Artificial Intelligence and Robotics Research 人工智能与机器人研究, 2026, 15(2), 538-547
Published Online March 2026 in Hans. https://www.hanspub.org/journal/airr
https://doi.org/10.12677/airr.2026.152052

Hans汉斯

一种基于机器学习的量子随机数预测方法

韩　宇*，费洋扬

河南省网络密码技术重点实验室，河南　郑州

收稿日期：2026年1月4日；录用日期：2026年3月9日；发布日期：2026年3月19日

摘　要

随机性是密码学等领域的关键资源，量子随机数被视为真随机性生成的重要标准，但现有NIST统计测试难以全面评估其不可预测性与脆弱性。本文提出一种基于机器学习的量子随机数预测方法，以激光相位噪声方案的量子随机数为研究对象，通过构建数据集、选取适配的全连接网络模型，利用历史随机序列预测未来输出，以预测概率量化其熵值上限与脆弱性关键因素。该方法可补充传统统计测试，为随机数安全性分析提供新工具，适用于密码学场景适配与设备优化。

关键词

量子随机数，机器学习，随机数预测方法

Leveraging Machine Learning for Quantum Random Number Prediction

Yu Han*, Yangyang Fei

Henan Key Laboratory of Network Cryptography Technology, Zhengzhou Henan

Received: January 4, 2026; accepted: March 9, 2026; published: March 19, 2026

Abstract

Randomness is a vital resource in fields such as cryptography, and quantum random numbers are widely recognized as the standard for generating true randomness. However, existing NIST statistical tests fail to fully assess their unpredictability and vulnerability. This article proposes a QRN prediction method based on machine learning, focusing on QRNs generated via the laser phase noise scheme. By constructing a dedicated dataset, selecting a suitable fully connected network model, and leveraging historical random sequences to predict future outputs, the upper bound of entropy and key vulnerability factors are quantified by means of prediction probability. This method serves

*通讯作者。

文章引用: 韩宇，费洋扬. 一种基于机器学习的量子随机数预测方法[J]. 人工智能与机器人研究, 2026, 15(2): 538-547.
DOI: 10.12677/airr.2026.152052

as a complement to traditional statistical tests, offers a novel tool for random number security analysis, and is well-suited for cryptographic scenario adaptation and device optimization.

Keywords

Quantum Random Numbers, Machine Learning, Random Number Prediction Method

1. 引言

1.1. 研究背景

随机性在基础科学和实际应用中均扮演着关键角色[1]。理论层面，内在随机性是否存在的问题引发了关于人类对自然界认知的无尽争论[2]。实际应用中，随机性是人类活动不可或缺的资源。特别是在密码学应用等信息处理任务中，对“优质”随机数的需求极为迫切——这类随机数不仅需满足统计均匀性和可验证隐私性，还必须不能被有效预测。即为实现最高级别的安全性，随机数应具备信息论意义上的隐私性，即任何观察者均无法对其进行预测[3]。量子随机数(Quantum Random Number, QRN)依赖激光相位噪声等量子随机过程，被广泛认为是生成真随机性的关键标准[4]。然而，对于量子随机数的评估仍面临难题：经典噪声污染和不当的后处理过程也可能引入相关性导致不可预测性的破坏，而对这些相关性进行精准量化仍是一项艰巨任务。

考虑到真随机性对于理解一些自然的本质规律至关重要，尤其在信息处理等任务中更是不可或缺的，因此在实践中评估真随机数质量是一项十分重要的工作内容，特别是在需要统计一致性和不可预测性的密码应用中。美国国家标准局开发的 NIST 统计估值器[5]用于测试生成随机数的统计特性，量子信息的最新进展也为评估具有良好特征的物理模型的随机性提供了手段。然而，统计测试受到计算机复杂性的极大限制，考虑到计算速度和计算资源的制约，终究无法穷尽所有的统计规律，而只能是提炼出最为精准的若干条统计规律来进行普遍性测试；同时很多物理模型也往往依赖于无法证明的假设，且其理论模型与实际情况的匹配程度也无法达到圆满的程度。

而现有的量子随机数发生器研究主要关注于如何进行有效设计与实现随机数生成，追求更高的量子随机数生成码率或者探索量子随机数发生器设备的小型化芯片化可能，其安全性分析的主流方法仅为对该量子随机数发生器生成的最终随机数据进行统计性测试，根据统计性测试结果来判断此类方法生成的随机数是否安全，而完全忽略了不可预测性指标，这无疑存在很大的风险隐患。近年来的安全性分析文章[6]表明，相关研究人员也对目前的安全性分析现状有所忧虑，但苦于无法找到除了 NIST 统计性测试之外的更有效的安全性评价工具。

1.2. 研究意义

目前在量子随机数研究领域特别是量子随机数实际安全性或脆弱性研究领域，一种重要的现实问题即，那些已经通过美国 NIST SP 800-22 统计集测试的量子随机数是否就真的是安全的，攻击者是否可以利用未被完全消除的部分经典信息实现某类量子黑客攻击行为，或者攻击者是否可以通过注入或者放大某类经典噪声而实现对量子随机数有效获取的目的。因此，亟需一种区别于 NIST 统计集测试或能有效

补充其不足的随机数脆弱性评估方法，这一问题具有重要研究价值。对于真随机数不可预测性的研究尤为重要，尝试建立一个区别于统计性检验机制的量子随机性预测方法以对量子随机数的脆弱性进行定量分析是尤为必要的。

近十年来，机器学习理论研究与实际部署取得了飞跃式的进展，尤其在计算机视觉和图像识别、自然语言处理、语音识别、预测分析和专家系统、知识图谱等领域全方位加深了人类认识世界、处理数据的能力[7]。因此考虑将机器学习的方法引入对量子随机数脆弱性的评估中来，通过利用机器学习算法对不同原理不同类型的量子随机数进行预测攻击尝试，实现对量子随机数脆弱性所关联的关键因素的锚定和分析。具体而言，提出了一种用于随机性评估的机器学习预测方法，利用量子熵源生成的随机序列的先前结果，设定相应的神经网络模型来预测未来的输出，预测结果提供了输出熵的上界并为物理随机性评估模型提供依据。该方法适用于非均匀统计，并可适用于量子随机数后处理过程中随机性提取协议的分析。同时，建立量子随机数的安全性分析模型并尝试利用机器学习技术对其进行预测攻击有利于更加清晰地度量量子随机数的安全强度并剖析潜在问题，以便进一步优化量子随机数发生器设备的生产参数或对使用不安全量子随机数的设备进行渗透尝试。

2. 量子随机数发生器预测分析的发展现状

随着量子信息技术的不断发展，量子随机数发生器的生成方案不断更新完善，其最终随机数成码率也达到 Gbps 乃至数百 Gbps 的量级，足以支撑信息安全领域算法与协议的随机数源码率需求。与之相对应的是关于随机数的经典统计性检测方法和标准是相对固定的，同时基于机器学习预测工具的量子随机数脆弱性分析也成为近年来的热点研究问题。

近两年来，得益于 GPU 等硬件的发展和机器学习技术的溢出效应，采用机器学习技术对随机数序列进行预测测试的方案相继出现。此类方法的主要手段在于构造以某类神经网络模型为核心的神经网络对随机数序列进行预测，将预测概率与统计概率(或者分布概率，结合具体的熵模型)进行对比；如果预测概率显著高于统计概率，则认为随机数序列无法通过该机器学习技术测试，或者表述为利用机器学习技术实现对随机数序列的有效预测。而部分被有效预测的序列能通过 NIST 测试，即表明基于机器学习构造的测试方法具有在精确度和严格性上超越传统 NIST 随机数统计测试方法的可能性。但同时该方法与机器学习技术的有效使用(网络模型、参数)密切相关，这也是重要问题所在。

上述机器学习的方法也可用于对量子随机数发生器产生的量子随机数序列进行测试。测试方法可以分为两个方面：一个是对最终生成的量子随机数序列进行测试，这可以用来评估最终量子随机数成码的质量；另一个是对整个量子随机数产生过程中每个阶段的数据的随机性进行测试，这可以用来评估不同阶段背景噪声(一般认为是经典熵)对量子熵的影响程度，主要原因在于不可避免且无法精确分离的经典熵可能会引入较多的关联且相关模式可以被深度学习方法捕获，可以用来指导对量子随机数每一步具体操作(如仪器精度、相位、混频电压、滤波设置等)的改进，也可以通过控制变量的方法评估一部分背景噪声(温度、电压、强度等)所引入关联性的大小(如果关联性较大的话，对于某些 QRNG 即可以实施有效攻击)。因此，基于机器学习构造的测试方法成为一种新兴的评估量子随机数性质的有效工具。

使用神经网络的方法对随机数的不可预测性研究最早可以追溯至 2003 年，英国学者 D.A. Karras 和 V. Zorkadis 等人于 2003 年发表题为《Improving Pseudorandom Bit Sequence Generation and Evaluation for Secure Internet Communications Using Neural Network Techniques》[8]文章中介绍了一种基于多层感知器(Multilayer Perceptron, MLP)的针对随机比特序列的预测性检验。文中将随机比特序列视为时间序列，则通过用确定长度的滑动窗口进行扫描，从而形成一系列适合于定义 MLP 训练任务的训练集，此类训练集包含某一长度比特的样本及下一个输出比特，即代表根据已知的若干个比特对下一个比特进行预测，尝

试根据过去和现在的数字对未来的数字进行预测。该方法认为，如果此类任务可以被 MLP 所学习，则待预测的数字很有可能被预测出来。该篇文章虽然无法实现较好的预测效果，但给出了基于神经网络进行随机数预测的方法雏形。

2018 年，澳大利亚悉尼大学的 Nhan Duy Truong 等人在《IEEE Transactions on Information Forensics and Security》发表题为《Machine Learning Cryptanalysis of a Quantum Random Number Generator》[9]的文章，作者等人开发了一种机器学习预测分析方法，以研究确定性经典噪声在光学连续变量量子随机数发生器的不同阶段中的影响，当确定性噪声源突出时，其机器学习模型成功地检测到固有的相关性；在进行适当的信号滤波和随机性后处理后，其机器学习模型无法对量子随机数发生器系统进行有效预测以此证明量子随机数发生器的鲁棒性;最后利用其机器学习模型对线性同余伪随机数进行预测并在周期为 228 时取得了略微超过理论猜测值得预测结果，以此表明机器学习在对量子随机数设备质量进行基准测试方面具有潜力。

其方法的主要思路仍在于使用机器学习技术来分析以前的输出，以猜测由随机数发生器生成的下一个输出，主要利用循环卷积神经网络(RCNN)用于学习在量子随机数发生器生成随机数的不同阶段中，长序列的生成数字之间可能存在的潜在关联性与模式。文中认为量子随机数发生器可以分为两个部分：熵源和后处理程序。熵源产生原始随机性，作为源的随机物理过程的直接结果。原始随机性包含量子和经典起源的不确定性。经典熵包括来自经典设备的噪声，例如外围测量设备和 ADC。为了提取内在随机性，必须从主随机比特流中去除可能不可信或有偏差的经典熵。

然而该文章虽然声称基于机器学习技术实现了对量子随机数的密码学分析，仅仅是在量子随机数的生成过程中捕捉到了一定的关联性，并无法对最终生成的量子随机数进行有效预测。但同时，该文章利用其 RCNN 模型对经典的线性同余随机数发生器进行预测，并在周期为 28 时以不可忽略的概率对随机数进行了预测，周期为 28 的线性同余随机数已可以确保完全通过 NIST 统计性测试，所以其声称在此场景内实现了比 NIST 更好的随机性检测结果。但其结果有待商榷，因为其所使用的测试集规模已经超出了周期为 28 的线性同余随机数的周期。

2020 年 4 月，加拿大滑铁卢大学的 Vadim Makarov 等人对 IDQ 公司生产的量子随机数发生器商用产品的硬件和固件进行了逆向工程、测试和分析[10]，结果表明其输出的随机数主要来源于物理随机过程，研究了具体电子元件对不可预测性造成的影响。其结果表明此类光学量子光学随机数发生器的质量可以得到初步保证，最终生成的量子随机数质量较高且是较为安全的，但是探测器电子元件等对不可预测性造成的影响有待进一步评估。

2022 年，德国杜塞尔多夫大学的 Sarnava Datta 等人利用机器学习技术开展对贝尔测试中猜测概率的估计[11]，提出了基于深度学习模型猜测概率估计新方法，能够有效绕过计算复杂和繁琐的半确定优化过程，比使用传统求解器速度要快，显示了机器学习在估计猜测概率和理解量子非局域性方面的能力。

3. 一种基于机器学习的量子随机数预测方法

内在随机性是量子理论的必要条件，这为基本量子测试提供了一种不同的方法。本文使用量子随机熵源的历史输出预测其未来输出，采用一种带有神经网络的机器学习算法来学习数据中的结构和潜在的隐藏模式而非简单的统计数据分析。此类方法可以广泛应用于实际应用，比如面向量子随机数预测的密码分析，以及测试商业量子随机数的质量。下面以基于激光相位噪声方案的量子随机数[12]-[14]为例，阐述该方案的基本思路和实现方法。

3.1. 基于激光相位噪声方案的量子随机数发生器基本原理

基于激光相位噪声原理的量子随机数发生器是目前速度较快的生成方案。该方案通过将激光器中的

自发辐射引起的相位涨落提取出来，通过干涉的方法转化为光强并测量来产生随机数。在经典的量子光学模型中，光子在受激辐射作用下具有固定的相位，但自发辐射产生的光子具有随机的初始相位，因此自发辐射光子的总相位一直随着时间波动，可以认为是一种量子随机源。

激光相位噪声量子随机数发生器的工作流程如下：连续波激光器发出的激光经过干涉仪，其相位涨落被转化为强度涨落；光电探测器测出强度涨落并输出一个相应的电信号；模数转换器将涨落的电信号离散化并转化为随机数；后处理过程将原始数据转化为无偏且均匀分布的随机数。在实验中，后处理过程分为两类：离线后处理，将产生的随机数原始数据收集起来，一起通过随机性提取器；实时后处理，原始数据实时地被转化为最终的随机数。

激光相位噪声量子随机数发生器的随机性来自于激光器内部介质的自发辐射带来的随机相位涨落。激光器内部工作介质的原子能级跃迁会产生两种辐射，受激辐射和自发辐射。后者由真空涨落引起，发射出的光子相位具有高度的随机性。在相位涨落量子随机数发生器中，整个装置将自发辐射的随机相位提取出来产生随机数。

3.2. 最小熵与机器学习预测概率

在高斯分布假设下，激光相位噪声量子随机数发生器的采样器的采样电压近似符合高斯分布。假设在实验中取多个样本，每个样本给出一个 256 位结果。在样本数量接近无穷大的渐近极限下，每个样本的条件最小熵的下限为分布的最高点。假设一个给定的信息源发出的信息能够用一个随机变量 $\mathcal{X}=\{x_1,x_2,\cdots,x_n\}$ 描述，香农熵定义为

$$\mathcal{H}(\mathcal{X})=-\sum p_i \log_2 p_i \tag{1}$$

其中 p_i 是 $\mathcal{X}$ 的取值为 x_i 的概率，即 $\mathcal{P}(\mathcal{X}=x_i)=p_i$。

最小熵刻画的是攻击者完全掌握随机序列统计分布时的最坏情况，此时攻击者能获取的信息量为所有熵度量中的上限，可作为密码学中随机性的安全评估指标，被定义为

$$\mathcal{H}_{\min}=-\log_2\left[\max \mathcal{P}(x_i)\right] \tag{2}$$

也即攻击者在已知量子随机数原始数据统计分布的情况下，能够获得明确随机数信息的最大概率。

由上文已知，攻击者盲猜成功的最大概率 $\mathcal{P}_{\text{guessing}}$ 即为量子随机数原始数据统计分布中最高点所对应的概率，即

$$\mathcal{P}_{\text{guessing}}=\max \mathcal{P}(x_i) \tag{3}$$

我们定义概率 $\mathcal{P}_{\text{ML}}$ 为利用机器方法对量子随机数原始数据预测正确的概率

$$\mathcal{P}_{\text{ML}}\equiv\mathcal{P}\left(\left(\mathcal{X}_j\middle|\mathcal{X}_h\right)\right) \tag{4}$$

其中 $\mathcal{X}_h$ 代表已经公布的历史随机数，$\mathcal{X}_j$ 代表下一位要预测的随机数。

3.3. 机器学习数据准备及基本流程

基于激光相位噪声方案的量子随机数原始数据是可以根据需要产生的。为了避免机器学习的训练时间过长，实验设计上原始数据为 8 位二进制的 1 GB 激光相位噪声采样数据，其中 500 MB 作为训练集，500 MB 作为测试集并且平均分为 5 个 100 MB 的测试单元。训练集中，定义选取 $\mathcal{M}$ 长的 $\mathcal{K}$ 位数 $\mathcal{X}_h$ 作为历史数据的输入，下一位 $\mathcal{K}$ 位数 $\mathcal{X}_j$ 作为所对应的标签，同时为了保证随机数据的选择顺次进行，定义每次间隔 $\mathcal{L}$ 的长度向后滑动以确定每次 $\mathcal{M}$ 长的 $\mathcal{K}$ 位数 $\mathcal{X}_h$ 的起始位置，如图 1 所示。

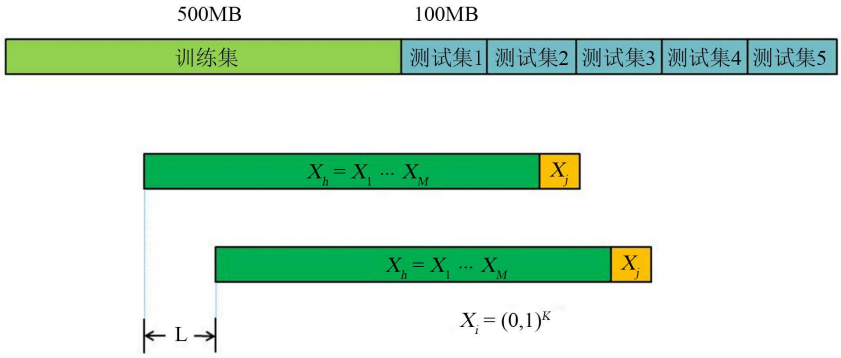

Figure 1. Schematic of ML dataset setup for quantum random number predictive analysis
图 1. 面向量子随机数预测分析的机器学习数据集设置示意图

下面形式化并明确量子随机数的预测过程：

首先，从量子随机源中收集量子随机数序列样本，然后用历史数据序列长度为 5 的窗口，构造机器学习预测的数据集，每 100 个 8 位二进制数作为输入，下一位 8 位二进制数作为预测目标。

然后，确定机器学习模型，可以是线性预测器、SVM、神经网络等，以便当随机数二进制时可以用更紧凑的方式进行表示。为确保模型学习到数据内在规律而非机械记忆，需将数据划分为互不重叠的训练集与测试集。

最后，使用训练数据对模型进行训练而至收敛，并使用测试数据对其进行测试以评估预测的准确性；最后通过对测试集数据的预测，我们可以得出数据中包含的随机性的下限，也即机器方法对量子随机数原始数据预测正确的概率 $\mathcal{P}_{\mathrm{ML}} \equiv \mathcal{P}\left(\left(\mathcal{X}_j \middle| \mathcal{X}_h\right)\right)$。

3.4. 机器学习网络结构选择与比较

对于通用预测器，目前所有通用的强大网络结构如 CNN、LSTM 和 Transformer 等，这些结构在随机数预测任务上表现出明显的领域偏差，不适合于随机数预测任务。可以尝试全连接网络结构，因为它包含最轻微的偏差，并且在捕获复杂相关性方面非常强大。

CNN (卷积神经网络)是一类前馈神经网络，包含卷积计算特点并具有深度结构特征。大致分为卷积层、池化层、全连接层三个部分，通过仿造生物的视觉感知机制来构架模型，核心特点体现在范围内的特征提取和代表值输出。然而对于随机数预测任务来说，其相邻或局部范围内的随机数数据并不存在如图像数据般的显性特征关联与空间关联性，因此卷积神经网络结构很难捕捉到这种非范围内紧密联系的没有特殊特征的关联，因此采用经典的卷积神经网络模型适用于图像识别和视觉处理等领域，但无法获得较好的随机数预测结果。

LSTM (长短时记忆网络)是一种特殊的循环神经网络(RNN)，其内部的神经元结构是一个神经网络块，除了包括一部分上一层神经网络的输出结果之外，还包含一部分本层神经元结构的状态信息，其状态信息在网络中循环传递，并通过遗忘门函数的设计使得在训练时能够控制梯度的收敛性，从而缓解梯度消失或者爆炸的问题并能保持长期的记忆性，因此适用于语音处理、机器翻译以及周期性时间序列预测任务等。但是对于量子随机数特别是实际生成的量子随机数来说，它虽然也是时间序列预测任务，但是其由于影响因素的不确定性，它无法呈现类似于机票淡旺季周期、股票周期或者期货周期的周期性时间序列特征，适用于大致确定性周期时间序列预测任务的 LSTM 神经网络结构也无法取得较好的随机数预测结果。

Transformer 结构于 2018 年由谷歌团队提出，用于生成词向量的 BERT 算法，该结构使用了注意力

机制，将序列中的任意两个位置之间的距离缩小，同时并非采用顺序结构从而具有更好的并行性，同时通过 Query 向量、Key 向量和 Value 向量赋予不同维度的信息来实现更好的信息检索和匹配，并通过残差网络结构来解决深度学习退化问题，因此非常适用于解决自然语言处理的机器翻译领域任务。但是随机数序列仅由 0 和 1 组成，无语义信息与上下文关联，而 Transformer 的注意力机制依赖于序列的语义关联与特征差异，因此其多维度信息检索与匹配能力在无特征的随机序列中无法发挥作用，因此 Transformer 结构也无法取得较好的随机数预测结果。

3.5. 全连接网络模拟结构与环境

本研究提供了全连接网络的初步模拟结构，设计为 4 层结构化网络，包含输入层、2 个隐藏层和输出层，隐藏层均采用 ReLU 激活函数缓解梯度消失问题，输出层采用 Softmax 激活函数输出概率分布，具体拓扑结构如下：

输入层：神经元数量 = 历史窗口长度 × 单段位宽 = 40，对应 5 个 8 位二进制数的历史输入，无激活函数；隐藏层 1：神经元数量 = 128，激活函数 = ReLU，提升非线性拟合能力；隐藏层 2：神经元数量 = 64，激活函数 = ReLU，提取数据高阶非线性特征；输出层：神经元数量 = 256，激活函数 = Softmax，对应 8 位二进制数的所有可能输出(2^8 = 256)，输出各结果的预测概率；网络损失函数采用交叉熵损失(Cross-Entropy Loss)，适配多分类预测任务；优化器采用 Adam，兼具自适应学习率和动量特性，保证训练收敛速度与稳定性。

实验基于通用 GPU 工作站开展，软件环境依托 Python 深度学习生态构建，无定制化组件，具体配置如下。硬件环境主要包括：处理器，Intel Core i9-13900K (3.0 GHz，24 核)；显卡，NVIDIA GeForce RTX 4090 (24 GB GDDR6X)；内存，64 GB DDR5 5600 MHz；存储，1 TB NVMe SSD (用于数据存储和模型缓存)。软件库主要包括：编程语言，Python 3.9.18；深度学习框架，PyTorch 2.1.0 (CUDA 12.1 加速)；数据处理，NumPy 1.26.0，Pandas 2.1.1；可视化，Matplotlib 3.8.0，Seaborn 0.12.2；统计测试，NIST SP 800-22 Test Suite 2.1.0。

所有超参数经多轮实验优化确定最优值，同时引入早停、L2 正则化等策略防止过拟合，具体设置如下：批次大小(Batch Size)为 256，用于平衡训练效率与拟合稳定性；学习率(Learning Rate)为 10^{-4}，用于控制参数更新步长，避免震荡；训练轮数(Epoch)为 100，用于保证模型充分学习；权重衰减(Weight Decay)为 10^{-6}，用于 L2 正则化，抑制过拟合；丢弃率(Dropout)为 0.2，用于随机丢弃神经元，提升泛化性；梯度裁剪(Gradient Clipping)为 1.0，用于防止梯度爆炸；初始化方法为 He 初始化，用于适配 ReLU 激活函数，提升收敛性；模型训练中采用早停(Early Stopping)策略，验证集损失连续 10 轮未下降则停止训练。

4. 核心经典噪声源作用机制分析

经典噪声是影响激光相位噪声 QRNG 输出随机数安全性的关键因素，其引入的时序相关性虽无法通过 NIST SP 800-22 统计测试识别，却能被机器学习模型捕获，进而降低量子随机数的实际不可预测性。下面围绕各噪声源的影响路径，系统分析 QRNG ADC 采样原始数据中核心经典噪声源的来源、本质特征及作用机制，为提升设备性能、增强量子随机数实际安全强度提供坚实的物理依据和技术支撑。

4.1. 电子学噪声作用机制

电子学噪声是影响量子随机数原始数据相关性的关键经典噪声源，其来源具有多样性，主要集中在量子随机数生成链路中的信号转换与放大环节，核心包括热噪声、散粒噪声和放大器固有噪声三类，三类噪声相互叠加，共同引入时序相关性。具体来看，热噪声主要产生于光电探测器的半导体材料和信号

放大器的晶体管内部，是由载流子的热运动引发的随机电信号波动，其强度与环境温度正相关，温度越高，载流子热运动越剧烈，热噪声的幅值越大；散粒噪声则源于光电探测器的光－电转换过程，光子撞击探测器光敏面产生光电子时，光电子的发射数量和发射时间具有随机性，导致输出电信号出现离散性波动，这类噪声虽与量子光信号的随机性存在本质区别，但会与激光相位噪声产生的量子涨落信号叠加，干扰原始数据的纯度；放大器固有噪声是信号放大器自身工作过程中产生的噪声，包括电流噪声、电压噪声等，源于放大器内部晶体管的非线性特性和电路寄生参数，会在信号放大过程中同步放大，进一步增强噪声的影响。

其作用机制具体为：激光相位噪声产生的量子光强涨落信号经干涉仪转化为光强信号后，进入光电探测器完成光－电转换，在此过程中，热噪声、散粒噪声与光－电转换后的电信号直接叠加；随后，叠加了噪声的电信号进入信号放大器进行幅值放大，放大器固有噪声进一步与信号叠加，导致最终输出至ADC 的电信号中包含大量经典噪声成分。更为关键的是，这类电子学噪声具有显著的短期时序相关性——其相关性主要由载流子寿命和放大器带宽决定：载流子在半导体材料中的寿命有限，相邻时刻的载流子运动存在关联性，导致热噪声与散粒噪声的波动并非完全独立随机，而是呈现出短期的连续变化特征；放大器的带宽决定了其能够处理的信号频率范围，超出带宽的噪声信号会被滤波，但带宽范围内的噪声信号会保持连续的时序关联，最终导致 ADC 采样得到的原始数据中存在连续的噪声关联，这种关联虽微弱，但被全连接网络模型精准捕获，成为模型提升预测准确率的核心依据。

4.2. 电源纹波作用机制

电源纹波是一类系统性经典噪声，区别于电子学噪声的随机性叠加，其具有明显的周期性特征，主要来源于为 QRNG 整个硬件系统供电的直流电源，是直流电源输出电压中夹杂的周期性交流分量，其幅值通常较小(一般为毫伏级)，但对精密电子设备和量子信号检测的影响不可忽视。QRNG 硬件系统中，激光器、光电探测器、ADC、信号放大器等核心器件均需要稳定的直流电源供电，而实际应用中的直流电源无法输出绝对纯净的直流电压，受电源内部整流电路、滤波电容性能、供电线路阻抗等因素影响，输出电压会在额定直流电压附近呈现周期性波动，形成电源纹波。电源纹波的频率通常与电网频率(50 Hz)或电源内部开关频率(kHz 至 MHz 级)相关，呈现固定的周期性，属于可预测的系统性噪声。

其作用机制具体为：电源纹波的周期性波动会同步影响 QRNG 各核心器件的工作状态，进而导致ADC 采样数据引入周期性时序相关性。对于激光器而言，电源纹波会导致其驱动电流出现周期性波动，进而影响激光器的输出功率和相位稳定性，使得激光相位涨落的检测出现周期性偏差；对于光电探测器，电源纹波会影响其偏置电压的稳定性，导致探测器的增益出现周期性变化，使得光－电转换的效率随纹波周期波动，输出电信号的幅值也随之呈现周期性变化；对于 ADC，电源纹波会影响其采样基准电压的稳定性，导致采样精度出现周期性偏差，使得采样得到的原始数据在纹波周期内呈现规律的波动趋势。尽管这类周期性相关性的幅值微弱，且被量子随机信号的涨落部分掩盖，无法通过 NIST SP 800-22 统计测试识别，但全连接网络模型能够通过非线性拟合，挖掘出这种隐藏的周期性关联，进而提升对未来随机数的预测成功率。

4.3. 温度漂移作用机制

温度漂移是一类慢变经典噪声，其影响具有累积性和缓慢性，主要来源于 QRNG 硬件系统的内部发热和外部环境温度变化，核心影响对象为激光器、干涉仪等对温度敏感的器件，其产生的时序相关性呈现慢变特征，与电子学噪声的短期关联、电源纹波的周期性关联形成明显区别。QRNG 工作过程中，激光器、信号放大器等器件会持续发热，导致系统内部温度逐渐升高；同时，外部环境温度的波动(如实验

室环境温度变化、设备散热不畅等)也会导致系统整体温度发生变化，这种温度变化虽缓慢(通常为每分钟0.1℃至0.5℃)，但会对核心器件的工作参数产生显著影响。其中，激光器的谐振腔长度对温度极为敏感，温度每变化1℃，谐振腔长度会产生微小形变(约10^{-6}量级)；干涉仪的光程差也会随温度变化而改变，因为干涉仪的光学元件(如镜片、光程臂)会随温度膨胀或收缩，导致光程差出现缓慢波动。

其作用机制具体为：温度变化导致激光器谐振腔长度形变后，会影响激光的输出波长和相位稳定性，使得激光相位涨落的随机性受到干扰，引入慢变的系统性偏差，该偏差引起光强涨落的检测基准发生缓慢变化，使得光电探测器输出的电信号呈现长期缓慢的趋势性波动。这种慢变的波动会导致ADC采样的原始数据在较长时间尺度上(如几分钟至几十分钟)呈现微弱的趋势性，形成慢变时序相关性——与电子学噪声的短期关联不同，这种相关性的变化速度缓慢，无法通过短期数据的统计分析发现，却能被训练后的全连接网络模型通过非线性拟合捕获，模型通过学习长期数据中的慢变趋势，进一步提升预测准确率。此外，温度漂移的累积效应会导致噪声相关性逐渐增强，长期运行后，这种慢变关联会更加明显，对量子随机数的不可预测性产生持续影响。

5. 结语

本文聚焦量子随机数的不可预测性与脆弱性评估问题，提出一种基于机器学习的预测方法。以激光相位噪声方案生成的量子随机数为研究对象，通过构建专用数据集、选取适配的全连接网络模型，利用历史随机序列预测未来输出，以预测概率量化熵值上限与脆弱性关键因素。该方法有效补充NIST统计测试，为量子随机数安全性分析提供了全新工具，可支撑密码学场景适配与发生器设备优化，为真随机数质量评估提供了非统计检验的新思路。

未来可从多维度深化研究：一方面，优化机器学习模型泛化能力，拓展至真空涨落、贝尔不等式违背等其他量子随机数生成方案，提升对不同类型随机数的脆弱性评估精度，同时探索数据质量与成码率的动态平衡策略。另一方面，强化跨场景适配，将该预测方法与金融加密、量子通信等领域的安全需求深度绑定，完善移动端后处理程序的实时性与低功耗设计。此外，需推动行业标准化建设，联合密码学与量子光学领域力量，建立机器学习驱动的量子随机数安全评估体系，加速量子随机数发生器的芯片化集成与抗量子密码算法的融合应用，助力其在政务数据安全、工业互联网等关键领域实现规模化部署。

致　谢

感谢李元昊等人提供的讨论与帮助。

参考文献

[1] Zhang, J., Zhang, Y., Zheng, Z., Chen, Z., Xu, B. and Yu, S. (2021) Finite-Size Analysis of Continuous Variable Source-Independent Quantum Random Number Generation. *Quantum Information Processing*, **20**, Article No. 15. https://doi.org/10.1007/s11128-020-02936-7

[2] Zhou, H., Yuan, X. and Ma, X. (2015) Randomness Generation Based on Spontaneous Emissions of Lasers. *Physical Review A*, **91**, Article 062316. https://doi.org/10.1103/physreva.91.062316

[3] Fei, X., Yin, Z., Cui, C., Huang, W., Xu, B., Wang, S., *et al*. (2018) Optimality of Quantum Randomness Certification with Independent Devices. *Journal of the Optical Society of America B*, **35**, Article 2186. https://doi.org/10.1364/josab.35.002186

[4] Michel, T., Haw, J.Y., Marangon, D.G., Thearle, O., Vallone, G., Villoresi, P., *et al*. (2019) Real-Time Source-Independent Quantum Random-Number Generator with Squeezed States. *Physical Review Applied*, **12**, Article 034017. https://doi.org/10.1103/physrevapplied.12.034017

[5] Rukhin, A., Soto, J., Nechvatal, J., *et al.* (2010) SP 800-22 Rev.1a: A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications. NIST Special Publication, National Institute of Standards and Technology, 1-105.

[6] Barrett, J. and Gisin, N. (2011) How Much Measurement Independence Is Needed to Demonstrate Nonlocality? *Physical Review Letters*, **106**, Article 100406. https://doi.org/10.1103/physrevlett.106.100406

[7] Wang, H., Fu, T., Du, Y., Gao, W., Huang, K., Liu, Z., *et al*. (2023) Scientific Discovery in the Age of Artificial Intelligence. *Nature*, **620**, 47-60. https://doi.org/10.1038/s41586-023-06221-2

[8] Karras, D.A. and Zorkadis, V. (2003) Improving Pseudorandom Bit Sequence Generation and Evaluation for Secure Internet Communications Using Neural Network Techniques. *International Joint Conference on Neural Networks*, Portland, 1367-1372. https://doi.org/10.1109/IJCNN.2003.1223895

[9] Truong, N.D., Haw, J.Y., Assad, S.M., Lam, P.K. and Kavehei, O. (2019) Machine Learning Cryptanalysis of a Quantum Random Number Generator. *IEEE Transactions on Information Forensics and Security*, **14**, 403-414. https://doi.org/10.1109/tifs.2018.2850770

[10] Petrov, M., Radchenko, I., Steiger, D., *et al*. (2020) Independent Security Analysis of a Commercial Quantum Random Number Generator. arXiv.2004.04996.

[11] Datta, S., Kampermann, H. and Bruß, D. (2020) Upper Bound on the Guessing Probability Using Machine Learning. arXiv.2212.08500.

[12] Lei, W., Xie, Z., Li, Y., Fang, J. and Shen, W. (2020) An 8.4 Gbps Real-Time Quantum Random Number Generator Based on Quantum Phase Fluctuation. *Quantum Information Processing*, **19**, Article No. 405. https://doi.org/10.1007/s11128-020-02896-y

[13] Nie, Y., Huang, L., Liu, Y., Payne, F., Zhang, J. and Pan, J. (2015) The Generation of 68 Gbps Quantum Random Number by Measuring Laser Phase Fluctuations. *Review of Scientific Instruments*, **86**, Article No. 2435. https://doi.org/10.1063/1.4922417

[14] Zhang, X., Nie, Y., Zhou, H., Liang, H., Ma, X., Zhang, J., *et al*. (2016) Note: Fully Integrated 3.2 Gbps Quantum Random Number Generator with Real-Time Extraction. *Review of Scientific Instruments*, **87**, Article No. 2435. https://doi.org/10.1063/1.4958663