huge collection of text, including Wikipedia and BookCorpus, for longer periods than the RoBERTa-base model, which helps improve the model's generalization ability. Additionally, RoBERTa Large uses a larger batch size during training, which allows the model to see more examples per iteration and improve its understanding of language. It can be fine-tuned on specific downstream tasks or used to extract contextualized word representations for other NLP applications. However, due to its large size, RoBERTa Large requires significant computational resources to train and use [40]. The proposed method uses UniProt and NCBI Dataset. The CD-HIT server is used to remove protein sequences with greater than 40% identity. Next, the 1-Gram is applied to the dataset. Then we applied a transformer-based model for the prediction of VGICs and LGICs (See Figure 1).

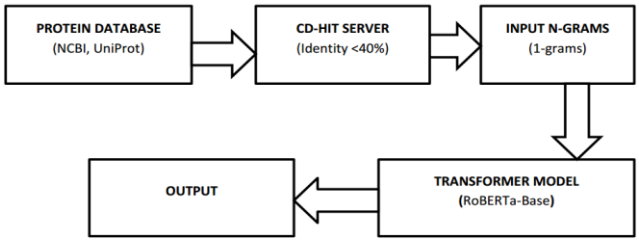

Fig. 1. Flowchart of the Pre-trained BERT Language Models based approach.

Table 3: The results of predicting LGICs and VGICs

Family	Model	Loss	Acc.	Precision	recall	AUC	F1
Overall	BERT-Base (Uncased)	17	92	92	100	88	96
	BERT-Base (Cased)	16	92	92	100	88	96
	BERT-Large (Uncased)	17	92	92	100	85	96
	BERT-Large (Cased)	17	92	92	100	90	96
	RoBERTa-Base	13	95	96	98	95	97
	RoBERTa-Large	18	92	92	100	90	96
VGICs	BERT-Base (Uncased)	24	90	90	100	81	95
	BERT-Base (Cased)	24	90	90	100	81	95
	BERT-Large (Uncased)	23	90	90	100	81	95
	BERT-Large (Cased)	23	90	90	100	80	95
	RoBERTa-Base	14	95	96	99	96	98
	RoBERTa-Large	23	90	90	100	88	95
LGICs	BERT-Base (Uncased)	00	100	100	100	100	100
	BERT-Base (Cased)	00	100	100	100	100	100
	BERT-Large (Uncased)	00	100	100	100	100	100
	BERT-Large (Cased)	00	100	100	100	100	100
	RoBERTa-Base	00	100	100	100	100	100
	RoBERTa-Large	00	100	100	100	100	100

5. RESULTS AND DISCUSSION

To predict VGICs and LGICs we used RoBERTa-Large, BERT-Base (Uncased), BERT-Large (Uncased), BERT-Base (Cased), BERT-Large (Cased), and RoBERTa-Base.

To predict VGICs and LGICs, the accuracy, loss, precision, recall, AUC, and F1- Score of the different classifiers are calculated. The BERT Base (Uncased) gives accuracy of 90%, 100%, 92%, with loss of 24%, 00%, 17%, precision of 90%, 100%,92%, recall of 100%,100% 100%, AUC of 81%, 100%, 88%, and F1-Score of 95%, 100%, 96% for the prediction of VGIC, LGIC and overall respectively. The BERT Base (Cased) gives accuracy of 90%, 100%, 92%, with loss of 24%, 00%, 16%, precision of 90%, 100%,92%, recall of 100%,100% 100%, AUC of 81%, 100%, 88%, and F1-Score of 95%, 100%, 96% for the prediction of VGIC, LGIC and overall respectively. The BERT Large (Uncased) gives accuracy of 90%, 100%, 92%, with loss of 23%, 00%, 17%, precision of 90%, 100%,92%, recall of 100%,100% 100%, AUC of 81%, 100%, 85%, and F1-Score of 95%, 100%, 96% for the prediction of VGIC, LGIC and overall respectively. The BERT Large (Cased) gives accuracy of 90%, 100%, 92%, with loss of 23%, 00%, 17%, precision of 90%, 100%,92%, recall of 100%,100% 100%, AUC of 80%, 100%, 90%, and F1-Score of 95%, 100%, 96% for the prediction of VGIC, LGIC and overall respectively. The RoBERTa Large gives accuracy of 90%, 100%, 92%, with loss of 23%, 00%, 18%, precision of 90%, 100%,92%, recall of 100%,100% 100%, AUC of 88%, 100%, 90%, and F1-Score of 95%, 100%, 96% for the prediction of VGIC, LGIC and overall respectively. Our proposed approach (RoBERTa-Base) showed 95% accuracy, 96% precision, 99% recall, 96% AUC, 98% F1, and 14% loss for the prediction of VGICs, 100% accuracy, 100% precision, 100% recall, 100% AUC, 100% F1 and 00% loss for the prediction of LGICs and 95% accuracy, 96% precision, 98% recall, 95% AUC, 97% F1 and 13% loss for overall.

The analysis of Table 3 shows that the RoBERTa BASE model gives better results in comparison with other approaches (see Table 3).

6. COMPARATIVE ANALYSIS

In the paper [41], the author proposed a Random Forest with an MRMD-based method using 188- dimensional feature and obtained accuracy values for VGICs, LGICs, and overall of 93.9%, 86.0%, and 89.93%, respectively. In [14], the author classified classes of glucose transporters with bidirectional encoder representations from transformer language models. They converted protein sequences into finite-length relevant vectors by treating them as sentences. They obtained accuracy values for GLUT, SGLT, and SWEET of 92.43%, 97.30%, and 92.97% respectively. For VGICs, LGICs, and overall, the suggested method of the RoBERTa BASE achieved an accuracy of 95%, 100%, and 95% respectively (See Table 4).

7. CONCLUSION

Ion channels are the cause of many neurological disorders. So the accurate prediction of ion channels is very important. If the ion channels will be predicted correctly, then tasks become easy for the drug analyst to discover the new drugs. The study of ion channels has directed significant progress in cellular physiology and the creation of novel therapies for a variety of diseases. BERT has recognized itself as a new and more effective language model that is simple and effective. Our proposed approach showed 95% accuracy, 96% precision, 99% recall, 96% AUC, 98% F1, and 14% loss for the prediction of VGICs, 100% accuracy, 100% precision, 100% recall, 100% AUC, 100% F1 and 00% loss for the prediction of LGICs and 95% accuracy, 96% precision, 98% recall, 95% AUC, 97% F1 and 13% loss for overall. Comparative analysis shows that the RoBERTa base-based method performed better than the other approaches. The future of ion channel prediction lies in integrating diverse datasets, including experimental measurements, structural information, genomic data, and clinical observations. Collaborations between experimentalists, computational biologists, and clinicians are crucial for building comprehensive models and validating predictions in real-world scenarios. Sharing data and developing standardized formats for data exchange will facilitate progress in ion channel prediction research.

Table 4: Comparative Analysis for the prediction of ion channel classes between the present and proposed method

Reference	Method	Class	Performance (%)
[41]	Random Forest with MRMD (188-dimensional feature)	VGICs	93.9
		LGICs	86.0
		overall	89.93
[14]	BERT-Large (Uncased)	GLUT	92.43
	BERT-Base (Uncased)	SGLT	97.30
	BERT-Large (Uncased)	SWEET	92.97
Proposed Method	RoBERTa BASE	VGICs	95
		LGICs	100
		overall	95

REFERENCE

[1] Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A. & Rush, A. M. 2020. Transformers: State-of-the-art natural language processing. Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations, pp. 38-45.

[2] Braşoveanu, A. M., & Andonie, R. 2020. Visualizing transformers for nlp: a brief survey. Proceedings of the 24th International Conference Information Visualisation (IV) pp.

270-279. IEEE.

[3] Gillioz, A., Casas, J., Mugellini, E., & Abou Khaled, O. 2020. Overview of the Transformer-based Models for NLP Tasks. Proceedings of the 15th Conference on Computer Science and Information Systems (FedCSIS), September. pp. 179-183. IEEE.

[4] Han, K., Xiao, A., Wu, E., Guo, J., Xu, C., & Wang, Y. 2021. Transformer in transformer. Advances in Neural Information Processing Systems, 34, 15908-15919.

[5] Tombola, F., Pathak, M. M., & Isacoff, E. Y. 2006. How does voltage open an ion channel? Annu. Rev. Cell Dev. Biol., 22, 23-52.

[6] Kefauver, J. M., Ward, A. B., & Patapoutian, A. 2020. Discoveries in structure and physiology of mechanically activated ion channels. Nature, 587(7835): 567-576.

[7] Huang, Y., Fliegert, R., Guse, A. H., Lü, W., & Du, J. 2020. A structural overview of the ion channels of the TRPM family. Cell Calcium, 85, 102111.

[8] Zheng, S. P., Huang, L. B., Sun, Z., & Barboiu, M. 2021. Self-assembled artificial ion channels toward the natural selection of functions. Angewandte Chemie International Edition, 60(2): 566-597.

[9] Catterall, W. A. 1995. Structure and function of voltage-gated ion channels. Annual review of biochemistry, 64(1): 493-531.

[10] Lehmann-Horn, F., & Jurkat-Rott, K. 1999. Voltage-gated ion channels and hereditary disease. Physiological reviews, 79(4): 1317-1372.

[11] Frank, H. Y., Yarov-Yarovoy, V., Gutman, G. A., & Catterall, W. A. 2005. Overview of molecular relationships in the voltage-gated ion channel superfamily. Pharmacological reviews, 57(4): 387-395.

[12] Hucho, F., & Weise, C. 2001. Ligand-gated ion channels. Angewandte Chemie International Edition, 40(17): 3100-3116.

[13] Keramidas, A., Moorhouse, A. J., Schofield, P. R., & Barry, P. H. 2004. Ligand-gated ion channels: mechanisms underlying ion selectivity. Progress in biophysics and molecular biology, 86(2), 161-204.

[14] Shah, S. M. A., Taju, S. W., Ho, Q. T., & Ou, Y. Y. 2021. GT-Finder: Classify the family of glucose transporters with pre-trained BERT language models. *Computers in biology and medicine*, *131*, 104259.

[15] Ho, Q. T., Le, N. Q. K., & Ou, Y. Y. 2021. FAD-BERT: improved prediction of FAD binding sites using pre-training of deep bidirectional transformers. *Computers in Biology and Medicine*, *131*, 104258.

[16] Shah, S. M. A., & Ou, Y. Y. 2021. TRP-BERT: Discrimination of transient receptor potential (TRP) channels using contextual representations from deep bidirectional transformer based on BERT. *Computers in Biology and Medicine*, *137*, 104821.

[17] Jie, Z., Qiu, W. R., & Xiao, X. 2022. DTI-BERT: Identifying drug-target interactions in cellular networking based on BERT and deep learning method. *Frontiers in Genetics*, 1189.

[18] Indriani, F., Mahmudah, K. R., Purnama, B., & Satou, K. 2022. ProtTrans-Glutar: Incorporating Features from Pre-Trained Transformer-Based Models for Predicting Glutarylation Sites. *Frontiers in Genetics*, 1201.

[19] Malik, A., Subramaniyam, S., Kim, C. B., & Manavalan, B. 2022. SortPred: The first machine learning-based predictor to identify bacterial sortases and their classes using sequence-derived information. *Computational and structural biotechnology journal*, *20*, 165-174.

[20] Zhang, L., Qin, X., Liu, M., Liu, G., & Ren, Y. 2021. BERT-m7G: A Transformer Architecture Based on BERT and Stacking Ensemble to Identify RNA N7-Methylguanosine Sites from Sequence Information. *Computational and mathematical methods in medicine*, *2021*.

[21] Dai, C., Feng, P., Cui, L., Su, R., Chen, W., & Wei, L. 2021. Iterative feature representation algorithm to improve the predictive performance of N7-methylguanosine sites. *Briefings in Bioinformatics*, *22*(4),: bbaa278.

[22] Hong, Y., Song, J., Ko, J., Lee, J., & Shin, W. H. 2022. S-Pred: protein structural property prediction using MSA transformer. *Scientific Reports*, *12*(1): 13891.

[23] Hong, Y., Lee, J., & Ko, J. 2022. A-Prot: protein structure modeling using MSA transformer. *BMC bioinformatics*, *23*(1): 1-11.

[24] Behjati, A., Zare-Mirakabad, F., Arab, S. S., & Nowzari-Dalini, A. 2022. Protein sequence profile prediction using ProtAlbert transformer. *Computational Biology and Chemistry*, *99*, 107717.

[25] Shang, J., Tang, X., Guo, R., & Sun, Y. 2022. Accurate identification of bacteriophages from metagenomic data using Transformer. *Briefings in Bioinformatics*, *23*(4): bbac258.

[26] Bhardwaj, S. K., Vishwakarma, S., Bihari, A., Tripathi, S., Agrawal, S., & Joshi, P. 2024. Protein Enzyme Sequence Class Prediction using Computational Model.GMSARN International Journal 18, 62-70.

[27] Verma, A., Singh, A., Bihari, A., Tripathi, S., Agrawal, S., Pandey, S. K., & Verma, S. 2023. Identification of Hate Speech on Social Media using LSTM. GMSARN International Journal 17, 468-474.

[28] Singh, A., & Tiwari, A. K. 2023. Machine learning-based approach for prediction of ion channels and their subclasses. Journal of Cellular Biochemistry, 124(1): 72-88.

[29] Cavnar, W. B., & Trenkle, J. M. 1994. N-gram-based text categorization. Proceedings of SDAIR-94, 3rd annual symposium on document analysis and information retrieval, April. Vol. 161175.

[30] Kondrak, G. 2005. N-gram similarity and distance. Proceedings of String Processing and Information Retrieval. Proceedings of 12th International Conference, SPIRE 2005, Buenos Aires, Argentina, November 2-4, 115-126. Springer Berlin Heidelberg.

[31] Marino, J. B., Banchs, R. E., Crego, J. M., de Gispert, A., Lambert, P., Fonollosa, J. A., & Costa-jussà, M. R. 2006. N-gram-based machine translation. Computational linguistics, 32(4): 527-549.

[32] Sun, F., Liu, J., Wu, J., Pei, C., Lin, X., Ou, W., & Jiang, P. 2019. BERT4Rec: Sequential recommendation with bidirectional encoder representations from the transformer. Proceedings of the 28th ACM international conference on information and knowledge management, November,1441-1450.

[33] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

[34] Geetha, M. P., & Renuka, D. K. 2021. Improving the

performance of aspect based sentiment analysis using fine-tuned Bert Base Uncased model. International Journal of Intelligent Networks, 2, 64-69.

[35] He, J., Zhao, L., Yang, H., Zhang, M., & Li, W. 2019. HSI-BERT: Hyperspectral image classification using the bidirectional encoder representation from transformers. IEEE Transactions on Geoscience and Remote Sensing, 58(1): 165-178.

[36] Syed, M. S. S., Syed, Z. S., Lech, M., & Pirogova, E. 2020. Automated Screening for Alzheimer's Dementia Through Spontaneous Speech. Proceedings of Interspeech, 2020, pp. 2222-6.

[37] Tenney, I., Das, D., & Pavlick, E. 2019. BERT rediscovers the classical NLP pipeline. arXiv preprint arXiv:1905.05950.

[38] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D. & Stoyanov, V. 2019. Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692.

[39] Taju, S. W., Shah, S. M. A., & Ou, Y. Y. 2021. Identification of efflux proteins based on contextual representations with deep bidirectional transformer encoders. *Analytical Biochemistry*, *633*, 114416.

[40] Zhu, C., Cheng, Y., Gan, Z., Sun, S., Goldstein, T., & Liu, J. 2019. Freelb: Enhanced adversarial training for natural language understanding. arXiv preprint arXiv:1909.11764.

[41] Han, K., Wang, M., Zhang, L., Wang, Y., Guo, M., Zhao, M., & Wang, C. 2019. Predicting ion channels genes and their types with machine learning techniques. Frontiers in genetics, 10, 399.