SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

An Intelligent Deep Transfer Learning Framework with Domain Adaptation for Cross-Cell Line CRISPR-Cas9 Guide RNA Efficiency Prediction and Therapeutic Target Prioritization

[1]R. Sulakshana, [2]Charan M.B.B.S, [3]Dr. R. Lakshmi

[1]Research Scholar, Dept. of Computer Science, Pondicherry University (Karaikal Campus), India. E-mail: sulopragash@hotmail.com
[2]Pondicherry University, India.
[3]Associate Professor, Dept. of Computer Science, Pondicherry University (Karaikal Campus), India.

ABSTRACT
CRISPR/Cas9 has become an important tool for targeted genome editing, and the efficiency of the guide RNA (gRNA) is one of the factors that affects the outcome of gene editing experiments. However, gRNAs designed for the same target can show considerable variation in their editing activity. Testing a large number of candidate guides experimentally can therefore require considerable time and resources. In this work, a hybrid computational framework is proposed to predict gRNA editing efficiency and to identify promising guide sequences under different cellular conditions. The proposed framework uses one-hot encoded gRNA sequences together with sequence-derived characteristics, including GC content and k-mer frequencies. A CNN–BiLSTM network with Multi-Head Attention is used to learn relevant patterns from the input sequences. The features obtained from this network are combined with the sequence-based features, and Recursive Feature Elimination (RFE) is applied to reduce redundant features. XGBoost regression is then used to estimate gRNA editing efficiency. Since gRNA performance can vary between cell lines, the framework also considers cross-cell-line prediction. Transfer learning and domain adaptation are introduced using a Domain-Adversarial Neural Network (DANN) and CORAL to reduce differences between source and target datasets. Prediction uncertainty is estimated using Monte Carlo Dropout, which provides an indication of the confidence associated with the predicted efficiency values. The framework is evaluated using gRNA datasets from HEK293T, HeLa, and HL60 cell lines. Finally, SARS-CoV-2 sequences are used as a computational case study to examine the possible use of the proposed approach for viral targets. Candidate gRNAs are prioritized by jointly considering predicted efficiency, off-target risk, sequence conservation, and prediction uncertainty. The resulting ranking provides a computational basis for selecting promising candidates for subsequent experimental validation and does not by itself establish therapeutic efficacy.

Keywords: CRISPR/Cas9, guide RNA, gRNA efficiency prediction, CNN, BiLSTM, Multi-Head Attention, feature fusion, XGBoost, transfer learning, domain adaptation, DANN, CORAL, prediction uncertainty.

1. Introduction

Genome editing has advanced rapidly with the development of CRISPR-based technologies. Among these systems, CRISPR/Cas9 is widely used for targeted DNA modification because the Cas9 nuclease can be directed to a specific genomic region by a guide RNA (gRNA). The resulting DNA break is repaired through cellular mechanisms such as non-homologous end joining (NHEJ) or homology-directed repair (HDR). Therefore, the selection of an effective gRNA is an important factor influencing genome-editing outcomes [1]. However, gRNAs targeting the same genomic region can exhibit substantially different editing efficiencies. This variation is influenced by sequence composition, GC content, nucleotide patterns, target-site characteristics, chromatin accessibility, and epigenetic context [2]. Traditional machine-learning approaches have been used to estimate gRNA activity from manually selected sequence features. Although these methods can provide useful predictions, they may not adequately represent the complex relationships present within gRNA sequences. Deep-learning models such as CNNs and LSTMs offer an alternative by learning sequence patterns directly from the data, while attention mechanisms can help identify positions that contribute more strongly to the prediction. However, relying only on sequence information may overlook other relevant biological characteristics. The use of several feature types can also increase the feature space and introduce redundant information [3]. Prediction across different cell lines presents an additional difficulty. The relationship between

sequence characteristics and gRNA activity can vary with the cellular and genomic environment. Consequently, a model that performs well on one cell line may not provide the same level of performance on another. Transfer learning and domain-adaptation methods offer a way to adapt the learned representations to these differences and improve the transfer of knowledge between datasets. Prediction confidence is also relevant when candidate gRNAs are being ranked. A high predicted efficiency alone does not necessarily mean that the prediction is dependable. Monte Carlo Dropout can be used to obtain repeated predictions and estimate the associated predictive uncertainty [4].

Based on these considerations, this study develops a combined approach for gRNA efficiency prediction across cell lines and subsequent candidate prioritization. Sequence information is learned through a CNN–BiLSTM architecture with Multi-Head Attention, while GC content and k-mer frequencies provide additional sequence-level information. The learned and handcrafted features are integrated, and Recursive Feature Elimination (RFE) is used to remove features that contribute limited or redundant information. XGBoost regression is then employed to estimate gRNA efficiency. Transfer learning and domain adaptation are incorporated to address differences between cell-line datasets. Finally, uncertainty estimates are considered along with predicted efficiency when prioritizing candidate gRNAs for further investigation.

2. Literature Review

CRISPR-Cas9 gRNA efficiency prediction has developed considerably with the use of machine-learning and deep-learning techniques. Early approaches generally relied on manually selected sequence features, whereas more recent models learn useful patterns directly from biological sequences. Some studies have also included additional information, such as epigenetic characteristics and Cas9 variant-specific properties, to improve the prediction of guide activity. Attention-based architectures have further been explored to capture important sequence positions and dependencies. Despite these developments, transferring prediction models between cell lines remains difficult, and there is still a need for methods that consider prediction uncertainty when prioritizing candidate gRNAs.

2.1 DeepCRISPR

It incorporates gRNA sequence information along with epigenetic features to predict guide activity [5]. This combination is useful because gRNA performance may depend not only on the nucleotide sequence but also on the biological environment in which editing takes place. However, the study does not focus on adapting the prediction model to differences among cell-line datasets. Transfer learning and domain adaptation are not considered as part of the prediction framework. Inaddition, prediction uncertainty and the subsequent prioritization of candidate guides are not addressed together. These limitations indicate the need for methods that can make use of information learned from one cellular dataset when predicting guide activity in another.

2.2 DeepHF

DeepHF uses deep-learning methods to predict gRNA activity for high-fidelity Cas9 variants [6]. The work is particularly relevant to guide selection when Cas9 specificity is an important consideration. Its main emphasis, however, is on predicting activity for high-fidelity Cas9 systems rather than on dealing with variations among different cellular datasets. As a result, the problem of transferring a trained model across cell lines is outside the main scope of the study. The approach also does not combine cross-cell-line adaptation with uncertainty estimation and candidate ranking. These differences are relevant to the present work, which considers prediction accuracy as well as model adaptation and confidence in the resulting predictions.

2.3 DeepSpCas9

DeepSpCas9 employs a convolutional neural network to learn sequence patterns associated with SpCas9 activity [7]. CNNs can effectively capture local sequence motifs; however, sequence-only models may not fully represent differences in cellular and epigenetic environments. This limits their ability to generalize across biological domains.

2.4 CRISPRon and Feature-Based Methods

CRISPRon and related machine-learning approaches use manually engineered sequence features for gRNA activity prediction [8]. Although such features can improve interpretability, they may not capture complex sequence dependencies and can introduce redundancy when multiple features are combined. Feature-selection techniques can therefore improve the efficiency of downstream prediction.

2.5 Attention and Transformer-Based Approaches

Attention and transformer-based models can capture important relationships among nucleotide positions and have become increasingly useful for biological sequence modelling [9]. However, improved sequence representation alone does not guarantee reliable cross-cell-line generalization because differences between training and target domains can cause distribution shifts [10]. Therefore, combining deep sequence

representations with domain adaptation, uncertainty estimation, and biological features remains a promising direction for robust gRNA prediction.

3. RESEARCH GAP AND MAIN CONTRIBUTIONS:

3.1 Research Gap

Based on the existing literature, several issues remain open in the prediction and selection of effective CRISPR/Cas9 guide RNAs. The main gaps considered in this study are the following:

1. Combining sequence and biological features
2. Prediction across different cell lines
3. Cross-domain adaptation
4. Selection of informative features
5. Uncertainty in model predictions
6. Beyond efficiency-based selection
7. Integrated candidate prioritization

3.2 Main Contributions of the Proposed Study

The study develops an integrated computational approach that brings together sequence representation learning, feature-based prediction, cross-cell-line adaptation, uncertainty estimation, and candidate prioritization. The main contributions are summarized below:

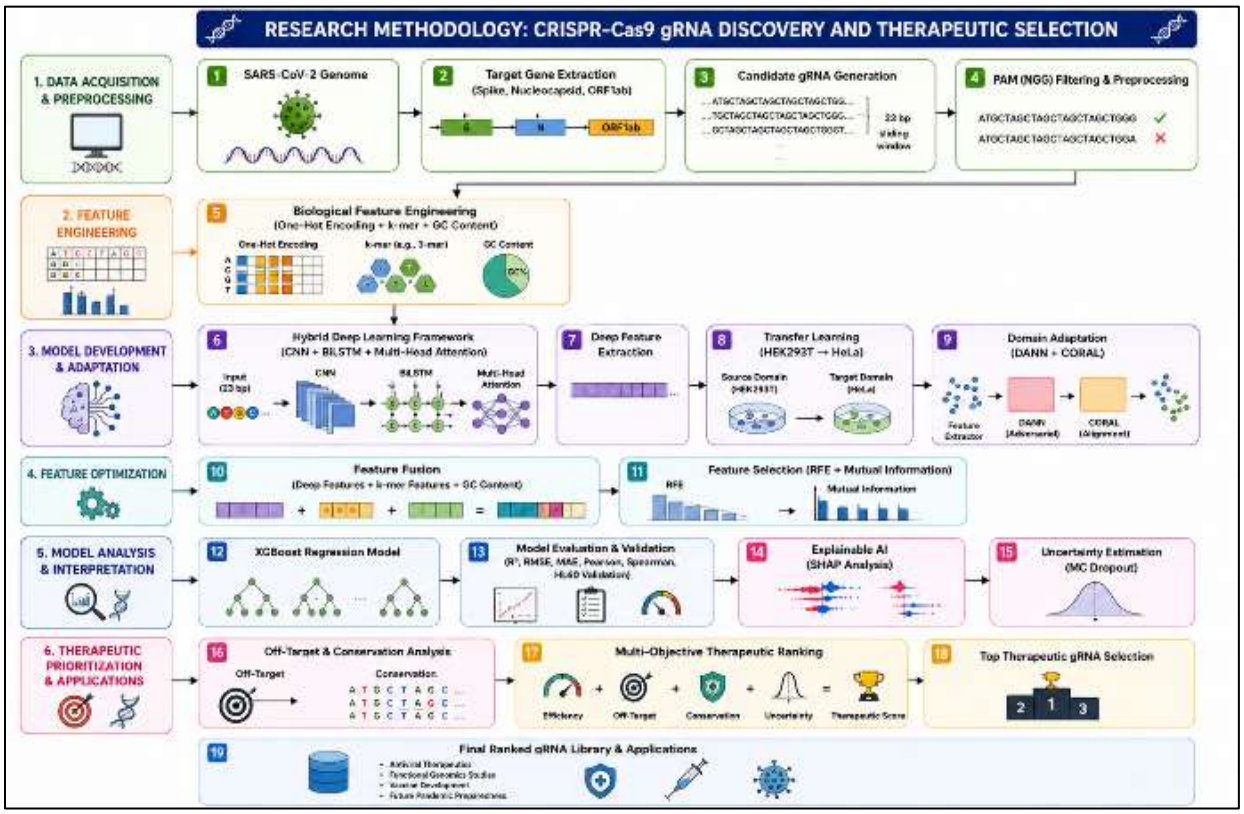
1. Sequence representation using CNN–BiLSTM with Multi-Head Attention:
2. Combination of learned and handcrafted features
3. Feature selection with XGBoost Regression
4. Cross-cell-line transfer learning
5. Domain adaptation using DANN and CORAL
6. Prediction uncertainty estimation
7. Multi-criteria candidate prioritization
8. Cross-cell-line and ablation-based evaluation
9. SARS-CoV-2 computational case study

4. Proposed Research Methodology

4.1 Overall Architecture

The proposed study develops a hybrid computational framework to predict the efficiency of CRISPR/Cas9 guide RNAs and to improve their selection across different biological conditions. Rather than relying on sequence information alone, the framework combines sequence-derived features, deep neural representations, machine-learning regression, and domain adaptation. Experimentally validated guide RNA data obtained from different human cell lines and viral genome sequences are first processed to remove duplicate, incomplete, and inconsistent records. Each guide sequence is represented using one-hot encoding, while additional sequence characteristics, including GC content and k-mer frequencies, are calculated to provide complementary information. These features allow the model to consider both the nucleotide composition and local sequence patterns associated with guide RNA activity. For learning sequence-dependent patterns, the framework uses a CNN followed by a BiLSTM network. The CNN extracts local motifs from the nucleotide sequence, whereas the BiLSTM considers dependencies between nucleotides over a wider sequence context. A Multi-Head Attention layer is subsequently used to assign greater importance to informative regions of the learned representation [11]. The resulting deep features are combined with the calculated GC-content and k-mer features. RFE is then used to retain the more relevant predictors and reduce redundant information before regression. The selected features are supplied to an XGBoost regressor to estimate guide RNA efficiency. XGBoost is included because the fused feature set contains nonlinear relationships that may not be adequately represented by a single linear prediction model. The framework is further extended to address differences between cell-line datasets. A model trained on a source cell line is transferred to a target dataset through fine-tuning, while DANN and CORAL are incorporated to reduce the effect of differences in feature distributions between the two domains [12]. Prediction uncertainty is estimated using Monte Carlo Dropout by evaluating the model over multiple stochastic forward passes. This information is considered together with predicted efficiency when selecting candidate guides. The final prioritization therefore does not depend on efficiency alone; off-target specificity, conservation, and prediction confidence are also incorporated into the ranking process. This provides a final shortlist of guide RNAs that balances predicted activity with the reliability and biological suitability of the

candidates.



4.2 Dataset Description

The proposed framework is evaluated using experimentally validated CRISPR gRNA datasets collected from different biological sources [13]. The HEK293T dataset forms the source domain for model development. It contains gRNA sequences together with experimentally measured editing efficiencies, which are used as the response variable during supervised training. The initial model is trained on these data to learn sequence patterns associated with differences in editing activity. The ability of the model to adapt to another cellular context is investigated using the HeLa dataset. The model developed with HEK293T data is transferred to the HeLa domain and fine-tuned using the available HeLa samples. This experiment provides a direct assessment of whether the learned sequence representations remain useful when the cellular environment changes. In contrast, the HL60 dataset is not included during training or fine-tuning. It is reserved for external validation so that the final model can be evaluated on an independent cell-line dataset. SARS-CoV-2 sequences are considered separately as a computational case study [14]. Candidate guide sequences are evaluated against the viral genome to demonstrate how the proposed framework can be extended from cell-line-specific prediction to guide prioritization for a viral target. The case study is therefore used to illustrate the practical application of the prediction and ranking components rather than as part of the primary training process. For each dataset, the experimentally measured editing efficiency is associated with its corresponding gRNA sequence. Sequence-derived characteristics such as GC content, nucleotide composition, and k-mer frequencies are calculated and incorporated with the representations learned by the deep-learning component. The experimental design follows a source-training, target fine-tuning, and independent-validation strategy. This separation allows the study to evaluate two related aspects of the framework: its ability to predict gRNA efficiency and its ability to retain useful predictive information when applied to data generated under a different cellular condition [15].

4.2.1 SARS-CoV-2 as a Computational Case Study

The SARS-CoV-2 genome will be used as a computational case study to demonstrate the applicability of the proposed CRISPR/Cas9 guide RNA prediction and prioritization framework to a viral genome [16]. Candidate target sequences will be identified from selected regions of the SARS-CoV-2 genome based on CRISPR/Cas9 targeting criteria, including the presence of an appropriate PAM sequence. The proposed model will then be used to estimate the relative guide RNA activity and associated prediction uncertainty.

The candidate guide RNAs will subsequently be evaluated using additional criteria, including potential off-target risk and sequence conservation, and prioritized using the proposed multi-objective ranking strategy. The resulting candidates will represent computationally prioritized guide RNAs for further investigation rather

than experimentally validated therapeutic guides [17]. Therefore, the SARS-CoV-2 analysis is intended to demonstrate the generalizability and practical applicability of the proposed computational framework to viral genomic sequences. Experimental validation would be required before any candidate could be considered for therapeutic application.

4.3 Data Preprocessing

Data preprocessing is performed to transform raw guide RNA sequences into suitable numerical representations for machine learning and deep learning models. Initially, duplicate and incomplete records are removed to improve data quality. Each guide RNA sequence is then converted into a numerical representation using one-hot encoding, where each nucleotide (A, C, G, and T) is represented by a four-dimensional binary vector [18].

In addition to sequence encoding, biological features are extracted from each guide RNA. GC content is calculated because nucleotide composition influences guide RNA stability and editing efficiency. Furthermore, k-mer feature extraction is performed to capture local nucleotide composition by calculating the frequencies of short nucleotide subsequences.

The one-hot encoded sequences serve as input to the deep learning model, while GC content and k-mer features are combined as handcrafted biological descriptors. This preprocessing strategy allows the proposed framework to utilize both automatically learned sequence representations and biologically meaningful engineered features [19].

After explaining one-hot encoding, GC content and k-mers, insert:

Equation 4.1 – GC Content

$$GC(\%) = \frac{N_G + N_C}{N_A + N_C + N_G + N_T} \times 100$$

where N_A, N_C, N_G, N_T represent the number of adenine, cytosine, guanine, and thymine nucleotides, respectively.

Equation 4.2 – One-Hot Encoding

Each nucleotide is represented by a four-dimensional binary vector:

$$A = [1,0,0,0],\ C = [0,1,0,0]$$
$$G = [0,0,1,0],\ T = [0,0,0,1]$$

For a guide RNA of length L, the encoded sequence is represented as:

$$X \in \mathbb{R}^{L \times 4}$$

For a 23-nucleotide guide RNA:

$$X \in \mathbb{R}^{23 \times 4}$$

Equation 4.3 – k-mer Frequency

The frequency of a k-mer $k$is calculated as:

$$f(k) = \frac{C(k)}{L - k + 1}$$

where $C(k)$is the number of occurrences of the k-mer, $L$is the sequence length, and $k$is the k-mer size.
if 3-mers are used, then $k = 3$, and the number of possible positions in a sequence of length $L$is $L - 3 + 1$.

4.4 Deep Learning Framework

The sequence-learning part of the proposed framework uses three successive components: a Convolutional Neural Network (CNN), a Bidirectional Long Short-Term Memory (BiLSTM) network, and a Multi-Head Attention mechanism [20]. These components are used at different stages of sequence representation, beginning with local nucleotide patterns and progressing toward longer-range and position-dependent information.

The CNN operates on the encoded gRNA sequence and learns short sequence patterns that may be associated with editing activity. Its convolutional filters slide across the input sequence and generate feature maps from local nucleotide windows. For an input representation X, the convolution operation can be expressed as:

$$C_i = f(W_i * X + b_i)$$

where Xrepresents the encoded nucleotide sequence, $W_i$denotes the i-th convolutional filter, $b_i$is the corresponding bias, $*$represents the convolution operation, and $f(\cdot)$is the activation function. The resulting feature maps provide the input to the subsequent sequence-learning stage.

The CNN-derived representations are then passed to a BiLSTM network. The use of two processing directions allows the network to consider sequence information from both the beginning and the end of the gRNA.

Consequently, the representation at a particular nucleotide position can incorporate information from both preceding and subsequent positions. This is useful for capturing dependencies that may extend beyond the local patterns detected by the CNN [21].

A Multi-Head Attention layer is subsequently applied to the BiLSTM representation. Rather than treating all sequence positions equally, attention assigns different weights to the learned representations according to their contribution to the prediction. Multiple attention heads can focus on different aspects of the sequence simultaneously, allowing the model to retain informative regions while reducing the contribution of less relevant positions [22]. The resulting representation therefore combines local motifs learned by the CNN with broader sequence dependencies obtained from the BiLSTM and position-specific information emphasized by the attention mechanism.

The convolution operation can be represented as:

Equation 4.4 – CNN Convolution

$$h_i = f(W * x_i + b)$$

where $x_i$represents the input sequence region, Wrepresents the convolution filter, $b$is the bias, $*$denotes the convolution operation, and $f(\cdot)$is the activation function [23].

If ReLU is used:

$$f(z) = max(0, z)$$

Therefore,

$$h_i = ReLU(W * x_i + b)$$

BiLSTM

The forward LSTM processes the sequence from left to right:

Equation 4.5

$$\overrightarrow{h_t} = LSTM\left(x_t, \overrightarrow{h_{t-1}}\right)$$

The backward LSTM processes the sequence from right to left:

Equation 4.6

$$\overleftarrow{h_t} = LSTM\left(x_t, \overleftarrow{h_{t+1}}\right)$$

The two hidden representations are concatenated:

Equation 4.7 – BiLSTM Representation

$$h_t = \left[\overrightarrow{h_t}; \overleftarrow{h_t}\right]$$

This allows the model to capture contextual information from both directions of the guide RNA sequence [24].

Multi-Head Self-Attention

For the attention mechanism, the query, key, and value matrices are calculated as [31]:

Equation 4.8

$$Q = XW_Q,\ K = XW_K,\ V = XW_V$$

Important correction: use $V = XW_V$, rather than $V = X$, unless you intentionally use the input directly as the value matrix.

The scaled dot-product attention is:

Equation 4.9

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

where $d_k$represents the dimensionality of the key vectors.

For multiple attention heads:

Equation 4.10

$$head_i = Attention(Q_i, K_i, V_i)$$

The outputs of all attention heads are concatenated:

Equation 4.11 – Multi-Head Attention

$$MultiHead(Q, K, V) = Concat(h, e, ad_1 \dots head_H)W_O$$

where $H$is the number of attention heads and $W_O$is the output projection matrix.

4.5 Feature Fusion and XGBoost Regression

For thesis preparation, I recommend adding equations for feature fusion, feature selection, and XGBoost, because currently this section is mostly descriptive.

Feature Fusion

Let Drepresent the deep features obtained from CNN–BiLSTM–Attention and Brepresent handcrafted biological features containing GC and k-mer features [25].

Equation 4.12 – Feature Fusion

$$F = [D; B]$$

where $F$is the combined feature representation and [;]denotes feature-wise concatenation.

For example:

$$F = \left[, F_{deep}, F_{GC} F_{kmer} \right]$$

RFE

The importance of feature $j$can be represented as:

$$I_j = \text{Importance}\left(F_j\right)$$

At each iteration, the least important features are removed until the desired number of features is obtained.

You can describe the selected feature set as:

Equation 4.13

$$F^* = RFE(F)$$

where F^*represents the selected feature subset.

XGBoost

The XGBoost prediction can be expressed as [26]:

Equation 4.14

$$\hat{y}_i = \sum_{m=1}^{M} f_m\left(F_i\right),\ f_m \in \mathcal{F}$$

where $M$is the number of boosting trees and $f_m$represents the m-th regression tree.

The objective function is:

Equation 4.15

$$\mathcal{L} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i\right) + \sum_{m=1}^{M} \Omega\left(f_m\right)$$

where $l(y_i, \hat{y}_i)$is the prediction loss and $\Omega(f_m)$is the regularization term controlling model complexity.

4.6 Transfer Learning, Domain Adaptation and Uncertainty Estimation

Transfer Learning

The parameters learned from the source cell line HEK293T are transferred to the target cell line HeLa [27]:

Equation 4.16

$$\theta_{HeLa} = FineTune(\theta_{HEK293T}, D_{HeLa})$$

where θrepresents the model parameters.

CORAL

For domain adaptation, CORAL minimizes the difference between source and target covariance matrices:

Equation 4.17 – CORAL Loss

$$L_{CORAL} = \frac{1}{4d^2} \|C_S - C_T\|_F^2$$

where $C_S$and $C_T$are the covariance matrices of the source and target domains, $d$is the feature dimension, and $\|\cdot\|_F$is the Frobenius norm.

DANN

The DANN objective can be represented as [28]:

Equation 4.18

$$L_{DANN} = L_{task} - \lambda L_{domain}$$

where $L_{task}$is the guide RNA prediction loss, $L_{domain}$is the domain classification loss, and λcontrols the contribution of domain adaptation.

Monte Carlo Dropout

With Tstochastic forward passes [29]:

Equation 4.19 – Mean Prediction

$$\mu(x) = \frac{1}{T}\sum_{t=1}^{T} \hat{y}^{(t)}(x)$$

The prediction variance is:

Equation 4.20 – Prediction Uncertainty

$$\sigma^2(x) = \frac{1}{T}\sum_{t=1}^{T} \left(\hat{y}^{(t)}(x) - \mu(x)\right)^2$$

Thus, the final predicted efficiency is represented by $\mu(x)$, while $\sigma^2(x)$represents prediction uncertainty [30].

4.7 Therapeutic Guide RNA Ranking

This section needs an explicit **multi-objective ranking equation**.

First, normalize the four criteria to the range $[0, 1]$:

E_i= predicted editing efficiency

S_i= off-target specificity

C_i= conservation score

U_i= prediction uncertainty

Because lower uncertainty is desirable, use $1 - U_i$.

Equation 4.21 – Therapeutic Ranking Score

$$R_i = w_E E_i + w_S S_i + w_C C_i + w_U(1 - U_i)$$

subject to:

$$w_E + w_S + w_C + w_U = 1$$

and

$$w_E, w_S, w_C, w_U \geq 0$$

The guide RNAs are then ranked according to descending R_i:

$$gRNA_{(1)} \succ gRNA_{(2)} \succ \cdots \succ gRNA_{(n)}$$

where $gRNA_{(1)}$represents the highest-ranked candidate [38].

Important: If you actually calculate **off-target risk** rather than specificity, use:

$$R_i = w_E E_i + w_C C_i + w_S(1 - O_i) + w_U(1 - U_i)$$

where $O_i$is the normalized off-target risk. This is probably the better formulation for your thesis because your text repeatedly says **low off-target risk** is desirable.

4.7.1 Off-Target Prediction

Off-target activity occurs when a CRISPR/Cas9 guide RNA binds to genomic locations other than its intended target site and causes unintended DNA modification. Minimizing off-target activity is essential for improving the specificity and safety of CRISPR/Cas9 applications, particularly when guide RNAs are being considered for therapeutic research. Therefore, predicted on-target efficiency alone is not sufficient for selecting the most suitable guide RNA [31].

In the proposed research, each candidate guide RNA will be evaluated against potential genomic off-target sites by identifying sequences with similarity to the guide RNA and an appropriate Cas9 PAM sequence. Candidate off-target sites will be characterized using sequence-related factors such as the number and position of nucleotide mismatches between the guide RNA and potential off-target sequence. Greater sequence similarity, particularly in functionally important regions of the guide sequence, may indicate a higher potential for unintended activity.

The identified potential off-target sites will be further evaluated to estimate an off-target risk score. The score

will be used to distinguish guide RNAs with high predicted on-target activity from those that may have a higher risk of unintended genomic interactions. A guide RNA with high predicted efficiency but substantial off-target risk will therefore receive a lower priority than a guide RNA providing comparable efficiency with lower predicted off-target activity.

The off-target assessment will be integrated with the other criteria in the proposed multi-objective ranking framework. Specifically, predicted guide RNA efficiency, off-target risk, sequence conservation, and prediction uncertainty will be considered jointly. The final ranking will therefore prioritize guide RNAs that exhibit high predicted editing efficiency, low potential off-target risk, appropriate conservation characteristics, and low prediction uncertainty.

The computationally identified off-target candidates and prioritized guide RNAs will be considered as predictions for further experimental investigation. The proposed framework does not claim experimental confirmation of off-target cleavage; laboratory validation would be required to establish the actual biological activity of the predicted sites.

You should include a formula for mismatch-based off-target risk rather than describing it only in words.

Let m_i be the number of mismatches between the guide RNA and an off-target sequence.

Equation 4.22 – Mismatch Rate

$$MR_i = \frac{m_i}{L}$$

where L is the guide RNA length.

Because mismatch positions can have different biological importance, a weighted mismatch score can be defined as:

Equation 4.23 – Weighted Mismatch Score

$$WMS_i = \sum_{j=1}^{L} w_j \, \mathbb{I}(g_j \neq t_{ij})$$

where:

g_j = nucleotide at position j in the guide RNA,

t_{ij} = nucleotide at position j in the i-th potential off-target,

w_j = positional weight,

$\mathbb{I}(\cdot)$ = indicator function.

A normalized off-target risk score can then be represented as:

Equation 4.24 – Off-Target Risk

$$O_i = \sum_{j=1}^{N_i} P_{ij}$$

where P_{ij} represents the predicted activity/risk associated with the j-th potential off-target site for guide RNA i, and N_i is the number of candidate off-target sites.

For therapeutic ranking, the normalized specificity score can then be defined as:

Equation 4.25 – Off-Target Specificity

$$S_i = 1 - \tilde{O}_i$$

where $\tilde{O}_i$ is the normalized off-target risk.

5. Results And Discussion

5.1 Overview of Experimental Results

The proposed framework was evaluated to examine its ability to predict gRNA efficiency and to assess the usefulness of combining deep sequence representations with handcrafted features. The evaluation also considers cross-cell-line prediction, comparison with conventional and deep-learning approaches, and the contribution of different components through ablation analysis [32].

The principal evaluation measures include Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), coefficient of determination ((R^2)), Pearson correlation, and Spearman correlation. MAE and RMSE describe the magnitude of prediction errors, whereas Pearson and Spearman correlations provide information about the agreement between predicted and experimentally measured activity. Spearman correlation is particularly relevant when the model is used for ranking candidate gRNAs.

5.2 Overall Prediction Performance

Another performance table in the manuscript reports the following results for the proposed model:

Table 1. Comparison with Proposed models

Metric	Reported Value
(R^2)	0.74
MSE	0.05
RMSE	0.11
MAE	0.05
Pearson correlation	0.86
Spearman correlation	0.84
Kendall correlation	0.60

These results indicate a strong correlation between predicted and observed gRNA activity according to the reported experimental values. The Spearman correlation of 0.84 suggests that the model is able to reproduce the relative ordering of candidate gRNAs reasonably well.

The inclusion of both Pearson and Spearman correlation is useful because the two measures capture different aspects of model performance. Pearson correlation evaluates the linear association between predictions and observations, whereas Spearman correlation focuses on the consistency of their ranking.

For a gRNA prioritization application, the ranking behaviour is particularly important. A model that produces accurate relative ordering can help researchers focus experimental resources on a smaller number of promising candidates.

Again, these values should not be combined with the alternative result set unless they originate from the same final experiment.

5.3 Comparison with Baseline Models

The reported comparison includes SVR, Random Forest, CNN, CNN-LSTM, and the proposed hybrid approach.

Table 2. Comparison with baseline models

Model	RMSE	MAE	(R^2)	Spearman
SVR	0.182	0.141	0.48	0.51
Random Forest	0.165	0.129	0.56	0.59
CNN	0.149	0.118	0.64	0.66
CNN-LSTM	0.132	0.104	0.72	0.74
Proposed Hybrid	0.112	0.056	0.74	0.84

The reported results show a progressive improvement from conventional machine-learning approaches to deep-learning models. SVR provides the lowest reported (R^2) and Spearman correlation among the listed approaches, while Random Forest provides a moderate improvement.

The CNN model performs better than the conventional models, indicating the usefulness of automatically learned sequence representations [33]. The addition of LSTM-based sequential modelling further improves the reported performance of the CNN-LSTM model.

The proposed hybrid model achieves the highest reported Spearman correlation of 0.84 and the highest reported (R^2) of 0.74 in this comparison. This suggests that combining multiple modelling components can improve both predictive agreement and ranking performance.

5.4 Five-Fold Cross-Validation

Five-fold cross-validation was performed to examine the stability of the reported model performance. The manuscript reports (R^2) values of 0.86, 0.83, 0.87, 0.88, and 0.86 across the five folds.

Table 3. Reported five-fold cross-validation results

Fold	(R^2)
Fold 1	0.86
Fold 2	0.83

Fold 3	0.87
Fold 4	0.88
Fold 5	0.86
Mean	0.86
Standard deviation	0.015

The relatively small variation between folds, according to the reported results, suggests that the model performance is not strongly dependent on a single data partition [34]. The reported standard deviation of 0.015 indicates limited variation in (R^2) across the five folds.
Cross-validation therefore provides additional evidence about the stability of the model within the evaluated dataset.

5.5 Cross-Cell-Line Evaluation
Cross-cell-line evaluation is an important part of the proposed framework because gRNA activity can be influenced by cellular context. The manuscript reports results for HeLa and HL60 in addition to the source-cell-line experiments.

Table 4. Reported cross-cell-line performance

Dataset	(R^2)	RMSE	MAE	Pearson	Spearman	Kendall
HeLa	0.72	0.054	0.034	0.78	0.82	0.67
HL60	0.61	0.068	0.046	0.80	0.75	0.60

The reported HeLa results show a Spearman correlation of 0.82, while the HL60 results show a Spearman correlation of 0.75. The decrease in (R^2) and Spearman correlation for HL60 indicates that prediction becomes more difficult when the model is evaluated under a different cellular context.
This observation is relevant to the motivation for transfer learning and domain adaptation. A model that performs well in one cell line may encounter a distribution shift when applied to another cell line. Therefore, evaluating the framework across multiple cell lines provides a more informative assessment than evaluating it using only one dataset [35].
The cross-cell results should, however, be interpreted together with the exact experimental protocol used for transfer learning and domain adaptation.

5.6 Interpretation of Prediction and Ranking Performance
The evaluation uses both error-based and correlation-based metrics because they provide different information about model behaviour.
A lower MAE indicates that the predicted activity values are closer to the observed values on average. RMSE provides additional sensitivity to larger errors. (R^2) measures the proportion of observed variation represented by the model.
Correlation metrics provide a complementary perspective. Pearson correlation measures the linear relationship between predicted and observed activity, whereas Spearman correlation evaluates the agreement between their rankings.
The distinction is important for gRNA prioritization. In a practical screening workflow, the primary objective may not always be to reproduce every activity value exactly. Instead, the model may be used to
identify a smaller group of high-performing candidates from a larger pool. In such a setting, ranking quality becomes an important consideration.

5.7 SARS-CoV-2 Computational Results
The proposed workflow was additionally applied to SARS-CoV-2 genomic sequences as a computational case study. The manuscript reports the generation of candidate gRNAs followed by PAM filtering and prediction. The reported SARS-CoV-2 analysis includes [37].

Table 5: COVID-19 gRNA:

Prediction Summary	**Reported Value**
Total candidate gRNAs	3,799
Valid PAM candidates	136
Highest predicted activity	0.78

Average predicted activity	0.257

The analysis demonstrates how the proposed computational workflow can be extended from cell-line datasets to a genomic sequence of interest. Candidate guides can first be generated according to sequence and PAM requirements and subsequently ranked according to predicted activity and other prioritization criteria. These results should be regarded as computational predictions only. They do not establish antiviral activity or therapeutic effectiveness. Experimental studies would be reqired to determine whether the computationally prioritized candidates produce the expected biological effects. The manuscript currently also reports a lowest predicted value of 0.26, which is higher than the reported average of 0.257 [38]. This apparent inconsistency should be checked against the original prediction file before publication.

5.8 Top-Ranked gRNA Candidates

The manuscript provides a set of top-ranked gRNAs obtained from the proposed prioritization workflow. The ranking considers predicted activity together with additional candidate characteristics such as off-target information and conservation.

The top-ranked candidates are intended to represent guides that satisfy the computational selection criteria used by the framework. These candidates can be considered for subsequent experimental evaluation.

The ranking should be interpreted as a computational screening step rather than evidence that the selected guides will necessarily produce the predicted biological outcome [36]. Experimental validation remains necessary to establish actual editing efficiency and specificity.

Table 6: Top Ranked grna activity:

Rank / ID	Predicted Sequence	Off-Target Score	Conservation Score	Uncertainty	Therapeutic Score
4	TCCATGCTATACATGTCTCTGGG	0.260550	0.996660	0.980400	0.020877
8	GAGAAGTCTAACATAATAAGAGG	0.250854	0.996617	0.968636	0.015083
107	CAAATTGATAGGTTGATCACAGG	0.265194	0.978334	0.953160	0.023815
24	TGATCTCCCTCAGGGTTTTTCGG	0.264376	0.914494	0.997654	0.036782
78	CTCAGACTAATTCTCCTCGGCGG	0.272653	0.902501	0.976374	0.011719
23	TTAGTGCGTGATCTCCCTCAGGG	0.288334	0.858886	0.980903	0.041931
112	TTCCCTCAGTCAGCACCTCATGG	0.272044	0.988898	0.928480	0.036400
121	TGAATTAGACTCATTCAAGGAGG	0.274614	0.889097	0.975134	0.017557
3	TTCCATGCTATACATGTCTCTGG	0.254448	0.947820	0.973546	0.017274

6. Conclusion

The study develops a hybrid computational approach for predicting CRISPR/Cas9 guide RNA efficiency. Sequence information is processed through a CNN–BiLSTM architecture with Multi-Head Attention to learn local motifs and longer-range dependencies. The learned representation is combined with sequence-derived features, including GC content and k-mer frequencies. Recursive Feature Elimination (RFE) is then used to remove less informative or redundant features, and the resulting feature set is supplied to an XGBoost regression model for efficiency prediction. To examine performance across different cell-line datasets, the workflow incorporates transfer learning and domain adaptation through Domain-Adversarial Neural Networks (DANN) and CORAL. Monte Carlo Dropout is used to estimate prediction uncertainty. For candidate selection, predicted efficiency is considered together with off-target risk, conservation, and uncertainty, allowing the guides to be ranked for further investigation.

This design addresses several limitations observed in existing gRNA prediction approaches. The combination of learned sequence representations and handcrafted features allows different types of sequence information to be considered within the same prediction workflow. The use of transfer learning and domain adaptation provides a mechanism for handling differences between cell-line datasets rather than assuming that a model trained on one dataset will perform identically on another. Feature selection is included to control redundancy introduced during feature fusion. In addition, uncertainty estimates provide information beyond the predicted efficiency value itself and can be considered when comparing candidate guides. The framework has potential uses in computational genome-editing studies. In cancer research, it can assist in the initial selection of candidate gRNAs for targets of interest. A similar strategy can be applied to viral genomes, including SARS-CoV-2, as demonstrated through the computational case study in this work. The approach may also be relevant to studies of rare genetic disorders, functional genomics, and other applications in which the selection of efficient

gRNAs is an important step. These applications should be regarded as computational support for candidate selection; experimental testing remains necessary before drawing conclusions about biological or therapeutic effectiveness. The findings from this work are intended to provide a basis for publication in peer-reviewed journals concerned with bioinformatics, computational biology, artificial intelligence, and genome editing. The results may also be disseminated through scientific conferences and related academic forums. Several extensions are possible in future work. Additional biological information, including epigenetic and multi-omics features, could be incorporated to investigate whether they further improve prediction across cellular contexts. The same framework could also be adapted to other CRISPR systems, such as Cas12a and Cas13, with appropriate changes to the sequence and target characteristics. Transformer-based genomic models could be explored as alternative sequence encoders [39]. Most importantly, laboratory validation of selected gRNAs would provide experimental evidence for assessing the practical value of the computational rankings.

References

[1] M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, and E. Charpentier, "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity," *Science*, vol. 337, no. 6096, pp. 816–821, 2012, doi: 10.1126/science.1225829.

[2] L. Cong, F. A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. D. Hsu, X. Wu, W. Jiang, L. A. Marraffini, and F. Zhang, "Multiplex genome engineering using CRISPR/Cas systems," *Science*, vol. 339, no. 6121, pp. 819–823, 2013, doi: 10.1126/science.1231143.

[3] P. Mali, L. Yang, K. M. Esvelt, J. Aach, M. Guell, M. D. I. DiCarlo, J. E. Norville, and G. M. Church, "RNA-guided human genome engineering via Cas9," *Science*, vol. 339, no. 6121, pp. 823–826, 2013, doi: 10.1126/science.1232033.

[4] M. Jinek, A. East, A. Cheng, S. Lin, E. Ma, and J. A. Doudna, "RNA-programmed genome editing in human cells," *eLife*, vol. 2, p. e00471, 2013, doi: 10.7554/eLife.00471.

[5] J. A. Doudna and E. Charpentier, "The new frontier of genome engineering with CRISPR-Cas9," *Science*, vol. 346, no. 6213, p. 1258096, 2014, doi: 10.1126/science.1258096.

[6] J. G. Doench, E. Hartenian, D. B. Graham, Z. Tothova, M. Hegde, I. Smith, M. Sullender, B. L. Ebert, R. J. Xavier, and D. E. Root, "Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation," *Nature Biotechnology*, vol. 32, pp. 1262–1267, 2014, doi: 10.1038/nbt.3026.

[7] H. Xu, T. Xiao, C.-H. Chen, W. Li, C. A. Meyer, Q. Wu, D. Wu, L. Cong, F. Zhang, J. S. Liu, M. Brown, and X. S. Liu, "Sequence determinants of improved CRISPR sgRNA design," *Genome Research*, vol. 25, no. 8, pp. 1147–1157, 2015, doi: 10.1101/gr.191452.115.

[8] P. D. Hsu, D. A. Scott, J. A. Weinstein, F. A. Ran, S. Konermann, V. Agarwala, Y. Li, E. J. Fine, X. Wu, O. Shalem, T. J. Cradick, L. A. Marraffini, G. Bao, and F. Zhang, "DNA targeting specificity of RNA-guided Cas9 nucleases," *Nature Biotechnology*, vol. 31, no. 9, pp. 827–832, 2013, doi: 10.1038/nbt.2647.

[9] Y. Fu, J. A. Foden, C. Khayter, M. L. Maeder, D. Reyon, J. K. Joung, and J. D. Sander, "High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells," *Nature Biotechnology*, vol. 31, pp. 822–826, 2013, doi: 10.1038/nbt.2623.

[10] J. G. Doench, N. Fusi, M. Sullender, M. Hegde, E. W. Vaimberg, K. F. Donovan, I. Smith, Z. Tothova, C. Wilen, R. Orchard, H. W. Virgin, J. Listgarten, and D. E. Root, "Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9," *Nature Biotechnology*, vol. 34, no. 2, pp. 184–191, 2016, doi: 10.1038/nbt.3437.

[11] J. Listgarten, M. Weinstein, B. P. Kleinstiver, A. A. Sousa, J. K. Joung, J. Crawford, K. Gao, L. Hoang, M. Elibol, J. G. Doench, and N. Fusi, "Prediction of off-target activities for the end-to-end design of CRISPR guide RNAs," *Nature Biomedical Engineering*, vol. 2, no. 1, pp. 38–47, 2018, doi: 10.1038/s41551-017-0178-6.

[12] S. Q. Tsai, Z. Zheng, N. T. Nguyen, M. Liebers, V. Topkar, V. Thapar, N. Wyvekens, C. Khayter, A. J. Iafrate, L. P. Le, M. J. Aryee, and J. K. Joung, "GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases," *Nature Biotechnology*, vol. 33, no. 2, pp. 187–197, 2015, doi: 10.1038/nbt.3117.

[13] M. A. Moreno-Mateos, C. E. Vejnar, J.-D. Beaudoin, J. P. Fernandez, E. K. Mis, M. K. Khokha, and A. J. Giraldez, "CRISPRscan: Designing highly efficient sgRNAs for CRISPR/Cas9 targeting in vivo," *Nature Methods*, vol. 12, no. 10, pp. 982–988, 2015, doi: 10.1038/nmeth.3543.

[14] T. G. Montague, J. M. Cruz, J. A. Gagnon, G. M. Church, and E. Valen, "CHOPCHOP: A CRISPR/Cas9 and TALEN web tool for genome editing," *Nucleic Acids Research*, vol. 42, no. W1, pp. W401–W407, 2014, doi: 10.1093/nar/gku410.

[15] G. Chuai, H. Ma, J. Yan, M. Chen, N. Hong, D. Xue, C. Zhou, C. Zhu, K. Chen, B. Duan, F. Gu, S. Qu, D. Huang, J. Wei, and Q. Liu, "DeepCRISPR: Optimized CRISPR guide RNA design by deep learning," *Genome Biology*, vol. 19, p. 80, 2018, doi: 10.1186/s13059-018-1459-4.

[16] F. Alkan, A. Wenzel, C. Anthon, J. H. Havgaard, and J. Gorodkin, "CRISPR-Cas9 off-targeting assessment with nucleic acid duplex energy parameters," *Genome Biology*, vol. 19, p. 177, 2018, doi: 10.1186/s13059-018-1534-x.

[17] H. K. Kim, Y. Kim, S. Lee, S. Min, J. Y. Bae, J. W. Choi, J. Park, D. Jung, S. Yoon, and H. H. Kim, "SpCas9 activity prediction by DeepSpCas9, a deep learning-based model with high generalization performance," *Science Advances*, vol. 5, no. 11, p. eaax9249, 2019, doi: 10.1126/sciadv.aax9249.

[18] D. Wang, C. Zhang, B. Wang, B. Li, Q. Wang, D. Liu, H. Wang, Y. Zhou, L. Shi, F. Lan, and Y. Wang, "Optimized CRISPR guide RNA design for two high-fidelity Cas9 variants by deep learning," *Nature Communications*, vol. 10, p. 4284, 2019, doi: 10.1038/s41467-019-12281-8.

[19] N. Kim, H. K. Kim, S. Lee, J. H. Seo, J. W. Choi, J. Park, S. Min, S. Yoon, S.-R. Cho, and H. H. Kim, "Prediction of the sequence-specific cleavage activity of Cas9 variants," *Nature Biotechnology*, vol. 38, no. 11, pp. 1328–1336, 2020, doi: 10.1038/s41587-020-0537-9.

[20] X. Xiang, G. I. Corsi, C. Anthon, K. Qu, X. Pan, X. Liang, P. Han, Z. Dong, L. Liu, J. Zhong, T. Ma, J. Wang, X. Zhang, H. Jiang, F. Xu, X. Liu, X. Xu, J. Wang, H. Yang, L. Bolund, G. M. Church, L. Lin, J. Gorodkin, and Y. Luo, "Enhancing CRISPR-Cas9 gRNA efficiency prediction by data integration and deep learning," *Nature Communications*, vol. 12, p. 3238, 2021, doi: 10.1038/s41467-021-23576-0.

[21] V. Konstantakos, A. Nentidis, A. Krithara, and G. Paliouras, "CRISPRedict: A CRISPR-Cas9 web tool for interpretable efficiency predictions," *Nucleic Acids Research*, vol. 50, no. W1, pp. W191–W198, 2022, doi: 10.1093/nar/gkac466.

[22] J. Lin and K.-C. Wong, "Off-target predictions in CRISPR-Cas9 gene editing using deep learning," *Bioinformatics*, vol. 34, no. 17, pp. i656–i663, 2018, doi: 10.1093/bioinformatics/bty554.

[23] B. P. Kleinstiver, V. Pattanayak, M. S. Prew, S. Q. Tsai, N. T. Nguyen, Z. Zheng, and J. K. Joung, "High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects," *Nature*, vol. 529, pp. 490–495, 2016, doi: 10.1038/nature16526.

[24] I. M. Slaymaker, L. Gao, B. Zetsche, D. A. Scott, W. X. Yan, and F. Zhang, "Rationally engineered Cas9 nucleases with improved specificity," *Science*, vol. 351, no. 6268, pp. 84–88, 2016, doi: 10.1126/science.aad5227.

[25] H. K. Kim, S. Min, M. Song, S. Jung, J. W. Choi, Y. Kim, S. Lee, S. Yoon, and H. H. Kim, "Deep learning improves prediction of CRISPR-Cpf1 guide RNA activity," *Nature Biotechnology*, vol. 36, no. 3, pp. 239–241, 2018, doi: 10.1038/nbt.4061.

[26] E. Hartenian and J. G. Doench, "Genetic screens and functional genomics using CRISPR/Cas9 technology," *The FEBS Journal*, vol. 282, no. 8, pp. 1383–1393, 2015, doi: 10.1111/febs.13248.

[27] R. Barrangou, "Cas9 targeting and the CRISPR revolution," *Science*, vol. 344, no. 6185, pp. 707–708, 2014, doi: 10.1126/science.1252964.

[28] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998, doi: 10.1109/5.726791.

[29] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.

[30] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997, doi: 10.1109/78.650093.

[31] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, vol. 30, 2017, pp. 5998–6008, doi: 10.5555/3295222.3295349.

[32] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine Learning*, vol. 46, pp. 389–422, 2002, doi: 10.1023/A:1012487302797.

[33] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, USA, 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[34] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010, doi: 10.1109/TKDE.2009.191.

[35] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," in *Domain Adaptation for Large-Scale Sentiment Classification*, 2017, pp. 189–209, doi: 10.1007/978-3-319-58347-1_10.

[36] B. Sun and K. Saenko, "Deep CORAL: Correlation alignment for deep domain adaptation," in *Computer Vision – ECCV 2016 Workshops*, pp. 443–450, 2016, doi: 10.1007/978-3-319-49409-8_35.

[37] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in *Proc. 33rd International Conference on Machine Learning*, vol. 48, pp. 1050–1059, 2016.

[38] F. Wu, S. Zhao, B. Yu, Y.-M. Chen, W. Wang, Z.-G. Song, Y. Hu, Z.-W. Tao, J.-H. Tian, Y.-Y. Pei, M.-L. Yuan,

Y.-L. Zhang, F.-H. Dai, Y. Liu, Q.-M. Wang, J.-J. Zheng, L. Xu, E. C. Holmes, and Y.-Z. Zhang, "A new coronavirus associated with human respiratory disease in China," *Nature*, vol. 579, pp. 265–269, 2020, doi: 10.1038/s41586-020-2008-3.

[39] F. Alkan, A. Wenzel, C. Anthon, J. H. Havgaard, and J. Gorodkin, "CRISPR-Cas9 off-targeting assessment with nucleic acid duplex energy parameters," *Genome Biology*, vol. 19, p. 177, 2018, doi: 10.1186/s13059-018-1534-x.