Predicting the evolutionary fitness effects of synonymous mutations

Seh Na Mellick[1,2]

smellick | sehna@cmu.edu

[1]Ray and Stephanie Lane Department of Computational Biology, Carnegie Mellon University, Pittsburgh, PA, USA

[2]Joint Carnegie Mellon University-University of Pittsburgh PhD Program in Computational Biology, Pittsburgh, PA, USA

Spring 2025

Abstract

Synonymous mutations, though they do not alter amino acid sequences, can significantly affect gene expression, translation dynamics, and cellular fitness. Despite increasing recognition of their functional relevance, predicting their fitness effects remains difficult due to subtle and context-dependent mechanisms. Here, we develop an interpretable machine learning framework to predict the fitness consequences of synonymous mutations in *Saccharomyces cerevisiae*. Using deep mutational scanning data, we construct a feature set capturing codon usage bias, mRNA secondary structure, ribosome profiling metrics, and evolutionary conservation. We benchmark several classifiers on a three-class task—strongly deleterious, mildly deleterious, and neutral—and evaluate performance using precision, recall, and confusion matrices. While ensemble models achieve high accuracy for neutral mutations, they struggle to recover deleterious variants due to class imbalance and overlapping features. SHAP analysis identifies codon optimality and RNA structural disruption as dominant predictive signals. Our results highlight the importance of biologically grounded features and underscore the need for interpretable models in capturing the hidden fitness costs of synonymous variation.

1. Introduction

Synonymous mutations are a class of genetic variants that do not alter the amino acid sequence of a protein. Traditionally, these mutations have been considered "silent" or selectively neutral, as they do not directly change the protein's primary structure [1, 6]. However, recent research has revealed that synonymous mutations can have significant effects on gene expression, protein folding, and overall cellular fitness [9]. Synonymous variants have been shown to impact fitness by altering the folding and post-transcriptional modification of mRNA structures and by modulating translation initiation and elongation rates [3, 9]. This realization has reshaped our understanding of molecular evolution, highlighting the importance of synonymous sites in evolutionary dynamics and fitness landscapes.

The downstream consequences of synonymous mutations arise from codon-specific interactions with the ribosome, tRNA pools, RNA-binding proteins, and RNA structural elements [2, 9]. Figure 1 illustrates several of these mechanisms. Panel 1A shows that even when a mutation does not change the encoded amino acid, it can alter the mRNA sequence and its properties. In panel 1B, changes to codon usage can influence translation kinetics by affecting how efficiently tRNAs are recruited. Panel 1C highlights how

synonymous variants can disrupt mRNA secondary structures, altering transcript stability and translation efficiency. These examples reflect broader principles; namely, that synonymous sites, despite lacking protein-coding changes, can be important targets of selection. This emerging understanding has catalyzed a shift in the study of molecular evolution, prompting renewed interest in the functional consequences of synonymous variation [2]. While early genome-scale evolutionary models largely ignored synonymous sites in fitness predictions, accumulating evidence has challenged the assumption that these changes are silent or selectively neutral. Importantly, these effects can vary across species, genes, and environmental conditions, introducing both complexity and richness into evolutionary predictions.

A 2022 landmark study by Shen et al. employed deep mutational scanning (DMS) to systematically quantify the fitness effects of synonymous mutations in the yeast *Saccharomyces cerevisiae* [8]. Their experiments involved high-throughput DMS via CRISPR-Cas9 gene editing of 150 bp segments of DNA coding sequences to generate all possible single nucleotide substitutions. As such, they created a library of strains containing all possible synonymous and nonsynonymous SNPs within those regions. Their DMS workflow was applied to a curated set of 22 nonessential *S. cerevisiae* genes known to influence organismal fitness. After performing competition assays, their results showed that a substantial fraction of synonymous mutations were deleterious, with effects often comparable in magnitude to nonsynonymous changes. The authors hypothesized that these outcomes were linked to mechanisms such as disrupted mRNA stability or altered co-translational folding. Their DMS fitness dataset is publicly available and thus offers a high-quality, experimentally validated ground truth for computational models of fitness effects. As such, it presents a unique opportunity to explore the mechanistic underpinnings of synonymous mutations and their consequences for fitness and evolution.

Despite the growing recognition of the functional importance of synonymous mutations, predictive models that can accurately forecast their fitness effects remain limited. Traditional evolutionary models have long disregarded the rich biological context surrounding synonymous sites, leading to oversimplified predictions. Conversely, machine learning approaches offer a promising avenue to integrate diverse biological features and learn complex relationships between sequence variation and fitness outcomes. However, many existing deep learning models lack interpretability and are trained predominantly on datasets curated solely from evolutionary genomics features. This makes it difficult to extract detailed mechanistic insights about how synonymous mutations influence fitness.

To address these challenges, we developed an interpretable machine learning framework trained on deep mutational scanning data from *S. cerevisiae* [8]. We curated the dataset with biologically meaningful features, including sequence context, codon usage metrics, mRNA structural disruption, and evolutionary conservation. Using these features, we trained classifiers to predict the fitness effects of synonymous mutations. Benchmarking showed that ensemble methods performed well for neutral mutations but struggled with deleterious ones, due in part to class imbalance and overlapping feature distributions. Through SHAP-based interpretation, we identified codon optimality and structural destabilization as dominant predictive signals, linking mechanistic features to functional outcomes. These findings highlight the potential of interpretable models to uncover the subtle but consequential effects of synonymous variation on cellular fitness.

2. Methods

Our modeling strategy aims to predict the evolutionary fitness effects of synonymous mutations using an interpretable, biologically grounded machine learning framework. The overall computational approach was organized into four stages: (i) generating mutation-centered sequence windows, (ii) extracting biologically informative features, (iii) integrating evolutionary and structural annotations into a unified representation, and (iv) predictive modeling with interpretable attribution techniques. Together, these components enabled us to identify molecular mechanisms by which synonymous mutations impact gene expression and cellular fitness despite conserving the original protein sequence.

2.1. Mutation-centered window generation

To provide consistent structural and sequence context for each mutation, we extracted 51-nucleotide win-

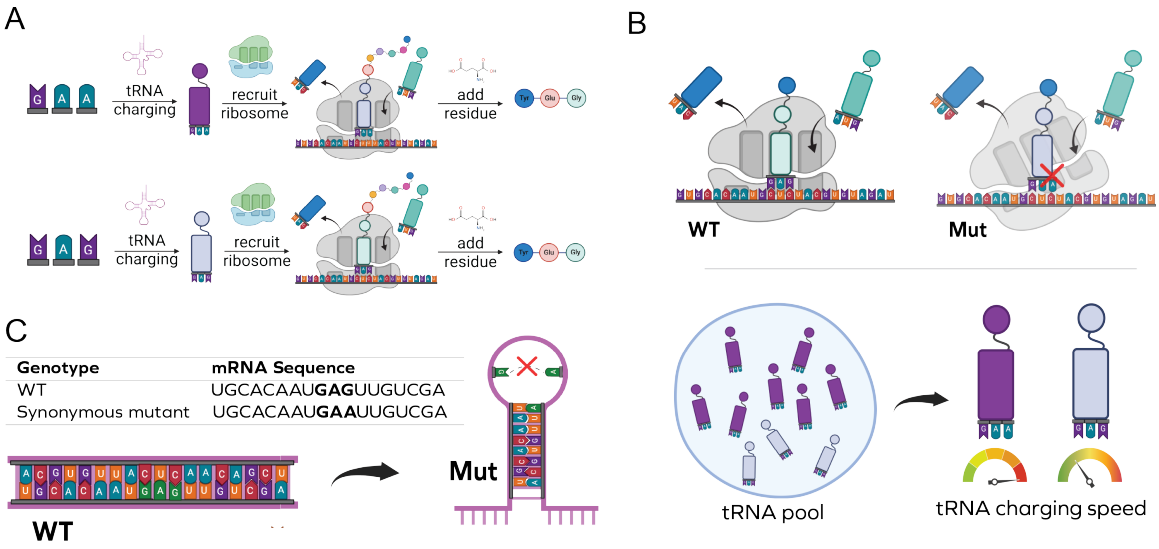

Figure 1: **Synonymous mutations can have significant effects on gene expression, protein folding, and overall cellular fitness. (A)** Synonymous SNP mutations code for the same amino acid but can alter the mRNA sequence. **(B)** Synonymous mutations can also affect the rate of translation elongation by altering codon usage bias. **(C)** Synonymous mutations can impact mRNA secondary structure, influencing stability and translation efficiency.

dows centered on the mutated base, spanning 25 nucleotides upstream and downstream. Both wild-type and synonymous mutant versions of these windows were used to compare structural and translational consequences of each mutation. This context-capturing design ensures that short-range mRNA secondary structures, codon positions, and surrounding motifs are consistently represented and aligned for analysis.

2.2. Feature extraction and annotation

Each mutation window was annotated with features reflecting codon usage bias and mRNA secondary structure. Codon usage was quantified using three metrics: Codon Adaptation Index (CAI), Frequency of Optimal Codons (FOP), and Relative Synonymous Codon Usage (RSCU). CAI, which captures the match between codon usage and optimal expression, is defined as:

$$\text{CAI} = \left(\prod_{i=1}^{L} w_i \right)^{1/L}, \quad w_i = \frac{f_i}{f_{\max, a_i}} \tag{1}$$

where w_i is the relative adaptiveness of codon i, f_i is its observed frequency, and $f_{\max, a_i}$ is the frequency of the most common codon for amino acid a_i.

RSCU compares the observed frequency of a codon to its expected usage among synonymous alternatives:

$$\text{RSCU}_i = \frac{X_i}{\frac{1}{n_a} \sum_{j=1}^{n_a} X_j} \tag{2}$$

We computed all metrics for both wild-type and mutant codons. Additionally, we included the directional difference in CAI:

$$\Delta\text{CAI} = \text{CAI}_{\text{mut}} - \text{CAI}_{\text{wt}} \tag{3}$$

Structural features were derived from thermodynamic folding of each sequence window. The minimum free energy (MFE) was used to quantify RNA stability, and the effect of a mutation on structure was measured by the MFE shift:

$$\Delta\text{MFE} = \text{MFE}_{\text{mut}} - \text{MFE}_{\text{wt}} \tag{4}$$

Codon-level pairing changes were assessed using dot-bracket notation. Each nucleotide of the codon was assigned a binary indicator $b_i \in \{0, 1\}$ representing unpaired or paired status, and changes in pairing triplets $\mathbf{b}_{\text{wt}}, \mathbf{b}_{\text{mut}}$ were used to classify mutations as disruptive, stabilizing, or neutral. Additional features included structural context labels such as "stem," "loop," or "bulge."

2.3. Integration of evolutionary signals

To account for evolutionary constraint, we included conservation scores derived from multi-species align-

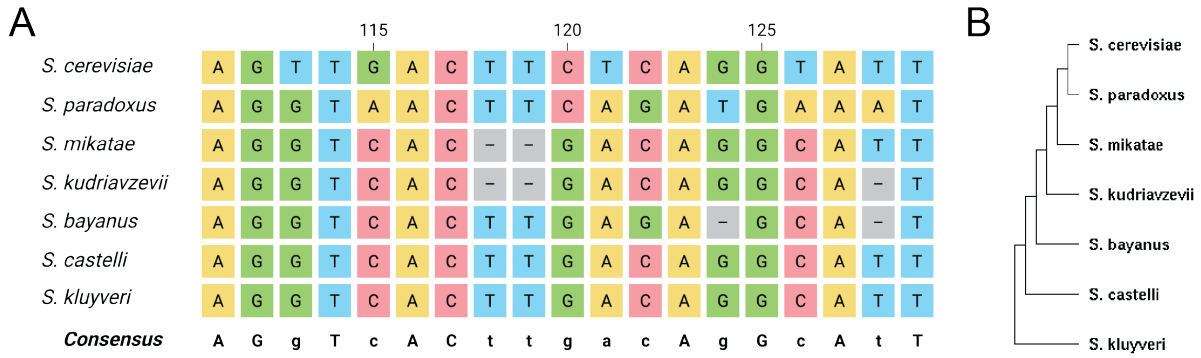

Figure 2: **PhastCons scores indicate evolutionary conservation across species, with higher values suggesting stronger purifying selection. (A)** Multiple sequence alignments reveal conserved regions across species, with PhastCons scores indicating the probability of conservation. **(B)** Phylogenetic tree of yeast species used for alignment.

ments. PhastCons scores were used to represent the probability that a given nucleotide base belongs to a conserved element. These conservation metrics were applied to the central nucleotide of each mutated codon, offering a proxy for purifying selection acting on synonymous sites. Figure 2 illustrates this approach.

2.4. Predictive and interpretable modeling

We used a combination of commonly used classification models to learn the relationship between annotated features and fitness. XGBoost was selected as the primary non-linear model due to its ability to model complex interactions and support interpretability. The objective function minimized the squared loss with regularization:

$$\mathcal{L} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \sum_{k=1}^{K} \Omega(f_k) \tag{5}$$

We also tested a shallow neural network with two hidden layers using ReLU activation and dropout regularization, trained via the Adam optimizer. All models were evaluated using mean squared error (MSE), mean absolute error (MAE), and coefficient of determination (R^2). Further, we employed SHapley Additive exPlanations (SHAP) [4] to interpret model outputs and feature importance.

3. Implementation

All analyses were conducted in Python 3.10 using a modular pipeline built with Jupyter notebooks and standalone scripts.

3.1. Dataset processing and feature matrix construction

The synonymous mutation dataset was obtained from the supplemental materials of Shen et al. (2022) [8]. A custom Python parser extracted the mutation indices, codon identities, and corresponding fitness scores from the raw files. Fitness values were normalized to wild-type levels across three biological replicates. To ensure reliability, we excluded mutations with missing values or high replicate variability, defined as a standard deviation greater than 0.05. After filtering, the final dataset consisted of 4,093 high-confidence synonymous mutations from 22 nonessential *S. cerevisiae* genes. For each retained mutation, we extracted a 51-nucleotide window for sequence context as described previously. Both the wild-type and synonymous mutant versions of each window were stored in FASTA format. This step facilitated direct pairwise comparisons for downstream structure prediction and feature extraction.

All computed features were consolidated into a structured matrix, with each row representing a single synonymous mutation. The matrix included `MFE` values, ΔMFE (Eq.4), base-pairing status, codon usage metrics including `CAI` (Eq.1), `RSCU` (Eq.2), and ΔCAI (Eq.3), conservation scores, and transcript metadata. Continuous features were standardized, and categorical variables such as codon identity were one-hot encoded. Missing values in numeric fields were imputed using feature-wise medians for all models tested except for XGBoost and Random Forest methods. This matrix served as the input to both dimensionality reduction and predictive modeling workflows.

3.2. Feature extraction and annotation

RNA secondary structure predictions were performed using RNAfold from the ViennaRNA Package (v2.5.1). RNAfold was run within a Docker container built from the official ViennaRNA image. RNAfold was executed with default energy parameters and the `--noPS` flag to suppress generation of graphical output. For each input sequence, RNAfold returned a minimum free energy (`MFE`) value in kcal/mol and the corresponding secondary structure in dot-bracket notation.

Structural features were extracted using custom Python scripts. For each pair of wild-type and mutant sequences, we recorded the `MFE` and computed ΔMFE (Eq.4) as a scalar measure of the thermodynamic impact. To extract codon-level features from dot-bracket structures, we identified whether each nucleotide in the mutated codon participated in base-pairing. Based on this, mutations were categorized as structurally neutral, disruptive, or stabilizing. Codons were also annotated with structural contexts—stem, loop, bulge, or unpaired. For a subset of sequences, RNAplfold was used to compute local pairing probabilities and ensemble diversity.

To quantify evolutionary constraint, we annotated each synonymous site with PhastCons scores from the 7-way yeast genome alignment (sacCer3). Conservation scores were retrieved using `pyBigWig` and aligned to the central nucleotide of each codon. Where gaps or mismatches arose, median gene-level scores were imputed. All conservation metrics were standardized before downstream use in correlation and SHAP analyses.

Codon usage bias was quantified using `CAI` (Eq.1) and `RSCU` (Eq.2). The CAI metric was computed for both wild-type and mutant codons using CAIcal [7]. RSCU was calculated manually according to Equation 2 using codon tables from the Kazusa database [5]. Directional effects were captured by ΔCAI (Eq.3). Each codon usage feature was stored independently.

3.3. Dimensionality reduction and predictive modeling

To visualize high-dimensional structure, we applied PCA (`scikit-learn` v1.3.0) prior to modeling. We then trained eight classifiers: XGBoost, Random Forest, k-NN, SVM, logistic regression, Lasso, Naive Bayes, and a shallow neural network. All used the same curated feature matrix. Models were trained on 80/20 train/test splits with z-score standardization, class weighting, and SMOTE-based oversampling to address class imbalance. Evaluation metrics included class-specific precision, recall, F1, accuracy, and ROC AUC. Hyperparameter tuning was done via five-fold cross-validation.

XGBoost performed best on the held-out set, using squared error loss with regularization as defined in Eq.5. Key parameters (η, tree depth, α, λ) were selected through randomized search.

3.4. Model interpretation with SHAP

We used SHAP (`shap v0.41.0`) with TreeExplainer to evaluate feature contributions in the trained XGBoost model. SHAP values were computed using the `Explainer` class [4] to yield interpretable global and local attributions. We generated summary plots, beeswarm plots, and feature dependence curves to visualize how features influenced predictions and to identify dominant biological signals.

4. Results

4.1. Dataset exploration

To evaluate the structure of the annotated feature space and its potential predictive value, we conducted an exploratory analysis visualized in Figure 3. Panel 3A displays a pairwise correlation matrix of selected features, where the diagonal shows kernel density estimates (KDEs) of each feature. These KDEs reveal non-normal, often multimodal distributions across features, indicating underlying biological heterogeneity among synonymous mutations. Notably, codon adaptation index (CAI) and context CAI exhibit distinct peaks, likely reflecting conserved codon usage patterns and local biases in translation efficiency across the yeast genome.

Several notable correlations emerge in Panel 3A. As expected, CAI and context CAI are moderately correlated, capturing overlapping signals of codon optimality. Additionally, minimum free energy (MFE) and its difference upon mutation (MFE_diff) are tightly linked, reflecting that mutations in already stable mRNA regions tend to introduce larger structural disruptions. These trends underscore the interdependence of features and motivate the need for either explicit decorrelation methods or model regularization to account for multicollinearity.

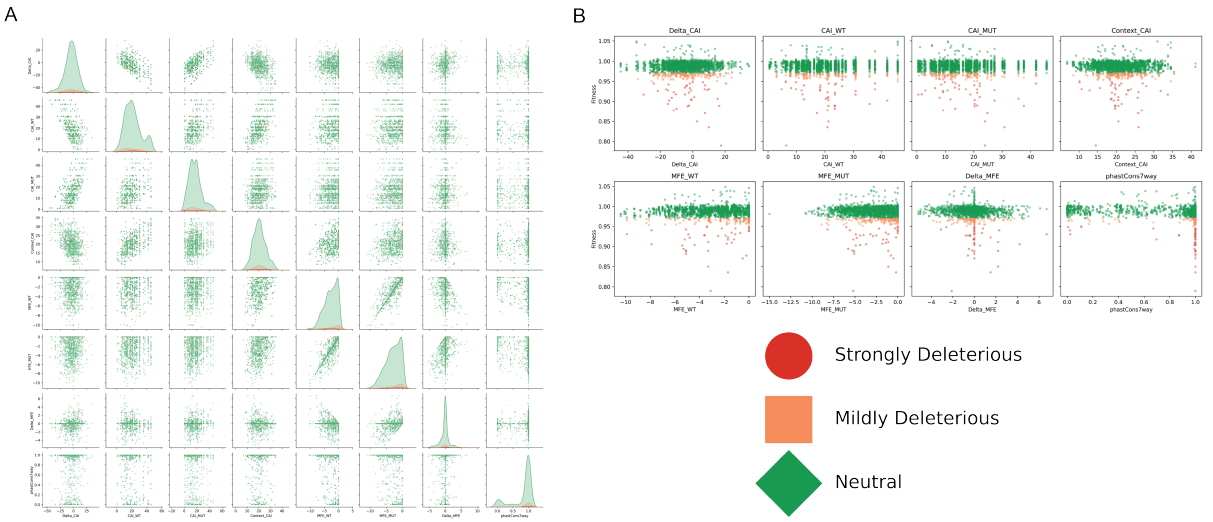

Figure 3: **Visualization of the dataset used for training and testing the model. (A)** Pairwise correlation reveal the distributions and p annotated features for synonymous mutations. The KDE plots (diagonal) show the distribution of each feature. **(B)** Scatterplots show the distribution of each feature against the fitness effect of synonymous mutations.

Panel 3B links these features to fitness outcomes by plotting each feature against measured fitness values, with points colored red, orange, and green by their fitness classes (strongly deleterious, mildly deleterious, or neutral, respectively). Several patterns emerge: CAI and context CAI show clear stratification by class, with strongly deleterious mutations (red) clustering in low-CAI regions, consistent with the biological intuition that rare codons impair translation efficiency. Similarly, high MFE_diff values—indicative of large disruptions to mRNA secondary structure—are enriched among strongly deleterious variants. In contrast, phastCons scores appear positively associated with fitness, suggesting that mutations at evolutionarily conserved sites tend to be less damaging or more robustly buffered.

4.2. Predicting fitness effects of synonymous mutations

To assess model performance, we benchmarked several widely-used traditional machine learning models on a classification task involving three fitness categories: strongly deleterious, mildly deleterious, and neutral synonymous mutations. Performance metrics are visualized in Figure 4.

Panel 4A summarizes the precision (top, blue heatmap) and recall (bottom, red heatmap) achieved by each model across the three classes. While overall classification accuracy was high for ensemble methods like XGBoost, Random Forest, and KNN (all around 91%), this performance was driven predominantly by correct predictions of the neutral class. For example, XGBoost achieved a precision of 0.93 and recall of 0.98 for neutral mutations, but these values declined sharply for mildly deleterious (precision 0.46, recall 0.19) and strongly deleterious mutations (precision 0.40, recall 0.25). This trend is consistent across models, suggesting that class imbalance and overlapping feature distributions present challenges in identifying non-neutral effects.

Panel 4B shows the confusion matrix for the best-performing model by accuracy (XGBoost). It reinforces the difficulty in predicting strongly and mildly deleterious mutations: most mildly deleterious mutations are misclassified as neutral (25/31), and nearly half of strongly deleterious mutations are also predicted as neutral. These misclassifications are critical, as they represent biologically meaningful variants that may be overlooked by automated prediction pipelines. However, the model demonstrates a relatively low false positive rate for neutral predictions, demonstrating that it can effectively filter out benign variants.

4.3. Interpreting feature contributions to predicted fitness effects

To better understand the decision-making process of the classifier, we applied SHapley Additive exPlanations (SHAP) to quantify the influence of each input

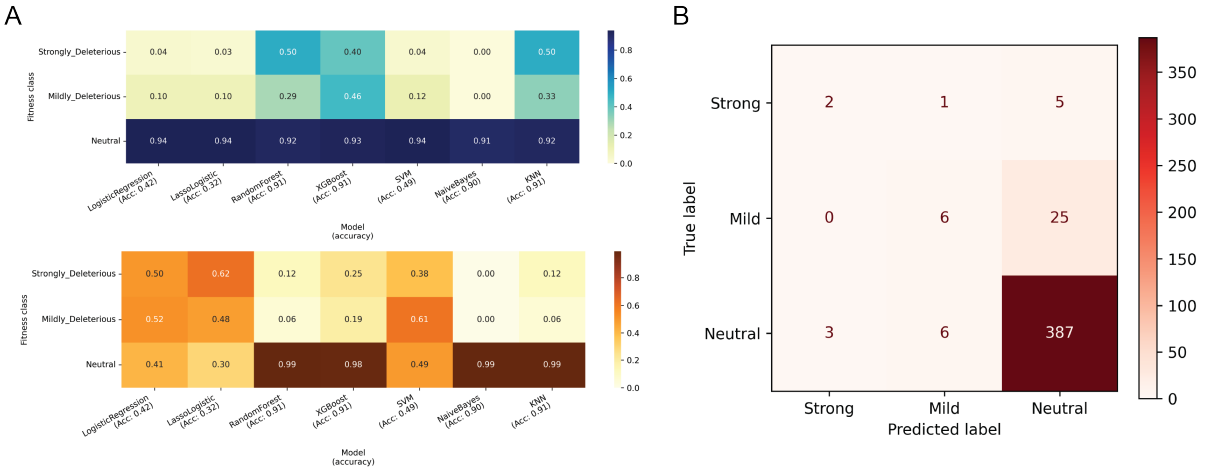

Figure 4: **Ensemble methods improve accuracy, although class imbalance remains a challenge. (A)** Heatmaps depicting the precision (top, blue) and recall (bottom, red) of each classifier's performance on the three classes. **(B)** Confusion matrix for the best-performing model, XGBoost, showing the distribution of true positive and false positive predictions across classes.

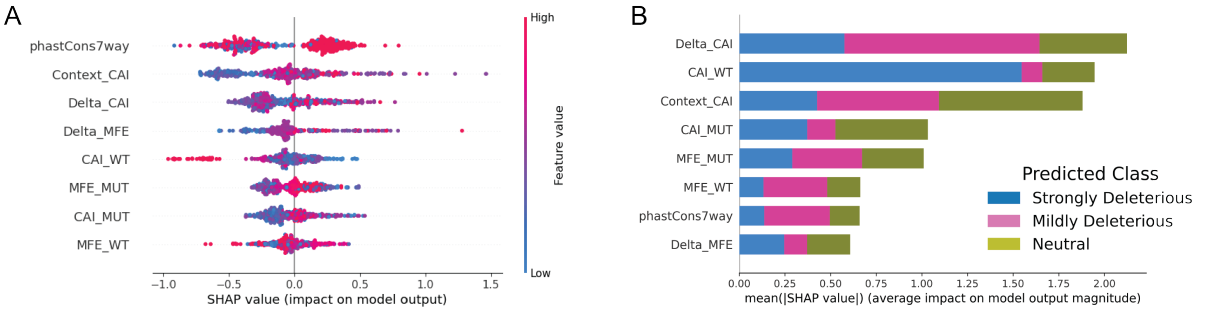

Figure 5: **Codon usage and mRNA structure are the most important features for predicting fitness effects. (A)** SHAP beeswarm plot showing the most important features for predicting fitness effects. Each point represents a single mutation; color indicates the feature value and the horizontal axis shows the SHAP value. **(B)** SHAP summary bar plot showing the mean absolute SHAP value for each feature across the dataset.

feature on model predictions. Figure 5 summarizes the results of this analysis.

Panel 5A presents a SHAP beeswarm plot, which visualizes how each feature contributes to the model's output for individual mutations. Features are ordered by overall importance, with each dot representing a single synonymous mutation. The color gradient reflects the feature value (red = high, blue = low), while the horizontal position corresponds to the SHAP value—i.e., the degree to which a feature pushes a prediction toward a particular fitness class. Notably, codon usage metrics such as CAI and context CAI consistently rank among the top contributors, with high CAI values generally pushing predictions toward neutrality. This supports the biological intuition that optimal codons correlate with minimal disruption to translation and fitness.

mRNA structural features also played a prominent role in determining the classification of synonymous mutations. High values of MFE_diff (ie, large disruptions to secondary structure upon mutation) were strongly associated with predictions of fitness loss. This reinforces the hypothesis that structural destabilization contributes to deleterious effects. Conversely, low MFE_diff values appeared to favor predictions of neutrality, aligning with prior observations that minor structural perturbations are often tolerated.

Panel 5B provides a global summary of feature importance, showing average absolute SHAP values across all predictions. The dominance of codon usage and RNA structural features confirms that these inputs drive the majority of predictive signal in the model. In contrast, evolutionary conservation features such as phastCons appear further down the ranking.

5. Discussion

This study presents an interpretable machine learning framework to predict the fitness consequences of synonymous mutations by integrating codon usage bias, mRNA structural features, and conservation metrics. Our findings reinforce the emerging consensus that synonymous mutations, while silent at the protein level, can carry nontrivial functional effects and thus contribute to the mutational fitness landscape.

Exploratory analyses of the annotated dataset (Figure 3) revealed non-Gaussian, often multimodal feature distributions and correlations. Metrics capturing codon optimality, such as CAI and CAI with local sequence context, exhibited clear separation between fitness classes. Meanwhile, features like MFE and MFE_diff highlighted the role of RNA structural stability. Notably, MFE_diff was higher in strongly deleterious mutations, supporting the hypothesis that disruption of local mRNA structure is a key mechanism of fitness loss. Moderate correlation between codon usage and mRNA stability features further suggests that translation and RNA structure jointly shape mutational effects.

Machine learning classifiers (Figure 4) were then benchmarked on the annotated synonymous mutation dataset. We observed strong performance in predicting neutral mutations, with ensemble models such as XGBoost achieving over 90% overall accuracy. However, this performance was disproportionately driven by the neutral class, which dominated the dataset. Precision and recall dropped markedly for mildly and strongly deleterious mutations, which were frequently misclassified as neutral. This highlights a key limitation; although classifiers readily identify mutations with no discernible effect on fitness, they struggle to detect subtle or context-dependent deleterious signals. We hypothesize that the inconsistent performance is due to overlapping feature space and class imbalance.

To mitigate this, future work may benefit from alternative preprocessing strategies. For instance, constructing a balanced training set by randomly subsampling the dominant neutral class to match the number of mildly and strongly deleterious examples could help equalize class representation without introducing synthetic noise. Alternatively, synthetic resampling techniques such as SMOTE may be replaced or complemented by class-conditional generative models that simulate biologically realistic mutations. Hierarchical classification, where the model first distinguishes neutral from non-neutral mutations before fine-grained subclassification, may also improve interpretability and robustness in the presence of imbalanced and ambiguous labels.

To interpret the model's decision-making process, we applied SHapley Additive exPlanations (SHAP) to quantify the influence of each input feature on predictions (Figure 5). The SHAP beeswarm plot (Figure 5A) revealed that codon usage features—particularly CAI and context CAI—were consistently among the top contributors. We further observed that high CAI values generally associated with predictions of neutrality. This supports the biological intuition that optimal codons are less disruptive to translation and are more likely to be tolerated. Similarly, high MFE_diff values were predictive of lower fitness, consistent with the role of secondary structure stability in mRNA regulation. In contrast, evolutionary conservation metrics such as phastCons appeared less influential (Figure 5B). Perhaps this could be attributed to either the limited resolution of conservation signals at synonymous sites and/or the lack of sensitivity to mechanistic nuance in traditional evolutionary conservation metrics.

Together, these results suggest that synonymous fitness effects arise from overlapping mechanisms involving translation dynamics and RNA folding. Moreover, they suggest that these effects can be captured by integrating biologically grounded features in predictive models. However, the misclassification of deleterious variants underscores the need for models that more effectively disentangle weak signals from background variation. Incorporating ribosome profiling or tRNA abundance data could further enhance model resolution. Additionally, leveraging deep learning approaches that learn hierarchical representations of sequence-structure-function relationships may improve predictive power.

Importantly, our interpretable modeling framework allows us to move beyond black-box predictions and extract mechanistic insight. By quantifying how each feature contributes to fitness prediction, we identify biological mechanisms that are most predictive—and thus potentially most causal—for deleterious fitness effects at synonymous sites. These insights can inform the design of more biologically informed priors in future models of molecular evolution.

Thus, our results underscore the importance of treating synonymous mutations as functionally rele-

vant and mechanistically diverse. As high-throughput mutational scanning data continues to grow, combining interpretable models with diverse molecular annotations will be essential for understanding how seemingly silent changes impact phenotype, fitness, and evolutionary trajectories.

References

[1] JV Chamary, JL Parmley, and LD Hurst. Hearing silence: non-neutral evolution at synonymous sites in mammals. *Nature Reviews Genetics*, 7(2):98–108, 2006.

[2] Richard C Hunt, V Sowmya Simhadri, Mattia Iandoli, Zuben E Sauna, and Chava Kimchi-Sarfaty. Exposing synonymous mutations. *Trends in Genetics*, 30(7):308–321, 2014.

[3] Grzegorz Kudla, Andrew W Murray, David Tollervey, and Joshua B Plotkin. Coding-sequence determinants of gene expression in escherichia coli. *Science*, 324(5924):255–258, 2009.

[4] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 2017.

[5] Yasukazu Nakamura, Takashi Gojobori, and Toshimichi Ikemura. Codon usage tabulated from international dna sequence databases: status for the year 2000. *Nucleic Acids Research*, 28(1):292, 2000.

[6] Joshua B Plotkin and Grzegorz Kudla. Synonymous but not the same: the causes and consequences of codon bias. *Nature Reviews Genetics*, 12(1):32–42, 2011.

[7] Pere Puigbò, Ignacio González Bravo, and Santiago Garcia-Vallvé. Caical: a combined set of tools to assess codon usage adaptation. *Biology Direct*, 3(38), 2008.

[8] Xiaofeng Shen, Sihan Song, Chao Li, and Jin Zhang. Synonymous mutations in representative yeast genes are mostly strongly nonneutral. *Nature*, 606(7915):725–731, 2022.

[9] Tamir Tuller et al. An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell*, 141(2):344–354, 2010.