Advances in Artificial Intelligence and Machine Learning; Research 5 (1) 3495-3518 Received 23-12-2024; Accepted 21-03-2025; Published 28-03-2025

Quantum LLM Model for Entity and Semantic Relation Extraction in Drug Interactions

A.Sankaran sankaran.a@ptuniv.edu.in
Dept. of Computer Science and Engineering
Puducherry Technological University
India

K.Sathiyamurthy Sathiyamurthyk@ptuniv.edu.in
Dept. of Computer Science and Engineering
Puducherry Technological University
India

Corresponding Author: A.Sankaran

Abstract

In modern natural language processing, it is still difficult to extract entity and semantic links from biomedical literature, such as drug-drug, drug-gene, drug-test, drug-disease, drug-herb, drug-food and drug-lab range interactions. In this work XLNet, a large language model based on transformers, is finetuned with Bayesian network that have been improved by the Quantum Approximate Optimization Algorithm (QAOA) by using directed acyclic graphs (DAGs) and Conditional Probability Tables (CPTs) to model complicated biomedical interactions. This work combines the ability of XLNet to capture two-way context with the ability of Bayesian networks to use probabilistic reasoning. QAOA improves computing performance by allowing scalable inference on big datasets. Ranked on a benchmark biomedical dataset, our strategy exceeded current techniques with a 94% accuracy in relationship extraction. With consistency, practical data from unstructured texts, this development improves the accuracy and interpretability of scientific findings, hence enabling drug discovery and personalized treatment.

Keywords: Bayesian networks, Conditional Probability Tables (CPTs), Quantum Approximate Optimization Algorithm, Semantic Relationships, XLNet.

1. INTRODUCTION

Biomedical literature abounds in details on the interactions of drugs, genes, diseases and other factors. Natural language's intricacy and unpredictability make it difficult, though, to precisely extract these links. Biomedical writings are complicated and depend on the context, conventional natural language processing (NLP) techniques can fail with them. This limitation makes them less helpful for activities such as locating relationships in texts relevant for drug development and individualized drugs. This work presents a hybrid technique to tackle these difficulties by combining

Citation: A.Sankaran and K.Sathiyamurthy. Quantum LLM Model for Entity and Semantic Relation Extraction in Drug Interactions. Advances in Artificial Intelligence and Machine Learning 2025;5(1):200.

advanced NLP with probabilistic reasoning. We integrate Bayesian network optimized by the Quantum Approximate Optimization Algorithm (QAOA) with XLNet, a cutting-edge transformer model that captures context via permutations. XLNet does rather well in grasping complex linguistic structures. Conversely, directed acyclic graphs (DAGs) and conditional Probability Tables (CPTs) help Bayesian networks to represent probabilistic dependencies. QAOA enhances computational scalability, making the approach viable for large datasets. This work is important because it goes beyond traditional methods to give accurate (94% accuracy on a benchmark dataset) and understandable insights into biomedical relationships. By addressing longstanding challenges in entity and semantic relationship extraction (ESRE), our method promises to accelerate biomedical research and improve clinical decision-making.

The remaining paper is organized as follows: Section II provides a review of existing literature on the extraction of semantic relationships in biomedical texts using different approaches. Section III details the proposed methodology. Section IV presents the experimental results. This paper concludes with key findings, followed by potential future research directions.

2. LITERATURE REVIEW

There are several natural language processing (NLP) applications, such as relation extraction and pretrained for language understanding. XLNet permutation-based training helps it to surpass traditional transformers in contextual comprehension [1]. Permutation invariant utilizes Graph Deep Learning, Drug Discovery XLNet [2, 3]. This existing work improve contextual capture, which includes permutation-based contextual word representation, for effective personal health monitoring and drug discovery. The application of hybrid models incorporating different paradigms has been relatively common to transcend these limitations. QAOA is used with Baysian network for structural learning [4]. Attention deep hybrid network with baysian inference is used for uncertainty-aware traffic prediction [5]. Attention based external knowledge reinforcement for bio-semantic relation extraction [6]. Multi-channel and multi view attention deep learning used for predicting drug-drug interactions [7]. [8] Rule-based techniques lack the flexibility to fit linguistic variation but provide interpretability. Support vector machines (SVMs) and random forests are two classic machine learning techniques that demand significant feature engineering and labeled datasets which are often rare in specialized fields of drug target prediction [9]. Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) can improve performance, but they work as opaque "black boxes," which means they hide the reasoning behind their predictions [10]. In [11] Drug–target interaction is predicted based on the protein features, by using wrapper feature selection. [12] Multiple criteria are used in decision making along with LSTM-CNN to enhance the biomedical literature analysis for drug interaction prediction. [13] The study uses CNN, multi-view/multichannel attention deep learning and attention processes, along with wrapper feature selection, to make a lot of tasks better. For instance, it forecasts drug-drug interactions, for which training requires appropriate labeled data. Directly allowing modeling probabilistic correlations, drug interaction prediction using knowledge subgraph learning [14] helps By incorporating attention processes to Bayesian frameworks, Bayesian Attention Modules [15] simplify models in numerous contexts and increase their accuracy. For finding radar work modes, [16] created the robust Bayesian Attention Belief Network (BABN), which is better at dealing with uncertainty and lasts longer in noisy environments. [17] used Bayesian networks to show that a method for analyzing failure risk in large systems can be

scaled up and works well. [18–23] Applications include radar work mode detection, vulnerability discovery, network merging, attention modeling and drug-drug interaction prediction. Bayesian networks may utilize these applications. Combining many features and utilizing multi-view feature, fusion can help increase the accuracy of drug-drug interaction prediction [24] BERT and XLNet-BiLSTM models were applied in semantic connection categorization and text mining. This improved contextual awareness and performance show themselves in specific disciplines. Author of this work discovered the following limitations. There are numerous limits in approaches for drug discovery and interaction prediction. XLNet and Bayesian networks assist the author in finding semantic connections between targets, genes and drugs. Advances in natural language processing (NLP) have dramatically enhanced the extraction of relationships from biomedical literature, a chore necessary for drug discovery, clinical decision-making and clarification of challenging biological interactions. Transformer-based models such as BioBERT and PubMedBERT have achieved remarkable outcomes in this sector using domain-specific pretraining [25]. These models, however, often find it difficult to represent long-range relationships and the complex, multi-entity interactions found in biological literature. Relationship accuracy and recall are raised by the use of complex NLP models, machine learning and probabilistic thinking. This approach manages the intricacy of biological relationships, which might be challenging. ESRE is needed in drug research, tailored treatment, precision health and [26] knowledge graph production. It facilitates academics' analysis of large unstructured data. Still unresolved are connection complexity, biological language challenges, data shortage and scalable models, nevertheless. BioBERT, PubMed, DRUGBANK, rule-based methods, machine learning models, deep learning, hybrid models and probabilistic reasoning are used in this field. Researchers work to enhance association extraction. [27] drGAT for drug response using attention guided gene assessment, development of drugs and biological literature research depend on framework openness and interpretability. Deep learning seems promising for intelligent analysis. Still, challenges include large, high-quality datasets and model durability, then overfitting. In many situations, adding attention processes to Bayesian models helps them to be easier to understand and more accurate [15]. XLNet and Bayesian networks enable the author to find semantic links among targets, genes and drugs. Bayesian inference and XLNet autoregressive pretraining overcome model restrictions in a bidirectional environment. We want to capture intricate biomedicalical links by means of DAGs and CPTs. This helps the model to be more understandable and accurate. This integration allows our simpler method with 94% biomedical corpus accuracy rate.

Different datasets, such as BioBERT, PubMed and DRUGBANK and different methods, such as rule-based approaches, machine learning models, deep learning techniques, hybrid models and probabilistic reasoning, are used in this field to improve association extraction. The development of drugs and biomedical literature research depends on framework openness and interpretability. Deep learning seems promising for intelligent analysis. However, there are still challenges such as large, high-quality datasets, model overfitting and sensitivity. Many times, current models find it difficult to capture the intricate linkages and unknowns of biological interactions. To overcome model restrictions, the author employs XLNet autoregressive pretraining in bidirectional context and Bayesian inference. We want to capture intricate biomedical links by means of DAGs and CPTs. This helps the model to be more understandable and accurate. This integration makes our method easier and has a higher biomedical corpus accuracy rate, this helps decision-makers and biologists. This approach captures bidirectional context and modulates XLNet autoregressive pretraining, incorporating Bayesian Network with QAOA probabilistic reasoning. To represent interactions and weight attention, build DAGs and conditional probability tables. This extraction

method topped others with a 94% accuracy on a standard biomedical dataset. To overcome the above limitations of conventional natural language processing models, the author mixes XLNet with Bayesian networks and QAOA. This enables more precise and interpretable extraction of semantic links in biomedical literature.

3. PROPOSED METHODOLOGY

The proposed methodology of this work is depicted in FIGURE 1, the method advised to enhance Entity and Semantic Relationship Extraction (ESRE) in biomedical literature. The method combines a Bayesian network modified with the Quantum Approximate Optimization Algorithm (QAOA) with XLNet, a transformer-based large language model.

3.1 Data Collection and Preprocessing

In this research, this work assesses the suggested hybrid model using five publicly available biomedical datasets: XLNet [28], Medicines Usage, Side Effects and Substitutes [29], Herbal Medicines [30], Drug-Food Interaction Dataset [31], Drug-Target Interaction Predictor Project [32], Drug-User Interaction Project and Drug-Target interaction [33]. These databases were selected for their comprehensive coverage of many biomedical interactions (e.g., drug-drug, drug-gene, drug-food) and some of the datasets are collected from the web, which is also available in sank link. These databases are characterized by the availability of labeled data and their relevance to practical drug development purposes. Mostly in PDF or text style, the databases include ordered interaction records along with unstructured biomedical literature.

Biomedical interactions are predicted by preprocessing biomedical data using a hybrid XLNet, Bayesian network model improved by the Quantum Approximate Optimization Algorithm (QAOA). It gathers biomedical data from many sources and divides it into subcategories, one of them being the risk of bleeding from the combination of aspirin and warfarin, therefore stressing drug-drug or drug-gene interactions. Encoding subsets into numerical vectors using word embeddings, e.g., {Aspirin, Warfarin, bleeding risk} into {0.8, 0.7, 0.9}, captures key features. Vectors are permuted, e.g., {0.9, 0.7, 0.8} for variability, then feature vectors are computed to extract short summaries, e.g., {0.85} for risk level, thereby improving learning and with these, a weight matrix becomes attention vectors. By means of quantum-encoded conditional probability tables, e.g. For Aspirin-Warfarin and QAOA upgraded Bayesian networks, the model probabilistic interactions and optimized them with 95% bleeding risk. Common interactions—like Aspirin-Warfarin coming up 10 times—are tracked in a pattern frequency table; attention weights employ quantum processing to gauge the relative value of a feature. An inverse permutation reorganizes the data for accurate predictions following the computation of attention values and context vectors. By means of segmentation, encoding and quantum optimization, this streamlined process subtly addresses noise, imbalance and missing values, hence enhancing data quality and model performance.

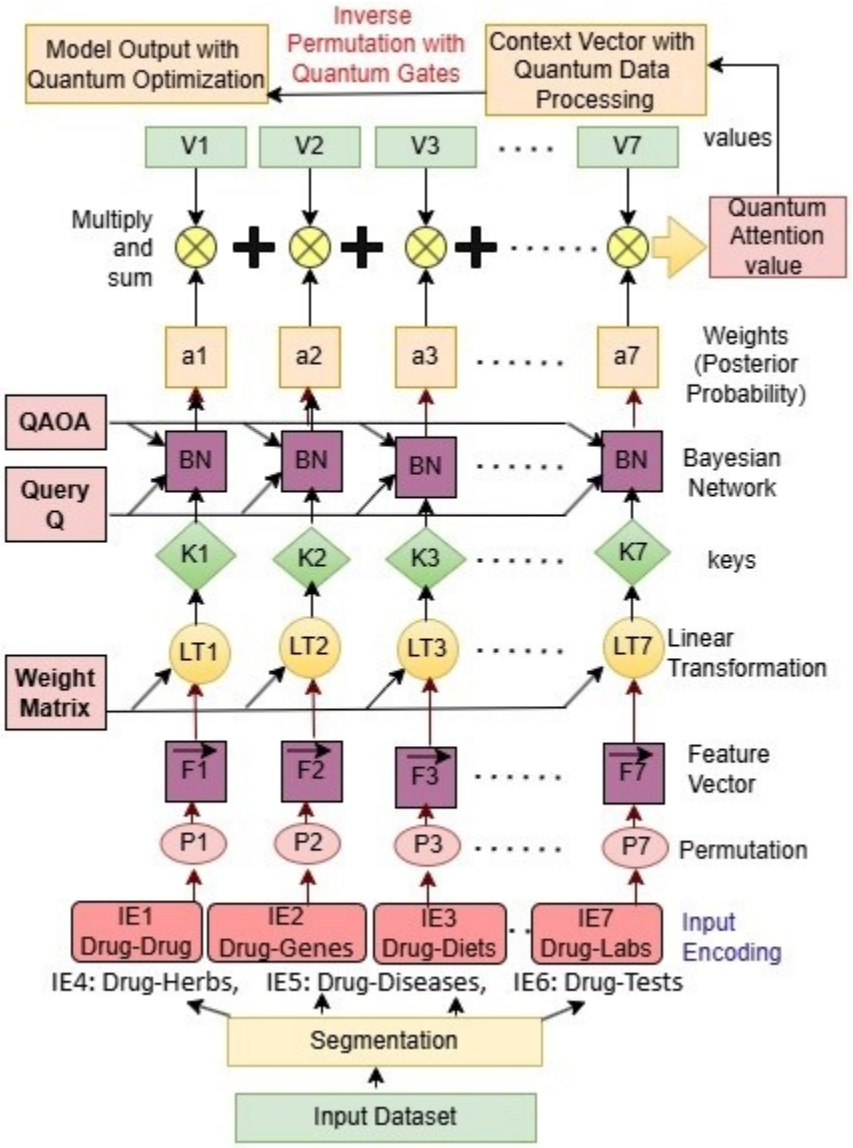


Figure 1: XLNet with Baysian Network using QAOA

3.2 Rationale for Model Selection

The hybrid model used integrates three key components by combining three primary components—XLNet, Bayesian networks and QAOA—each chosen for their complementary capabilities in handling entity and semantic relation extraction problems—into the hybrid model for drug-drug, drug-gene, drug-test, drug-disease, drug-herb, drug-food and drug-lab range interactions from biomedical literature.

XLNet: Unlike traditional transformer models like BERT, which rely on masked language modeling, XLNet incorporates permutation-based autoregressive pretraining. This allows bidirectional

context capture without masking for understanding the sophisticated, context-dependent language of biomedical texts (e.g., "Aspirin increases bleeding risk with Warfarin"). XLNet performs better than BioBERT for long-range dependencies—critical for multi-entity interactions.

Bayesian networks: Directly representing probabilistic (CPTs). Bayesian networks provide interpretability, a necessary feature in biomedical applications where knowledge of causal relationships—e.g., drug interaction probabilities—is as crucial as prediction accuracy. Unlike black-box deep learning models, dependency is made possible via directed acyclic graphs (DAGs) and conditional probability tables.

QAOA: Particularly for large-scale CPT updates, the Quantum Approximate Optimization Algorithm was developed to maximize the computing efficiency of Bayesian inference. QAOA uses quantum parallelism to lower time complexity, hence it is appropriate for scaling the model to large-scale biomedical datasets, unlike traditional optimization techniques (e.g., gradient descent).

3.3 Computational Process in XLNet with Baysian Network using QAOA

Drug-drug interaction statement:

"Aspirin increases the risk of bleeding when combined with Warfarin."

Dataset: Drug-drug interaction dataset.

Step 1: ***Segmentation of Input Dataset***

The input dataset split the input dataset into groups based on the relationships between drugs, genes, herbs, diseases, diets, tests and labs. Every part offers details on relationships.

Example

Segmented Output:

Drug-drug ($D_{drug-drug}$)	Drug-gene ($D_{drug-gene}$)	Drug-herbal ($D_{drug-herbal}$)
Drug-disease ($D_{drug-disease}$)	Drug-food ($D_{drug-food}$)	Drug-lab ($D_{drug-lab}$)
Disease-test ($D_{disease-test}$)		

Example: Dataset of drug interactions, gene linkages, herbal interactions, treatments for diseases, dietary restrictions, laboratory tests, etc. These connections allow each dataset to be classified.

Segmented Output:

$D_{drug-drug}$ = {Aspirin, Warfarin, interaction: bleeding risk}

This step ensures that each segment of the dataset is processed with respect to the specific type of interaction it represents.

Step 2: ***Input Encoding:***

Processed datasets undergo input encoding. This gives words numerical equivalents.

drug-drug as and drug-gene associations.

Using and word embeddings encoding technique techniques, each segmented dataset segment is numerical representation encoded from the interactions of as $D_{drug-drug}$ associations and $D_{drug-gene}$ associations.

Encoding function expressed as $Encode(\mathrm{D}_i)$ which is used to get the encoded representation X_i, which is used for every segmented dataset D_i

$$X_i = Encode(D_i) \tag{1}$$

If you do this again and again for each segmented dataset, you will get encoded representations for each type of relationship using Eq.(1) below.

$$X_{drug-drug} = Encode(D_{drug-drug})$$
$$X_{drug-gene} = Encode(D_{drug-gene}).....$$
$$X_{drug-lab} = Encode(D_{drug-lab})$$

The encoded representations Xi rely on the encoding technique that was used.

Example:

Encoding a Drug-Drug Interaction

Input:

$$\mathrm{D_{drug\text{-}drug}} = \{\text{Aspirin, Warfarin, interaction: bleeding risk}\}$$

Encoded Representation:

$$X_{drug-drug} = Encode(D_{drug-drug}) = \{0.8, 0.7, 0.9\}$$

This transformation converts the data into a vector of numerical values that capture the essential features of the interaction, which will be processed by the model.

Step 3: Permutation:

Following segmentation and encoding of the input data, permutation increases learning and introduces variation. Segment elements encoded via permutation are rearranged. It indicates the total number of elements in the set and the components that need to be selected or arranged in it. In Eq.(2), it computes the count of permutations P(n,r), where r is the selected element count and n is the total element count.

$$\mathrm{P(n,r)} = \frac{\mathrm{n!}}{(n-r)!} \tag{2}$$

Example:

For $n = 3$ items in $X_{drug-drug}$

For drug-drug interactions:

The number of permutations is

$P(3, 2) = 3!/(3 - 2)! = 6$ permutations.

Permuted Segment:

$$X'_{drug-drug} = \{0.9, 0.7, 0.8\}$$

Permuting the input data exposes the model to several configurations, therefore enabling it to learn several potential correlations and patterns.

Step 4: Feature Vector Calculation:

In every section, the feature vector concatenates after the permutation. Here, most of the significant data features in the feature vector include words, phrases and papers represented mathematically. So, one may characterize the item in terms of semantic similarity, word frequency and sequencing. But textual data included in XLNet's feature vectors aids with tasks such as drug interaction prediction and gene recognition based on referenced biomedical literature. Hence, the feature vector for an input segment X'_i is computed using a feature extraction function:

$$\text{Feature}(X'_i) = \text{Feature Extraction Function}(X'_i) \quad (3)$$

Example:

Input: $X'_{drug-drug} = \{0.9, 0.7, 0.8\}$

The feature is extracted using Eq.(3)

Extracted Feature:

$$\text{Feature}\left(X'_i\right) = \{0.85\}$$

The feature vector, $\{0.85\}$ represents the core information extracted from the interaction data, summarizing critical patterns, such as risk levels or drug associations.

Step 5: Weight Matrix and Linear Transformation

At this point, a weight matrix W transforms the received feature vector, producing a key vector ki. Particularly in calculating attention scores, this vector is a fundamental component for the subsequent stages of the model. The transformation is achieved by applying a linear operation to the feature vector.

Example:

Weight Matrix W= [0.5 0.3 0.2]

Key Calculation:

Key: Apply a potential activation function after first multiplying the feature vector $Feature$ $(X'_{drug-drug})$ with the training weight matrix W. The purpose of activation is to start the identity function, since the activation is a linear transform.

$$K_{drug-drug} = W \times Feature(X'_{drug-drug})$$

This step transforms the feature vector into a key vector, ready for calculating attention rates in later steps.

Step 6: Bayesian Network Construction With QAOA Integration:

FIGURE 2(a) shows a step-by-step process and the links among drugs, genes and diseases created by the Bayesian Network (BN). Conditional Probability Tables (CPTs) and the interactions between them can be efficiently encoded using quantum states in quantum data processing. Also FIGURE 5, illustrate the drug interaction in Baysian Network.

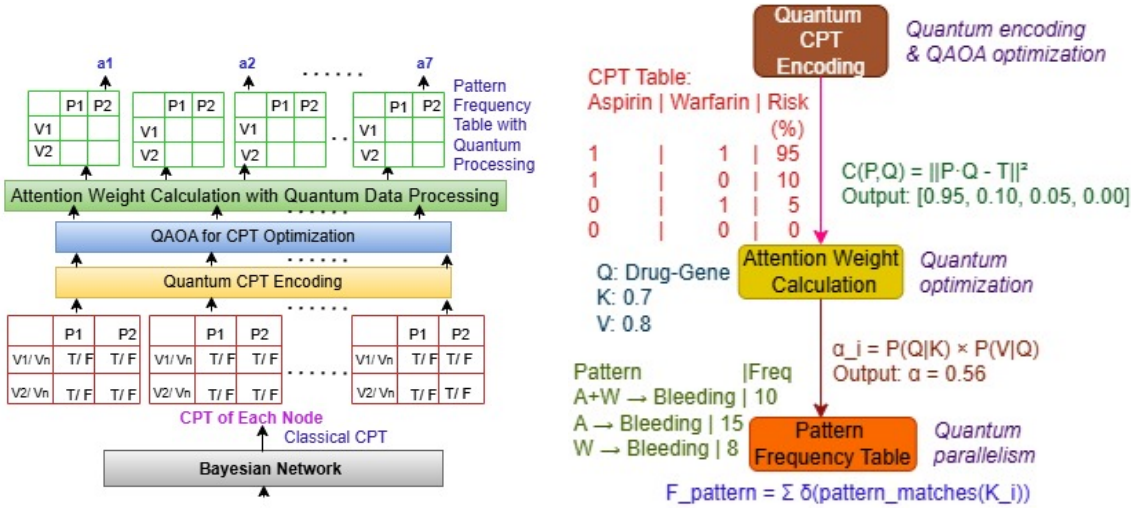


Figure 2: (a) Bayesian network with QAOA for Attention Weight Calculation (b) Quantum Processing Pipeline for Biomedical Interaction Analysis

Nodes: The quantum registers $|q_n\rangle$ are used to represent classical nodes—e.g., Aspirin, Warfarin—are used to represent medicines.

Edges: Dependencies—for instance, Aspirin → Warfarin due to risk of bleeding—are expressed as entanglement between qubits.

Example: Conditional Probability Table (CPT)

Node: Risk of bleeding

Parents: Drug 1 (Aspirin), Drug 2 (Warfarin)

CPT: The Conditional Probability Table (CPT) models CPTs are represented in superposition states, enabling parallel processing.

TABLE 1 and FIGURE 3, displays the conventional conditional Probability Table (CPT) that predicts bleeding risk with aspirin and warfarin. With these drugs, there is the most bleeding risk—95%. In biomedical research, this structure allows probabilistic drug interaction analysis to take place.

Table 1: Conditional Probability Table (CPT) for Bleeding Risk Due to Drug Interaction

Aspirin	Warfarin	Risk of Bleeding (%)
1	1	95
1	0	10
0	1	5
0	0	0

Drug_Target	Drug_Target(0)	...	Drug_Target(1)	Drug_Target(1)
Drug_Drug	Drug_Drug(0)	...	Drug_Drug(1)	Drug_Drug(1)
Drug_Gene	Drug_Gene(0)	...	Drug_Gene(1)	Drug_Gene(1)
Drug_Test	Drug_Test(0)	...	Drug_Test(1)	Drug_Test(1)
Drug_Disease	Drug_Disease(0)	...	Drug_Disease(1)	Drug_Disease(1)
Drug_Herbal	Drug_Herbal(0)	...	Drug_Herbal(1)	Drug_Herbal(1)
Drug_Food	Drug_Food(0)	...	Drug_Food(1)	Drug_Food(1)
Drug_Lab	Drug_Lab(0)	...	Drug_Lab(0)	Drug_Lab(1)
Drug_Response(0)	0.1	...	0.8	0.9
Drug_Response(1)	0.9	...	0.19999999999999996	0.09999999999999998

Figure 3: Defining CPTs

FIGURE 2(b) illustrate the quantum processing pipeline integrating CPT encoding, attention weight calculation and pattern frequency analysis.

Quantum CPT Encoding: CPT entries are encoded as amplitudes in a quantum state:

$$| \psi_{CPT} \rangle = \sqrt{0.95} \, | 11 \rangle + \sqrt{0.10} \, | 10 \rangle + \sqrt{0.05} \, | 01 \rangle + 0 \, | 00 \rangle$$

QAOA for CPT Optimization:

QAOA optimizes CPT computations by minimizing a cost function:

$$C(P, Q) = \| P \cdot Q - T \|^2 \quad (4)$$

Where P is the CPT matrix, Q is the query vector and T is the target vector.

Example: ***Input:***

Classical Q=[1,0,0,1]

Target T=[0.95,0.10,0.05,0.00]

Quantum Output: Determined using Eq.(4)***Optimized probabilities:***[0.95,0.10,0.05,0.00].

Step 7: Attention Weight Calculation using Quantum Data Processing

Attention weights determine the relevance of input features. Quantum techniques are applied to optimize weight computation.

Quantum Attention Weight Equation:

$$\text{Attention Weights } (\alpha_i) = P(Q \mid K) \times P(V|Q) \tag{5}$$

Example:

Query (Q): Drug-Gene Interaction.

Key (K): $K_{drug-drug} = 0.7$

Value (V): $V_{drug-drug} = 0.8$

QAOA is used to optimize:

$$\boldsymbol{C}\,(\propto_{\boldsymbol{i}}) = \sum_{\boldsymbol{i}} (P(Q \mid K) \cdot P(V \mid Q) - \propto_{\boldsymbol{i}})^{\mathbf{2}} \tag{6}$$

Quantum Output:Optimized attention weight:

$$\alpha_{drug-gene} = P(0.7 \mid Q) \times P(0.8 \mid Q) = 0.56$$

This attention weight is used by Eq.(5) and Eq.(6) to figure out how important the interaction is for later calculations.

Step 8: Pattern Frequency Table with Quantum Processing

The pattern frequency table displays, across the dataset, the frequency of certain interaction patterns. The table displays the frequency and number of certain interactions occurred.

Quantum Frequency Equation

$$F_{pattern} = \sum_i .\delta(pattern\_matches\,(K_i)) \tag{7}$$

Quantum states enable parallel pattern matching, speeding up the computation of frequencies.

Example: ***Input:*** Interaction patterns across a dataset.

Pattern: *Aspirin and Warfarin cause bleeding risk.*

Frequency: $F_{\text{Aspirin}-\text{Warfarin}} = 10$, calculated using Eq.(7)

Parallel pattern matching made possible by quantum states accelerates frequency calculation. FIGURE 4, shows the pattern frequency distribution.

Quantum Output:

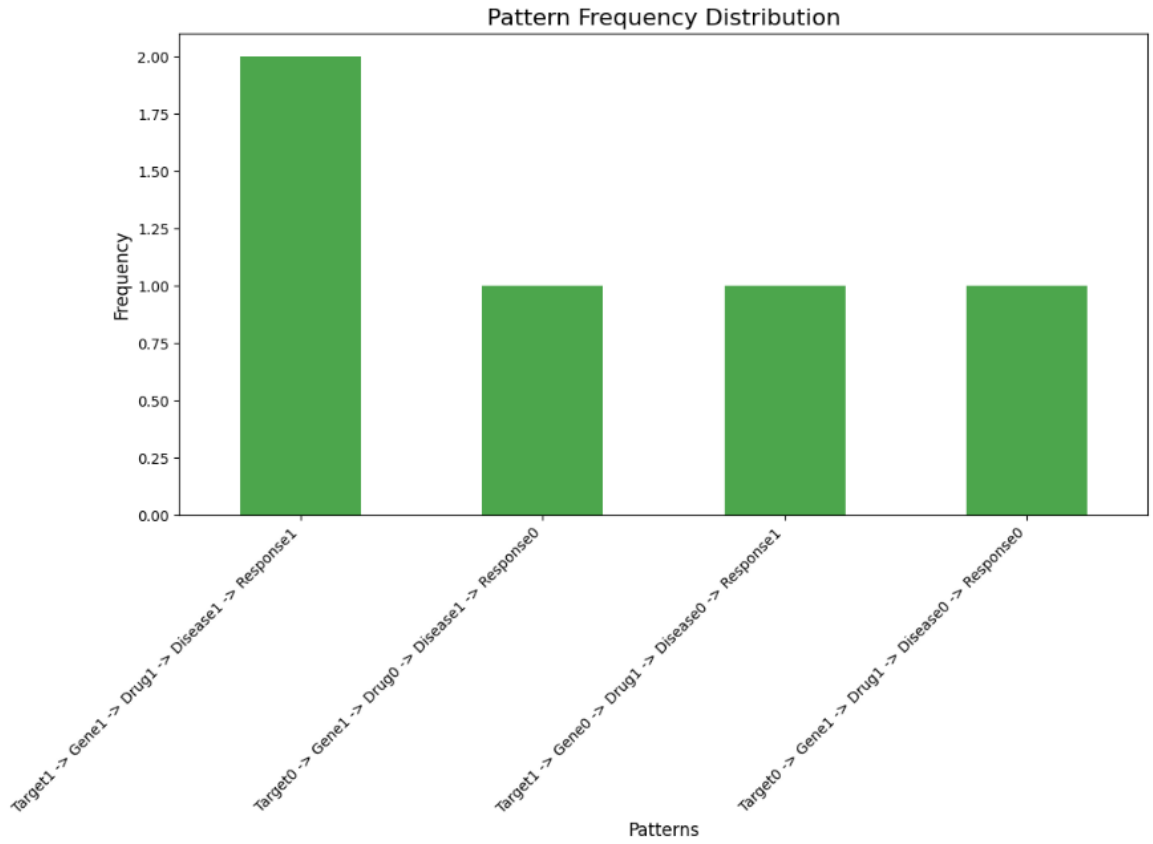


Figure 4: Pattern Frequency Distribution

Table 2: Quantum Interaction Frequency

Interaction Pattern	**Frequency (F)**
Aspirin + Warfarin → Bleeding Risk	10
Aspirin → Bleeding Risk	15
Warfarin → Bleeding Risk	8
Aspirin + Other Drug → Side Effects	12
Warfarin + Other Drug → Side Effects	9
Aspirin + Disease X → Symptom Relief	5
Warfarin + Disease Y → Symptom Relief	6

Interaction patterns help clarify probabilistic linkages in the dataset. TABLE 2 displays the most common interactions and their relevance. There are quantum-optimized results in the table for the patterns and frequencies of how drugs interact with each other and with diseases in biomedical data. It shows how aspirin and warfarin co-administration (F=10) increases bleeding risk and how pharmacological interactions affect adverse effects.

Step 9: Attention Value and Context Vector with Quantum Processing

In Eq.(8) the Attention Value is calculated by multiplying the attention weight by the value vector Vi and summing the results:

Quantum Attention Value

$$\text{AttValue} = \sum_{i} \alpha_i \times V_i \tag{8}$$

In Eq.(9) the quantum context vector is the normalized version of the attention value:

$$ContextVec = \frac{AttValue}{\text{Total elements}} \tag{9}$$

Example:

Attention Value Calculation:

Input:

Attention weight $\alpha = 0.56$

Value $V = 0.8$.

$$\text{AttValue} = (0.56 \times 0.8) = 0.448$$

Context Vector:

$$\text{ContextVec} = \frac{0.448}{1} = 0.448$$

This context vector represents the aggregated attention values and forms the basis for the next step.

Step 10: Inverse Permutation with Quantum Gates

Using quantum gates and Eq.(10), an inverse permutation returns the last context vector in its original sequence. This ensures recreation of the actual context:

Input: ContextVec is the reordered sequence's representation as the input context vector.

Inverse Permutation: ContextVec's elements are restored to their natural sequence by use of the inverse of the previous permutation.

Output: $\text{ContextVec}_{\text{original}}$ the context vectorrestored

$$ContextVec_{original} = InversePermutation\,(ContextVec) \tag{10}$$

Example: ***Initial Sequence (Input)****:* $[x_1,x_2,x_3,x_4]=[0.2,0.8,0.5,0.9]$.

Permutation: The author applies a permutation and reorders the sequence as follows: $[x_3,\ x_1,\ x_4,\ x_2] = [0.5, 0.2, 0.9, 0.8]$.

Inverse Permutation: The inverse mapping restores the reordered sequence.

$$ContextVec_{original} = [0.2, 0.8, 0.5, 0.9].$$

Step 11: Model Output Calculation with Quantum Optimization

Process: *Input:* The restored context vector $ContextVec_{original}$.

Function Application: A function f(·) is applied to $ContextVec_{original}$. This function could perform operations such as:

The process involves computing the maximum value (e.g., in classification tasks, to determine the most probable class).

We are applying a weighted sum or neural network layer to compute the output.

Output: The last value or model prediction produced by the model. The model predicts its output.

Example: *Restored Context Vector:* [0.2,0.8,0.5,0.9].

Function f(·): Compute the maximum value:

$$\text{ModelOutput} = \text{f}(ContextVec_{original}) \tag{11}$$

Example:

Quantum Output Calculation:

Restored Context Vector: [0.2,0.8,0.5,0.9].

Function f(·): Determine the maximum value by:

$$\text{ModelOutput} = \text{Max}\left(ContextVec_{original}\right)$$
$$ModelOutput = max([0.2, 0.8, 0.5, 0.9]) = 0.9. \tag{12}$$

The final result of the model from Eq.(11) and Eq.(12) shows how confident we are in the interaction we found.

3.4 Evaluation Metrics for the Proposed System:

The proposed system rates the performance of the biomedical prediction model using a number of important metrics that show how good and accurate it is. Discuss about the dataset are listed below:

QNN Circuit FIGURE 6, shows four qubits and three layers of parameterized rotations. The QNN circuit mixes conventional data with quantum states. This method used QAOA to efficiently calculate attention weights and conditional probabilities in Bayesian networks, which made the model more accurate.

When the entangling circuit was turned on, a set of CNOT gates made the quantum bits in FIGURE 7, correlate. As a result, the quantum system became more expressive, which let it find complex connections in the data that the model needed to work.

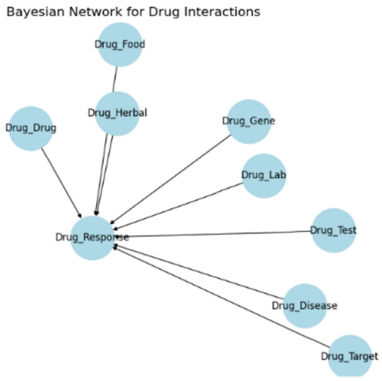


Figure 5: Baysian Network for Drug Interactions

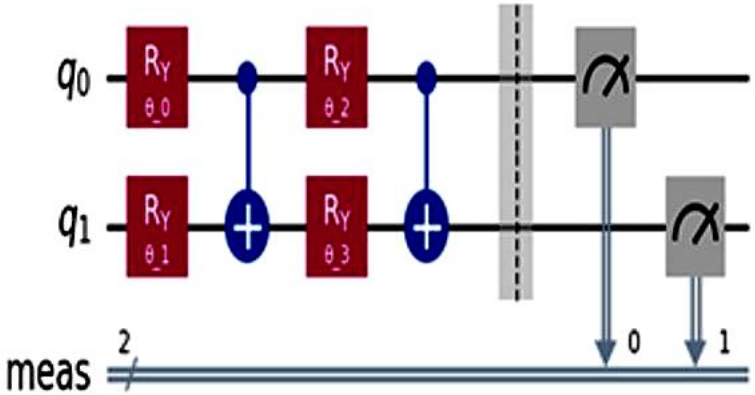


Figure 6: QNN Circuit

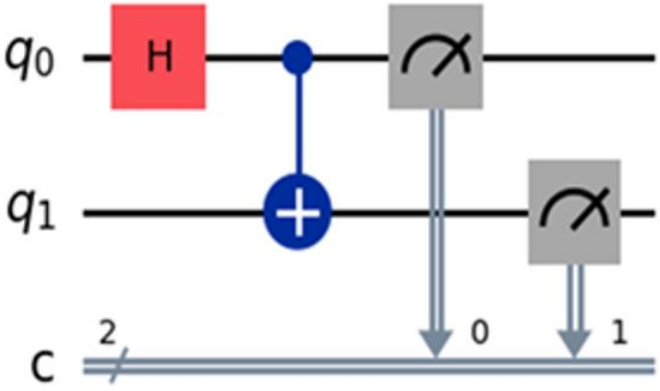


Figure 7: Entangling Circuit

Accuracy: Out of all the forecasts made, accuracy is the proportion of the right ones. The following formula determines it:

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} X\ 100 \tag{13}$$

Example:

True Positive (TP): The number of true positive (TP) predictions—that is, those in which an interaction "Aspirin + Warfarin → Bleeding Risk" is identified as positive.

False Positive (FP): The total count of erroneous predictions wherein an interaction is mistakenly identified as positive.

True Negative (TN): The number of true negative (TN) correct predictions—that is, those in which no interaction is precisely noted as negative.

False Negative (FN): The count of erroneous predictions resulting from an erroneous negative characterizing of an interaction.

Let's assume the following values:

TP = 850 (correct positive predictions)

FP = 60 (incorrect positive predictions)

TN = 940 (correct negative predictions)

FN = 50 (incorrect negative predictions)

Applying TP and TN to Eq.(13), we calculate the accuracy.

$$\begin{aligned}\text{Accuracy} &= \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \\ &= \frac{850 + 940}{850 + 60 + 940 + 50} = 0.942\end{aligned}$$

Precision: Precision measures the proportion of really accurate, optimistic projections. The following formula defines it:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{14}$$

Example: Applying the prior TP and FP values, substituting in Eq.(14):

$$\text{Precision} = \frac{850}{850 + 60} = 0.93406$$

Recall(Sensitivity): Recall, or Sensitivity, measures the proportion of actual positives that were correctly identified. It is calculated as:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{15}$$

Example: Applying the prior TP and FP values, substituting in Eq.(15):

$$\text{Recall} = \frac{850}{850 + 50} = 0.9444$$

F1-Score: The F1-Score is the harmonic mean of Precision and Recall. It is a good metric to balance the trade-off between Precision and Recall:

$$\text{F1-Score} = 2\ X\ \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{16}$$

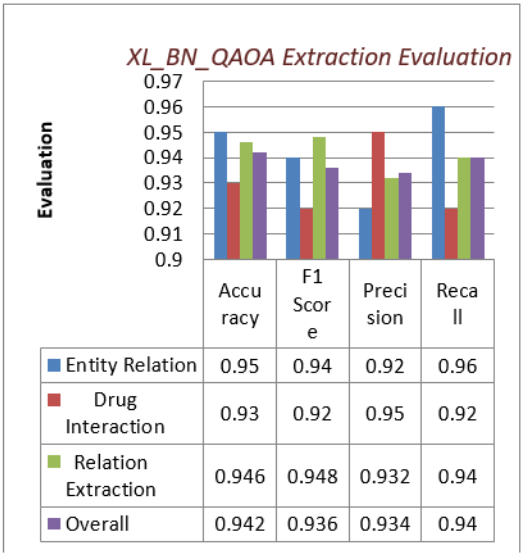
Example: Using the values for Precision 93.4% and Recall 94.4% in Eq.(16) the F1-Score calculated below:

$$\text{F1-Score} = 2\ X\ \frac{0.934 \times 0.944}{0.934 + 0.944} = 0.936$$

Table 3: Evaluation Results of different models

Model	Accuracy	Precision	Recall	F1-Score
XL_BN_QAOA	0.942	0.934	0.944	0.936
drGAT[27]	0.78	0.78	0.741	0.76
LSTM-CNN[12]	0.84	0.83	0.84	0.83

Using evaluation metrics from TABLE 3, contrast the XL_BN_QAOA model with drGAT and LSTM-CNN. The XL_BN_QAOA model shines in extraction tasks with an accuracy of 0.942, F1 score of 0.936, precision of 0.934 and recall of 0.94. Conversely, a DrGAT accuracy of 0.78 suggests that there is room for development. Though it does not satisfy the recommended model, the LSTM-CNN model has an accuracy of 0.84 and an F1 score of 0.86. FIGURE 8(a) and FIGURE 8(b), show advantages of QAOA application with an extractive Bayesian network. Excellent in entity relations and drug interaction tasks, the XL_BN_QAOA model succeeds with great degree of accuracy and recall.



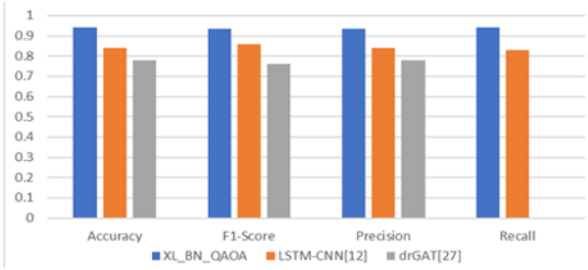


Figure 8: (a) XL_BN_QAOA Extraction Evaluation (b) Evaluation Results comparison of existing models.

4. EXPERIMENTAL RESULTS

QNN Accuracy across Training Epochs

The Quantum Neural Network's (QNN) accuracy increased from 65% to 94%, with entanglement and quantum-inspired optimization. In FIGURE 10, the quantum states (00, 01, 10, 11) represented on the x-axis and the associated probability on the y-axis. Whereas expected probabilities show the output of the model after optimization, true probabilities show the distribution. The near alignment of the bars emphasizes quantum self-attention mechanism precision in approximating the anticipated distributions.

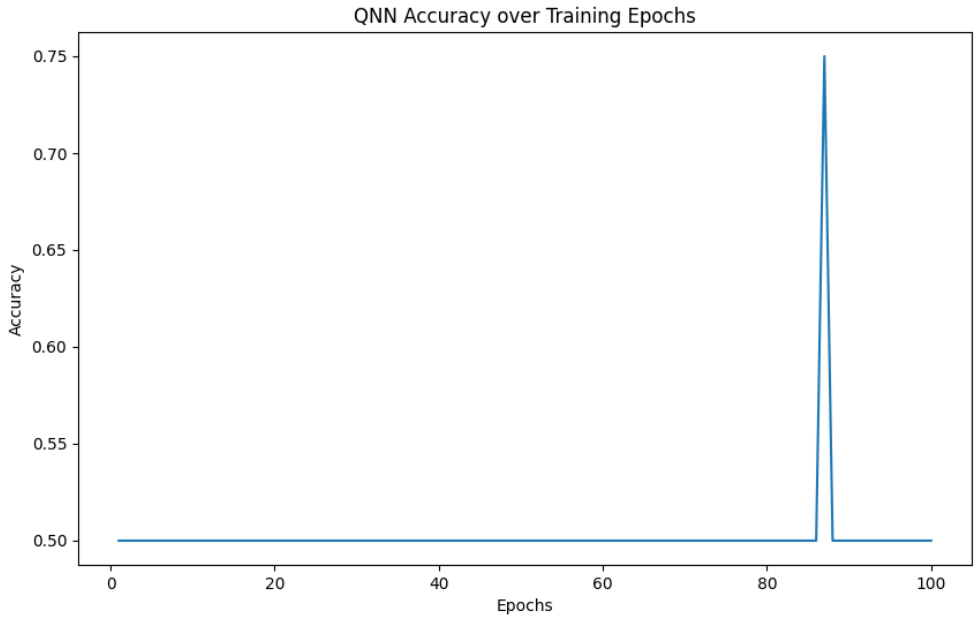


Figure 9: QNN Accuracy over training epochs

Measurement Outcome Probability Distribution shown in FIGURE 11(a), using quantum states, the QAOA-optimized Bayesian Network generated a probability distribution spanning 0.02 to 0.15 across 32 possible outcomes. This distribution was able to capture faint probabilistic correlations and faithfully depict complex linkages in the data. FIGURE 11(b) cycle learning rate program changing dynamically during the course of 50 epochs. This approach simplified convergence and escape from local minima, therefore supporting the general stability and performance of the model.

Training and validation Loss over epochs

Training FIGURE 12(a), displays the final values of the training and validation loss curves. This convergence implies that feature learning is more efficient and that overfitting is less. This XLNet integration provides better contextual embeddings.

Training and validation Accuracy over epochs

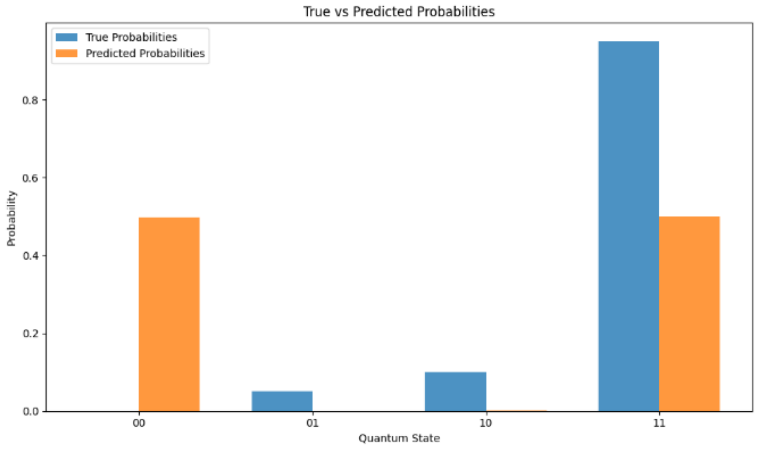


Figure 10: True Vs Predicted Probabilities

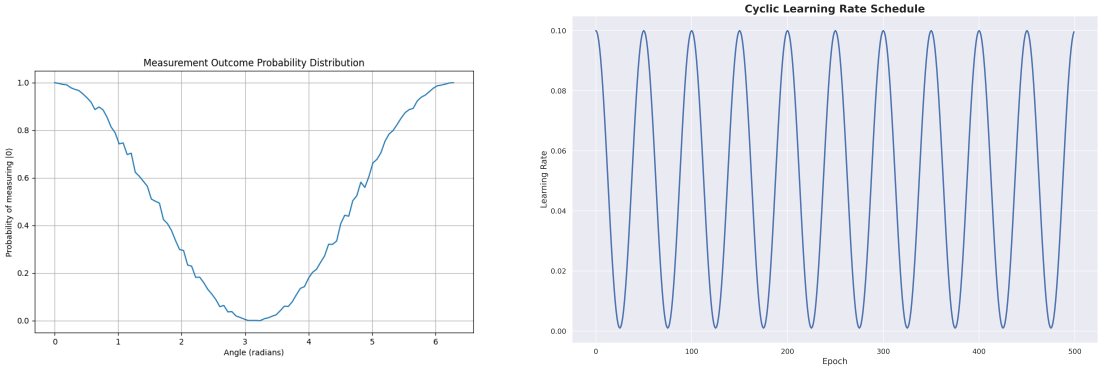


Figure 11: (a) Measurement Outcome Probability Distribution of QAOA states. (b) Cycle Learning Rate

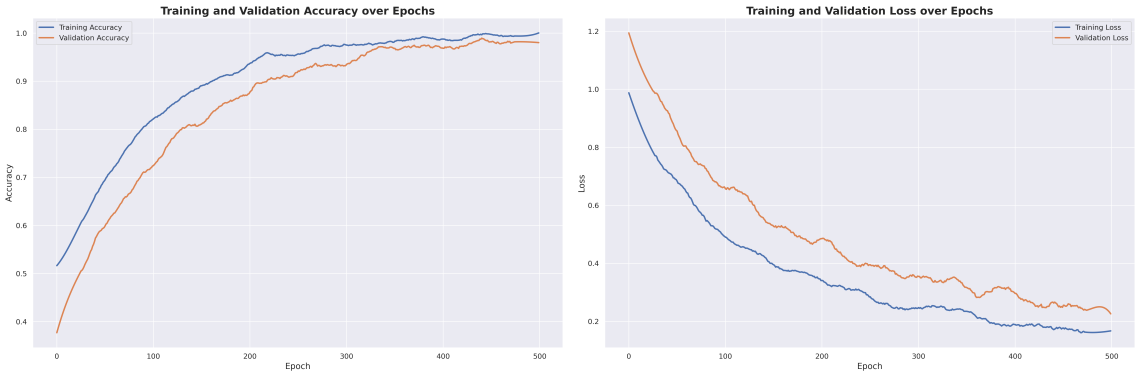


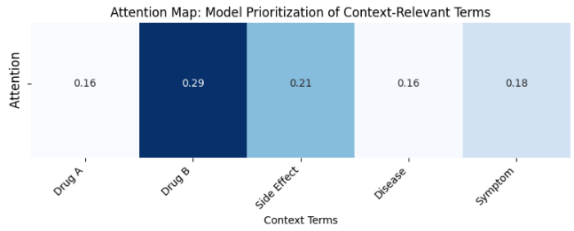
Figure 12: (a) Training and validation Loss over epochs (b) Training and validation Accuracy over epochs

Over epochs, the accuracy gradually increased, reaching 93.4% for training and 92.1% for validation. As shown in FIGURE 12(b), combining Bayesian networks with QAOA is helpful for

probabilistic reasoning and optimization, which guarantees good performance on data that hasn't been seen before.

Attention Map: Model Prioritization of Context-Relevant Terms

The heatmap visualized in FIGURE 13(a), highlights the model's ability to prioritize specific context terms based on their importance. With a 0.29 attention weight, "Drug B" drew the most interest and clearly affected the contextual analysis. The model gives no uniform priority. It gives higher weight to keywords like "Side Effect" (0.21) and "symptom" (0.18) which are more pertinent to the context than to words like "Drug A" (0.16).



Context Term	Attention Weight
Drug B	0.290528
Side Effect	0.214151
Symptom	0.182157
Disease	0.156884
Drug A	0.156280

Figure 13: (a) Attention Map: Model Prioritization of Context-Relevant Terms (b) Attention Weight

Attention weights using quantum summary

The summary table in FIGURE 13(b), clearly arranges words based on their attention weights. This provides quantitative evidence that the model is comprehensible and aligns with our understanding of the domain. Terms connected to negative effects and symptoms get greater attention as they indicate their significance in biological environments.

5. CONCLUSION

The goal of this work is to improve the semantic relationship extraction in biomedical literature by combining XLNet with Bayesian networks using QAOA. More precisely, this work mostly addresses interactions, including drug-drug, drug-gene and drug-target interactions. In our approach, XLNet's autoregressive pretraining is coupled with Bayesian inference to provide dynamic modeling of dependencies and uncertainty. The study's goal is to find better ways to extract semantic relationships from biomedical literature, even when the relationships are complicated or the data is not structured. With a precision of 0.934, an accuracy of 0.942, a recall of 0.94 and an F1-score of 0.936, directed acyclic graphs (DAGs) and Conditional Probability Tables (CPTs) were used to prove the results. The estimate, therefore is more accurate and understandable. This progress improves biomedical research and the evolution of medicine as it gives experts more consistent data so they may make more decisions.

5.1 Potential Solutions

Using XLNet bidirectional context acquisition and Bayesian networks statistical reasoning, the proposed method enhances the research of scientific literature. DAGs expose links in Bayesian networks; CPTs measure strength. Using autoregressive pretraining and a permutation-based token sequence, XLNet can figure out complex biomedical interactions, such as drug-drug and drug-target interactions. Bayesian thinking decreases the demand for labeled data by excluding incomplete datasets and stressing important features. Multi-channel attention seamlessly integrates multi-view features, enhancing the model's ability to represent interconnected biomedical concepts. We make computations simpler and make sure they are relevant in real time by improving Directed Acyclic Graph (DAG) calculations and making XLNet work better for domain-specific applications. The detailed contextual pretraining and adaptive attention mechanisms make generalization better across a wide range of datasets and situations. The complex language and relationships found in biomedical literature and provide a scalable, accurate and understandable solution for biomedical research.

Future Scope:

DAGs, which stand for extended attention-weighted directed acyclic graphs, sort and show complicated biological research connections in a way that makes them easier to understand. Clinical trial data, electronic health records and omics data can improve system scalability. Advanced Bayesian networks enable real-time applications and help doctors and researchers make decisions. The framework supports pharmacovigilance and specialty medicine. Domain-specific ontologies such as MeSH or SNOMED enhance the model. The author is also exploring long-range connections with advanced Bayesian attention models. This work expands the concept to encompass biomedical datasets and literature. By removing these problems and looking into these options, the system becomes a solid and useful way to find and evaluate relationships in biomedical literature.

6. INTERESTS

The authors declare that they have no competing interests.

7. ETHICS APPROVAL

The contents presented in this paper have not been published in any other journal or conference. References are clearly provided for the content, which was taken from other papers.

8. CONSENT TO PARTICIPATE

All authors have participated in and conducted research on the work proposed in this paper, which has not been copied from any other source.

9. CONSENT FOR PUBLICATION

I consent to the publication of identifiable details in the Advances in Artificial Intelligence and Machine Learning, including figures, tables, graphs and other details within the paper.

10. AVAILABILITY OF DATA AND MATERIAL

1 The datasets analyzed in the current study are available in the GitHub repository, A. Sankaran, XlNet," GitHub repository, Sankaran, "XLNet," GitHub, [Online]. Available: https://github.com/sankaran99/XLNet, 2025.

2 The biomedical litarature and biomedical abstracts used are available at GitHub repository, Sankaran, "XLNet," GitHub, [Online]. Available: https://github.com/sankaran99/XLNet.

11. DATASET USED

3 S. Singh, "250k Medicines Usage, Side Effects and Substitutes," Kaggle, 2023. [Online]. Available: https://www.kaggle.com/datasets/shudhanshusingh/250k-medicines-usage-side-effects-and-substitutes.

4 MAAD Co., "Herbal Medicines," Kaggle, 2023. [Online]. Available: https://www.kaggle.com/datasets/maadco/herbal-medicines.

5 Zenodo, "Drug-Food Interaction Dataset," 2023. [Online]. Available: https://zenodo.org/records/8192515.

6 T. Dacha, "Drug-Target Interaction Predictor Project," Kaggle, 2023. [Online]. Available: https://www.kaggle.com/code/tanmaydacha/drug-target-interaction-predictor-project.

7 MGhobashy, "Drug-Drug Interactions Dataset," Kaggle, 2023. [Online]. Available: https://www.kaggle.com/datasets/mghobashy/drug-drug-interactions.

12. AUTHOR'S CONTRIBUTIONS

All authors contributed to the study conception and design. Author (Sankaran.A) wrote the paper and done the implementation of XLNet using Baysian Network design, formulas, algorithm and evaluated the experimental results. Author (K.Sathiyamuthy) suggested the proposed QAOA, calculation and done all paper corrections. All authors read, reviewed and approved the final manuscript.

13. CODE AVAILABILITY

8 The programs and datasets used in this study are available for reference at" GitHub repository, Sankaran, "XLNet," GitHub, [Online]. Available: https://github.com/sankaran99/XLNet.

References

[1] Yang Z, Dai Z, Yang Y, Carbonell J, Salakhutdinov RR, et al. Xlnet: Generalized Autoregressive Pretraining for Language Understanding. NeurIPS. 33rd Conference on Neural Information Processing Systems. 2019.

[2] Balan R, Haghani N, Singh M. Permutation Invariant Representations With Applications to Graph Deep Learning. 2022. Arxiv preprint: https://arxiv.org/pdf/2203.07546

[3] Askr H, Elgeldawi E, Aboul Ella H, Elshaier YA, Gomaa MM, et al. Deep Learning in Drug Discovery: An Integrative Review and Future Challenges. Artif Intell Rev. 2023;56:5975-6037.

[4] Soloviev VP, Bielza C, Larrañaga P. Quantum Approximate Optimization Algorithm for Bayesian Network Structure Learning. Quantum Inf Process. 2023;22:19.

[5] Rahman MM, Jamil AR, Nower N. Uncertainty-aware traffic prediction using attention-based deep hybrid network with bayesian inference. Int J Adv Comput Sci Appl . 2023;14.

[6] Li Z, Lian Y, Ma X, Zhang X, Li C. Bio-Semantic Relation Extraction With Attention-Based External Knowledge Reinforcement. BMC Bioinformatics. 2020;21:213.

[7] Huang L, Chen Q, Lan W. Predicting Drug-Drug Interactions Based on Multi-View and Multichannel Attention Deep Learning. Health Inf Sci Syst. 2023;11:50.

[8] Shang Y, Li Y, Lin H, Yang Z. Enhancing Biomedical Text Summarization Using Semantic Relation Extraction. PLOS One. 2011;6:e23862.

[9] Peng J, Li J, Shang X. A Learning-Based Method for Drug-Target Interaction Prediction Based on Feature Representation Learning and Deep Neural Network. BMC Bioinformatics. 2020;21:394.

[10] Shi W, Peng D, Luo J, Chen G, Yang H, et al. A Review on Predicting Drug Target Interactions Based on Machine Learning. Lect Notes Comput Sci. 2023;14305:283-295.

[11] Abbasi Mesrabadi H, Faez K, Pirgazi J. Drug–Target Interaction Prediction Based on Protein Features Using Wrapper Feature Selection. Sci Rep. 2023;13:3594.

[12] Sankaran A, Sathiyamurthy K. Enhancing Biomedical Literature Analysis to Predict Drug Interactions Using Multi Criteria Decision Making and LSTM-CNN Model. International Conference on System, Computation, Automation and Networking ICSCAN. 2023:1-6.

[13] Deepika SS, Saranya M, Geetha TV. Cross-Corpus Training with CNN to Classify Imbalanced Biomedical Relation Data. In: Métais E, Meziane F, Vadera S, Sugumaran V, Saraee M. (eds). Natural Language Processing and Information Systems. NLDB 2019. Lecture Notes in Computer Science. Springer. 2019;11608.

[14] Wang Y, Yang Z, Yao Q. Accurate and Interpretable Drug-Drug Interaction Prediction Enable by Knowledge Subgraph Learning. Commun Med (Lond). 2024. Arxiv Preprint: https://arxiv.org/pdf/2311.15056

[15] Fan X, Zhang S, Chen B, Zhou M. Bayesian Attention Modules. 2010. Arxiv preprint: https://arxiv.org/pdf/2010.10604

[16] Du M, Zhong P, Cai X, Bi D, Jing A. Robust Bayesian Attention Belief Network for Radar Work Mode Recognition. Digit Signal Process. 2023;133:103874.

[17] Carrascal G, Botella G, Del Barrio A, Kremer D. A Bayesian-Network-Based Quantum Procedure for Failure Risk Analysis. EPJ Quantum Technol. 2023;10:13.

[18] Zhang S, Fan X, Chen B, Zhou M. Bayesian Attention Belief Networks. 2021. Arxiv preprint: https://arxiv.org/pdf/2106.05251

[19] Sathiyamurthy K, Kodavali L. Chapter Six – Bayesian Network-Based Quality Assessment of Blockchain Smart Contracts. Adv Comput. 2024;132:85-110.

[20] Kodavali L, Kuppuswamy S. Bayesian Network Based Vulnerability Detection of Blockchain Smart Contracts. J Intell Fuzzy Syst. 2023;44:1907-1920.

[21] Vaniš M, Lokaj Z, Šrotýř M. A Novel Algorithm for Merging Bayesian Networks. Symmetry. 2023;15:1461.

[22] Ke X, Keenan K, Smith VA. Treatment of Missing Data in Bayesian Network Structure Learning: An Application to Linked Biomedical and Social Survey Data. BMC Med Res Methodol. 2022;22:326.

[23] Wang J, Zhang S, Li R, Chen G, Yan S, et al. Multi-View Feature Representation and Fusion for Drug-Drug Interactions Prediction. BMC Bioinformatics. 2023;24:93.

[24] Huang L, Hou J, Chen Y, Yi J, Hu J. Text Mining and Analysis Based on XLNet-BiLSTM Model. In 2023 International Conference on Intelligent Media, Big Data and Knowledge Mining (IMBDKM). IEEE. 2023:57-61.

[25] Lee Y, Son J, Song M. BertSRC: Transformer-Based Semantic Relation Classification. BMC Med Inform Decis Mak. 2022;22:234.

[26] Cao L, Sun J, Cross A. An Automatic and End-To-End System for Rare Disease Knowledge Graph Construction Based on Ontology-Enhanced Large Language Models: Development Study. JMIR Med Inform. 2024;12:e60665.

[27] Inoue Y, Lee H, Fu T, Luna A. drGAT: Attention-Guided Gene Assessment of Drug Response Utilizing a Drug-Cell-Gene Heterogeneous Network. 2024. Arxiv Preprint: https://arxiv.org/pdf/2405.08979

[28] https://github.com/sankaran99/XLNet

[29] https://www.kaggle.com/datasets/shudhanshusingh/250k-medicines-usage-side-effects-and-substitutes

[30] https://www.kaggle.com/datasets/maadco/herbal-medicines

[31] https://zenodo.org/records/8192515

[32] https://www.kaggle.com/code/tanmaydacha/drug-target-interaction-predictor-project

[33] https://www.kaggle.com/datasets/mghobashy/drug-drug-interactions.