TireDiff: A Diffusion-based Tire Footprint Image Generation Framework for High-fidelity Prototyping

Sol Lee[1], Jisu Shin[1], Sungrae Hong[1], Chanjae Song[1],
Youngbin You[2], Jeongheon Park[2], Jungsoo Oh[2], Mun Yi[1]*
[1]Korea Advanced Institute of Science and Technology (KAIST)
[2] Hankook Tire & Technology Co.,Ltd, Republic of Korea
{leesol4553, jisu3389, sr5043, chan4535, munyi}@kaist.ac.kr
{ybyou11, jeongheon, jungsoo}@hankooktech.com

Abstract

Recent advances in conditional image generation have revolutionized visual synthesis across various domains, yet manufacturing applications face unique challenges in transforming complex tabular specifications into accurate visual representations. We introduce TireDiff, a framework for sophisticated tire footprint image generation, which is essential for the performance evaluation of newly designed tires. The framework, built upon conditional Latent Diffusion Models, generates tire footprint images directly from manufacturing specifications without requiring their prototypes to be built. Our approach includes a Dual-Stream Embedding method to effectively process hybrid manufacturing data, a physics-aware Tire Custom Loss to ensure physical fidelity, and a Tire-Reference-Free Predictor for quality assessment without ground truth data. Experiments on the TireEval dataset, containing real tire manufacturing data, show strong performance with an average contact error of 6.8% and high structural similarity to ground truth. The proposed framework has the potential to substantially reduce the costs and environmental burdens associated with traditional prototype-based testing in tire development cycles.

1 Introduction

Recent advances in computer vision have substantially improved conditional image generation, enabling precise control over the synthesis process (Zhan et al. 2024). These advancements have shown practical value across various domains, such as art creation and medical imaging, and are also expected to have a significant impact on manufacturing processes (Agnese et al. 2020; Kusiak 2024). In particular, conditional generative models integrate engineering expertise with data-driven insights, facilitating the automation of design processes in product development and prototyping (Howland et al. 2023).

Meanwhile, the tire manufacturing industry has a high demand for prototype production to evaluate performance and safety of real tires (Ridha and Curtiss 2018). This job begins with engineers issuing prototype tire specification tabular data for manufacturing facilities. Then, the engineers capture footprint images, which reveal tire-road interactions, of produced prototype tires under test conditions. These footprint images help engineers evaluate tire performance and iteratively refine the specifications. However, this traditional approach has significant drawbacks: substantial time and cost investments in repeated prototyping, the environmental burden from material and energy waste, and slower innovation cycles in tire development (Balakina et al. 2019).

Several attempts have been made in the industry to address this issue. (Ribeiro et al. 2020) proposed a time delay neural network (TDNN) for real-time estimation of the tire-road friction coefficient (TRFC) using lateral force data, bypassing mathematical models by learning vehicle-road interactions. Similarly, (Pearson, Blanco-Hague, and Pawlowski 2016) introduced TameTire, a physics-based model that predicts tire mechanical and thermal behavior, accurately calculating forces, moments, and thermal states under varying conditions such as slip angle, load, and temperature.

However, these attempts still face subsequent challenges. (1) ***Technological Constraints*** : Previous works have not benefitted from contemporary generative models. While existing studies estimate tire-road interactions, they lack the capability to generate visual representations of contact patterns directly. Consequently, engineers cannot access the visual outputs necessary for design optimization and performance evaluation. (2) ***Effective Tabular Embedding*** : Previous tabular condition embedders, such as one-hot encoding and conventional deep learning-based models, struggle to capture the semantic meaning of variables in text format (Singh and Bedathur 2023; Prokhorenkova et al. 2018; Arik and Pfister 2021; Huang et al. 2020). Although language models are capable of conditioning semantic meanings, their ability to encode fine-grained differences in numerical values is still limited (Singh and Bedathur 2023; Naseem et al. 2021; Thawani et al. 2021). (3) ***High Physical Fidelity*** : Unlike generating general domain images, tire footprint images must accurately reflect physical properties for prototype tire performance evaluation. (4) ***Absence of Ground Truth on Test*** : Since the primary goal of prototype image generation is virtual simulation rather than physical tire production, it is difficult to evaluate the quality of generative model outputs in practice (*i.e.*, there are no ground truth images for comparison).

To address these limitations, we propose TireDiff, a framework based on Latent Diffusion (Rombach et al.

[*] Corresponding Author

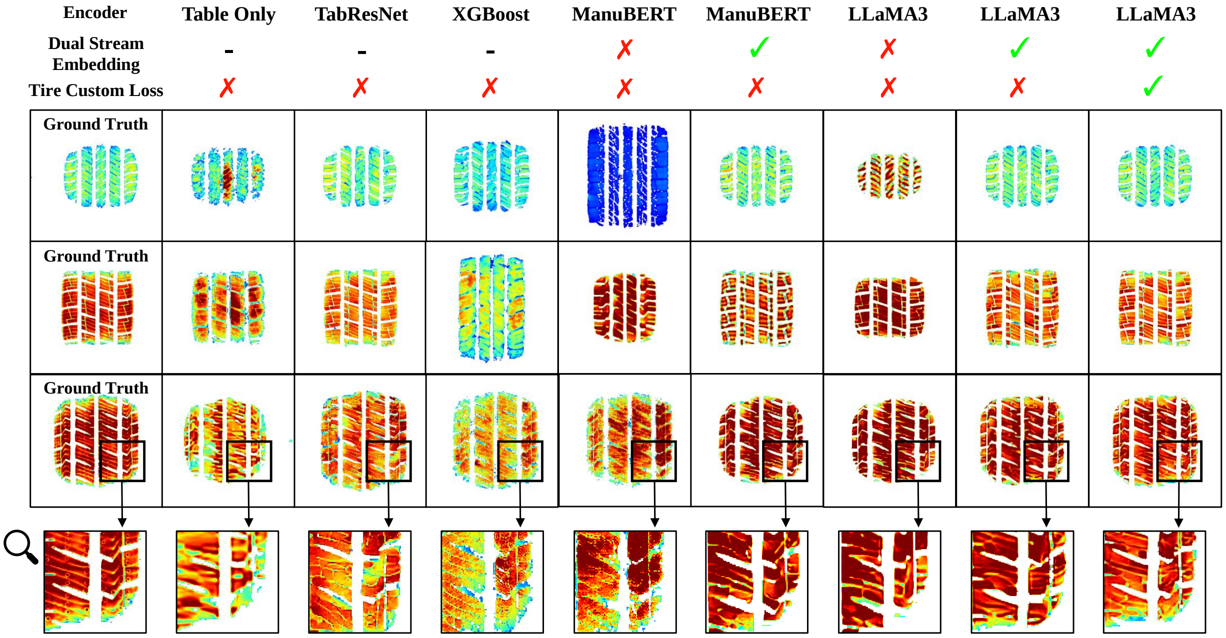


Figure 1: **Qualitative comparison of different encoder architectures for tire footprint generation.** The results demonstrate the effects of Dual-Stream Embedding and Tire Custom Loss applications. Note that Dual-Stream Embedding is applicable to language model-based encoders (ManuBERT and LLaMA3 (Dubey et al. 2024)) and is marked as '-' for others. All images are visualized with Pressure distribution colormap (red: high pressure, blue: low pressure) over gray footprint images, as detailed in Figure 6. Best view in color.

2022) for generating tire footprint images, without prototype production, from the tire specifications and test conditions. Through exploring optimal embedding approaches for tabular-to-image generation, we found that processing numerical data separately while using a language model to extract semantic features from variable names significantly outperforms feeding all data directly into language models or conventional deep learning-based models. Additionally, including variable descriptions helps improve the quality of tire footprint generation. To enhance the physical fidelity of generated footprints, we propose a Tire Custom Loss that explicitly optimizes key physical attributes such as contact length, width, and area, thereby improving both realism and utility. Furthermore, we develop the Tire-Reference-Free Predictor (TRFP), which quantifies the similarity between generated footprint images and predicted physical characteristics without requiring ground truth footprints. This framework demonstrates a practical application of generative models in tire manufacturing while presenting a new direction for sustainable innovation in product design and prototype automation across the manufacturing industry.

We propose our contributions as follows:

1. We propose TireDiff, a framework for generating tire footprint images from tabular specifications and test conditions. It enables rapid tire performance evaluation without physical prototyping, reducing both costs and environmental impact.
2. Through exploring optimal tabular embedding strategy, we observed that combining numerical values separately with language model-based semantic extraction using variable descriptions performs best, which we term Dual-Stream Embedding. And, we introduce Tire Custom Loss to preserve physical properties in generated images
3. We present Tire-Reference-Free Predictor (TRFP) for reference-free quality assessment, tackles that it is difficult to evaluate the generated image because there is no ground truth in the field.
4. The effectiveness of the framework has been validated through experiments using real tire manufacturing data, providing valuable insights for future research.

2 Related Work

2.1 Tire-Road Interaction Estimation

Tire-road interaction significantly affects vehicle stability and fuel efficiency, key factors in tire performance evaluation. Researchers have explored various methods to predict and analyze tire behavior under road conditions. (Ribeiro et al. 2020) developed a time delay neural network (TDNN) to estimate tire-road friction coefficient from lateral force data, bypassing traditional models. (Pearson, Blanco-Hague, and Pawlowski 2016) introduced TameTire, a physics-based model, and estimated thermal behavior under varying conditions such as slip angle and load. (Barbosa et al. 2021) developed a lateral force prediction model using Gaussian Process Regression (GPR) based on intelligent tire data, achieving reliable performance even at high slip angles.

However, these approaches primarily rely on numerical data or physical measurements without generating visual representations. We aim to predict footprint images, enabling more intuitive analysis of tire-road interactions and improving design optimization through visual insights.

2.2 Conditional Diffusion Model

Conditional generative diffusion models learn data distribution through noise addition and removal via Markov chains (Dhariwal and Nichol 2021; Ho, Jain, and Abbeel 2020) using condition embeddings. The methods can be categorized into pixel and latent space diffusions. (1) *Pixel*: GLIDE (Nichol et al. 2021) uses guided diffusion and classifier-free guidance with a transformer text encoder for high-resolution image generation, while Imagen (Saharia et al. 2022) optimizes computation with a pre-trained language model. (2) *Latent*: Stable Diffusion uses VQ-GAN (Esser, Rombach, and Ommer 2021) to generate latent representations, learning the reverse process for realistic image creation. DALL-E2 (Ramesh et al. 2022) employs a multimodal contrastive model CLIP (Radford et al. 2021) to encode images and text in the same space.

Despite their use in a variety of applications, there has been limited research in the tire industry, posing implementation challenges. This study aims to develop a conditional generation model for tire footprint images to enhance manufacturing efficiency.

2.3 Tabular Data Embedding Techniques

Machine Learning and Deep Learning Approaches Various machine learning and deep learning approaches have been proposed for tabular data embedding, showing consistent performance in classification and regression tasks. Decision Trees (Song and Ying 2015) capture nonlinear relationships through recursive splitting, while XGBoost (Chen and Guestrin 2016) excels in predictive accuracy and speed via gradient boosting. CatBoost (Prokhorenkova et al. 2018) simplifies preprocessing by automating categorical data encoding. These tree-based models provide robust results and interpretability, even on small datasets (Singh and Bedathur 2023). Their strength in handling numerical features makes them effective for tasks requiring precise numerical processing, though they have limitations in incorporating text information, like variable name meanings.

Meanwhile, deep learning methods can be divided into non-transformer and transformer-based approaches. Non-transformer models like TabMLP (Zaurin and Mulinka 2023) use basic neural networks for continuous features, while TabResNet (Zaurin and Mulinka 2023) adds residual connections to enhance learning. In contrast, transformer-based models such as TabNet (Arik and Pfister 2021) and FT-Transformer (Huang et al. 2020) leverage attention mechanisms to capture complex interactions in the data. These models excel at identifying complex patterns in numerical features, though they share tree-based models' limitations in incorporating textual information.

Language Model Approaches Language models have emerged as a powerful approach for transforming tabular data into embeddings by treating it as structured text sequences. Table2Vec (Zhang, Zhang, and Balog 2019) pioneered this approach, which adapted Word2Vec's skip-gram methodology (Radford et al. 2021) for table retrieval. BERT (Devlin 2018) marked a significant advancement in this field, leveraging transformer architecture self-trained on extensive text corpora. BERT-based methods generate Table Vectors through systematic row linearization (e.g., "[**column name**] | [**column type**] | [**cell value**] [SEP]"), demonstrating effectiveness in table classification tasks (Koleva et al. 2021; Agarwal et al. 2019). These developments have led to domain-specific variants, including ManuBERT (Kumar, Starly, and Lynch 2023) for manufacturing applications and MedBERT (Rasmy et al. 2021) or medical applications.

While these language models excel in processing textual information, research reveals limitations in numerical data handling (Sennrich 2015; Wu 2016). Given the numerical processing challenges, we explore embedding techniques for manufacturing tabular data containing both precise numerical information and textual descriptions. As embedding quality directly impacts conditional image generation performance, we comprehensively evaluate various embedding methods to find an optimal approach that effectively captures both types of information.

3 Method

This section introduces the key components of our tire footprint generation framework: exploration of tire tabular data embedding strategies, conditional latent diffusion model, physics-aware loss function, and Tire-Reference-Free Predictor. The framework is illustrated in Figure 2.

3.1 Embedding Strategy for Tire Specification-Condition Tabular Data

Dataset Description Our framework leverages the TireEval dataset from *Hankook Tire & Technology Co.,Ltd, Republic of Korea*. The TireEval dataset consists of tire specification with 82 variables (*e.g.*, including dimensions, tread design, and mold specifications) and 3 test conditions (*i.e.*, load, inflation pressure, and rim width), pairing real footprint images. We evaluated 843 unique tires under various combinations of these test conditions, generating 15,524 samples with their corresponding footprint images. The dataset was split into 11,838 training and 3,686 testing sets. Figure 3. depicts the TireEval dataset.

Embedding Strategies for Language Model We propose a new row linearization strategy that incorporates detailed column descriptions, based on (Yin et al. 2020) and (Chen et al. 2019). For example, the column [**Bead Type**] has the description [**the shape and structure of the tire bead, influencing wheel fit and driving stability**]. These descriptions are formatted into two structured ways: "[**column name**] ([**column description**]) | [**column type**] | [**cell value**]" with [SEP] tokens (Eq.(1)), and "[**column name**] ([**column description**]) is [**cell value**]" with semicolons (;) as separators (Eq.(2)). In this work, we denote these linearizations as [Eq.(1)] and [Eq.(2)]. Furthermore, for nu-

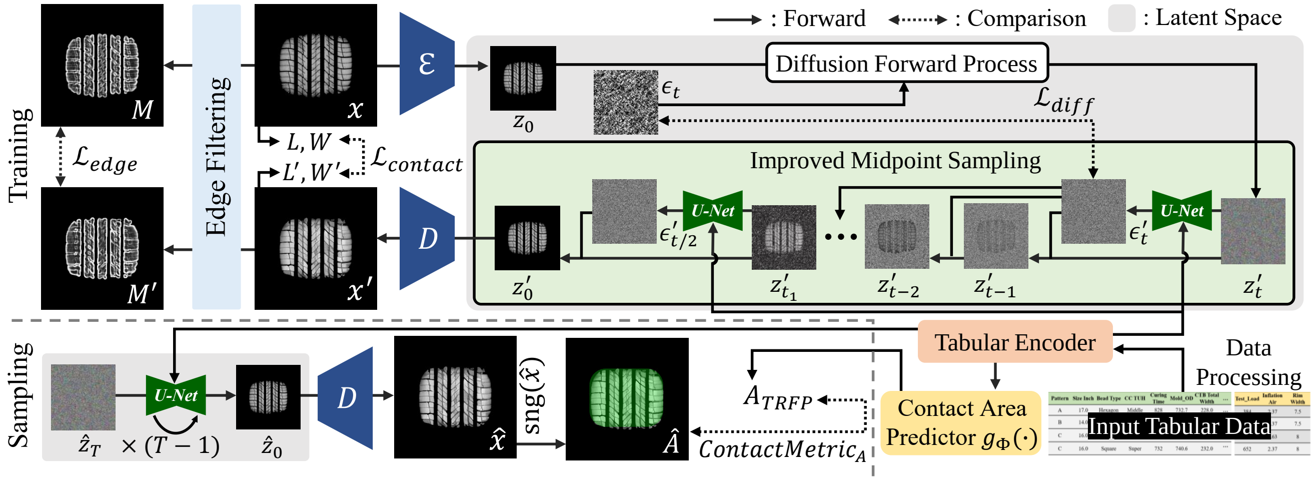
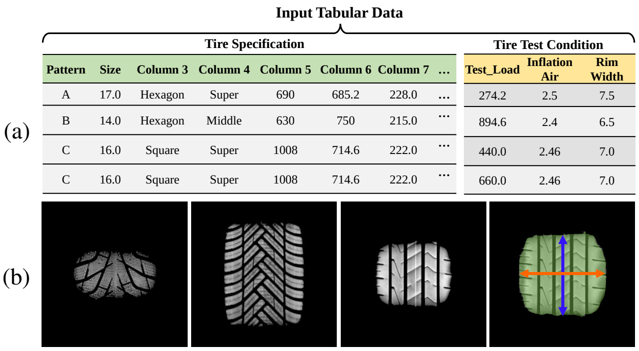


Figure 2: **Overview of the proposed TireDiff framework** Although the diffusion model operates in the latent space during training and testing, we visualize all processes in the original image space for clarity.



Input Tabular Data

Tire Specification								Tire Test Condition		
Pattern	Size	Column 3	Column 4	Column 5	Column 6	Column 7	...	Test_Load	Inflation Air	Rim Width
A	17.0	Hexagon	Super	690	685.2	228.0	...	274.2	2.5	7.5
B	14.0	Hexagon	Middle	630	750	215.0	···	894.6	2.4	6.5
C	16.0	Square	Super	1008	714.6	222.0	···	440.0	2.46	7.0
C	16.0	Square	Super	1008	714.6	222.0	···	660.0	2.46	7.0

Figure 3: **Example of the TireEval dataset**: (a) Tire specifications and test conditions in tabular format. (b) Corresponding tire footprint images, where ↔ shows the contact width W, ↕ indicates the contact length L, and the green area denotes the contact area $A = \mathrm{sgn}(x)$.

merical variables, which language models have limitations in processing, we propose a straightforward but effective approach named Dual-Stream Embedding. It simply concatenates column-wise normalized real numerical value of table next to Eq.(1) or Eq.(2). In this case, we denote Dual-Stream Embedding on [Eq.(1)] and [Eq.(2)] as [Eq.(1)||Numeric] and [Eq.(2)||Numeric] respectively. Through experiments, we demonstrate two key improvements: the new row linearization strategy which incorporates column descriptions outperforms conventional approaches, and the Dual-Stream Embedding improves upon the previous encoding method using language models, which were strong in textual information but less robust to precise numerical information (Singh and Bedathur 2023; Naseem et al. 2021).

$$\underbrace{\texttt{Bead Type}\ \Big(\texttt{the shape and ...}\Big)}_{\text{Column Name (Column Description)}} \Big|\ \underbrace{\texttt{text}}_{\text{Column Type}}\ \Big|\ \underbrace{\texttt{Hexagon}}_{\text{Cell Value}} \tag{1}$$

$$\underbrace{\texttt{Bead Type}\ \Big(\texttt{the shape and ...}\Big)}_{\text{Column Name (Column Description)}}\ \texttt{is}\ \underbrace{\texttt{Hexagon}}_{\text{Cell Value}} \tag{2}$$

3.2 Tire Footprint Image Generation Model

Preliminary: Stable Diffusion Stable Diffusion (Rombach et al. 2022) generates images by progressively denoising noise in the latent space of a pre-trained VAE (Kingma 2013). In this study, we extend this framework by converting input tabular data into embedding $\mathbf{e}$ to conditionally generate tire footprint images. The model utilizes a fixed encoder $\mathcal{E}$ to transform the input image x into latent features; $z_0 = \mathcal{E}(x)$ and decoder $D(\cdot)$ to generate estimated image $\hat{x}$. The diffusion forward process is predefined according to variance schedule β_t, where $t \in \{1, \ldots, T\}$ indicates the steps of the process, with T being the total number of steps. We denote $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$. This process is described by the following equation:

$$q(z_t \mid z_0) = N(z_t; \sqrt{\bar{\alpha}_t} z_0, (1 - \bar{\alpha}_t)\,\mathbf{I}), \tag{3}$$

The noise prediction network $\epsilon_\theta(\cdot)$ employs a U-Net (Ronneberger, Fischer, and Brox 2015) architecture, which uses t and embedding $\mathbf{e}$ as conditions. The training loss is given by:

$$\mathcal{L}_{diff} = \mathbb{E}_{\mathcal{E}(x),\mathbf{e},\epsilon\sim N(0,1),t}\left[\left\|\epsilon - \epsilon_\theta\left(z_t, t, \mathbf{e}\right)\right\|_2^2\right], \tag{4}$$

Custom Loss Function for Tire Footprint Image Generation Training a diffusion model for tire footprint generation requires accurate reflection of tire physical properties. Unlike general image generation tasks, our application demands precise representation of physical features - contact length, width, and area - which are critical metrics for performance evaluation (Figure 5(b)). These properties directly influence tire grip, stability, and durability, and their inaccurate representation would compromise the model's reliability and utility. Therefore, we designed a custom loss function called Tire Custom Loss to accurately reflect the physical properties of tire footprints. This loss integrates several components that represent important physical characteristics of the tire footprint, as outlined below:

- **Contact Length and Width Loss**: This term measures the absolute difference between the contact length L_i and width W_i of the ground truth image x, and the contact length L_i' and width W_i' of the generated image x', where N is the number of samples in a batch:

$$\mathcal{L}_{contact} = \frac{1}{N} \sum_{i=1}^{N} (|L_i - L_i'| + |\, W_i - W_i'|) \tag{5}$$

- **Edge Preservation Loss**: This term computes edge differences between x and x' using a Sobel filter (Kanopoulos, Vasanthavada, and Baker 1988) to generate fine details. The Sobel filter detects gradients in horizontal and vertical directions to enhance edges, enabling precise identification of contact surface details. The edge maps M_i and M_i' are extracted from the x and x', respectively:

$$\mathcal{L}_{edge} = \frac{1}{N} \sum_{i=1}^{N} |M_i - M_i'| \tag{6}$$

- **Tire Custom Loss**: The Tire Custom Loss function is defined as a weighted sum of the $\mathcal{L}_{contact}$, and the $\mathcal{L}_{edge}$ with weights $\lambda_{contact}$ and λ_{edge}, respectively:

$$\mathcal{L}_{tire} = \lambda_{contact} \times \mathcal{L}_{contact} + \lambda_{edge} \times \mathcal{L}_{edge} \tag{7}$$

Finally, the final loss term $\mathcal{L}_{tirediff}$, called tire diffusion loss, is defined as follows:

$$\mathcal{L}_{tirediff} = \mathcal{L}_{diff} + \mathcal{L}_{tire} \tag{8}$$

Midpoint Sampling for Training Since $\mathcal{L}_{tire}$ requires generated image $x' = D(z_t')$, z_t' must be denoised to z_0' during training. In the original DDPM (Ho, Jain, and Abbeel 2020), generating z_0' requires multiple sampling steps across random timesteps t, leading to a high computational cost. To address this, we adopted an improved midpoint sampling method from (Liu et al. 2024) that generates an approximate z_0' in just two U-Net inferences. The closed form of forward process is derived as Eq.(9) from Eq.(3). If we assume trained $\epsilon_\theta(\cdot)$ estimates proper noise at t, Eq.(9) can be depicted as Eq.(10). At the same time, based on Bayesian rules, Eq.(11) is derived from posterior $q(z_{t-1} \mid z_t)$. It shows the inverse process from z_t and z_{t-1} which is allowed when we move to z_0, starting from z_t. Therefore, improved midpoint sampling consecutively denoises from z_t' to z_{t_1}', where $t_1 = \lfloor t/2 \rfloor$. $t - t_1$ times of repetition estimates z_{t_1}' and now we can compute z_0' using Eq.(10) from z_{t_1}' directly.

$$z_t = \sqrt{\bar{\alpha}_t} z_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon, \text{where } \epsilon \sim \mathcal{N}(0, \mathbf{I}) \tag{9}$$

$$z_0 = \frac{z_t - \sqrt{1 - \bar{\alpha}_t}\epsilon_\theta(z_t, t, \mathbf{e})}{\sqrt{\bar{\alpha}_t}} \tag{10}$$

$$\begin{aligned} q(z_{t-1} \mid z_t, z_0) = \mathcal{N}(z_{t-1}; \tilde{\mu}_t(z_t, z_0), \tilde{\beta}_t \mathbf{I}), \\ z_{t-1} = \tilde{\mu}_t(z_t, z_0) + \sqrt{\tilde{\beta}_t}\epsilon, \\ \text{where } \begin{cases} \tilde{\mu}_t(z_t, z_0) = \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t} z_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t} z_t \\ \tilde{\beta}_t = \frac{1-\alpha_{t-1}}{1-\bar{\alpha}_t}\beta_t \end{cases} \end{aligned} \tag{11}$$

3.3 Tire-Reference-Free Predictor

In practical applications of tire footprint generation models, ground truth images are typically unavailable, making quality evaluation challenging. This limitation significantly impacts the decision-making process in tire design and manufacturing, where reliable quality assessment is crucial. To address this, we propose a Tire-Reference-Free Predictor for quantitatively assessing generated images without reference data. Taking inspiration from dialogue quality assessment using pre-trained language models (Mehri and Eskenazi 2020) and document summarization evaluation using unsupervised models (Lee, Park, and Kang 2024), we train a separate model to predict contact area from tire specifications and test condition embeddings (*i.e.*, input tabular data). The prediction model $g_\Phi(\cdot)$ consisted of a simple architecture of 9-layer MLP with ReLU activation functions. The predicted A_{TRFP} from $g_\Phi(\cdot)$ is used to calculate $ContactMetric_A$ in Eq. (12), which is described later, when the ground truth x is not available.

4 Experiment Setup

4.1 Tabular Encoder Setup

We trained several tabular encoders using the TireEval dataset. TabMLP, TabResNet, and TabNet followed the (Arik and Pfister 2021) approach, utilizing a symmetric encoder-decoder that masks and reconstructs input features to learn feature importance. FT-Transformer was trained with contrastive learning, incorporating row-level attention to learn feature interactions, based on the SAINT (Somepalli et al. 2021). The aforementioned encoders were implemented using the PyTorch-WideDeep library (Zaurin and Mulinka 2023). XGBoost and CatBoost were trained via self-supervised learning using temporary labels generated by K-means clustering, based on libraries [1,2]. On the other hand, we leveraged Hugging Face Transformer library (Wolf 2019) to adapt pre-trained BERT, ManuBERT parameters. We used the original LLaMA3 with the Ollama platform [3]. Table Only method was used as a baseline, which combines categorical features with one-hot encoding and numerical features without an encoder.

[1] https://catboost.ai/

[2] https://xgboost.readthedocs.io/en/stable/

[3] https://ollama.com/

Tabular Encoder	Row Linearization	CL ↓	CW ↓	CA ↓	C-Avg ↓	LPIPS ↓	SSIM ↑
Table Only	-	9.7029	7.4140	19.6410	12.2526	0.2068	0.7430
TabMLP	-	8.9750	5.5659	15.1719	9.9043	0.1844	0.7579
TabResnet	-	8.7600	5.5050	14.9297	9.7316	0.1858	0.7578
TabNet	-	9.0070	6.1469	16.0829	10.4123	0.2257	0.7326
FT-Transformer	-	12.6043	6.9277	16.2259	11.9193	0.2769	0.6686
CatBoost	-	37.7531	10.7431	67.9539	38.8167	0.2769	0.6686
XGBoost	-	38.7217	10.5187	69.8111	39.6838	0.2811	0.6648
BERT	Eq.(2)||Numeric	9.7330	6.8108	17.1705	11.2381	0.1870	0.7581
ManuBERT	Eq.(2)||Numeric	8.2972	6.1392	14.5364	9.6576	0.1667	0.7596
LLaMA3	Eq.(2)||Numeric	**7.2755**	**4.8211**	**12.2013**	**8.0993**	**0.1372**	**0.7614**

Table 1: **Quantitative Performance Comparison of Various Encoders in Tire Footprint Image Generation**

4.2 Tire Footprint Generation Model Setup

All images were resized to 256 × 256, and the Stable Diffusion architecture was used. The pre-trained VAE on Stability AI was used to reconstruct high-quality images. For all experiments, we used batch size 64, learning rate $1e-4$, and AdamW optimizer (Loshchilov and Hutter 2017). The noise steps were set to 1k, and training was conducted for 1.5k epochs. For experiments involving $\mathcal{L}_{tire}$, we set the $\lambda_{contact} = 0.05$ and $\lambda_{edge} = 0.1$. All experiments were conducted on a single NVIDIA® RTX A6000 with 48GB memory.

4.3 Evaluation Metrics

We evaluated our tire footprint generation model using both physical property metrics and visual similarity metrics.

Contact Metrics It measures the physical properties of generated tire footprint images that influence vehicle stability and performance, as illustrated in Figure 5(b). The mean absolute percentage error evaluates the alignment between $\hat{x}$ and x using $\hat{L} \leftrightarrow L$, $\hat{W} \leftrightarrow W$, and $\hat{A} \leftrightarrow A$. In Eq.(12), C can be one of L, W and A. Here, average contact metric C-Avg $= \frac{1}{3} \times (\sum_{C\in\{L,W,A\}} ContactMetric_C)$ is used to assess the overall physical properties of the tire footprint.

$$ContactMetric_C = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\hat{C}_i - C_i}{C_i}\right| \times 100 \qquad (12)$$

Visual Similarity Metrics We employ two complementary metrics to evaluate the visual quality of generated images. LPIPS (Zhang et al. 2018) measures perceptual similarity using deep neural network features with a pre-trained VGG (Simonyan 2014) network, where lower scores indicate better quality. SSIM (Wang et al. 2004) evaluates structural similarity by considering luminance, contrast, and structural information, with higher scores indicating better quality. These metrics provide a comprehensive assessment of both perceptual and structural aspects.

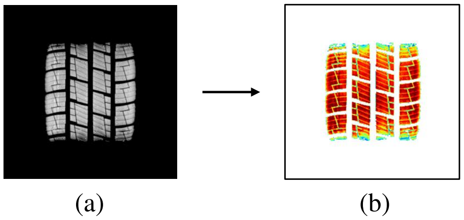


Figure 6: **Tire footprint pressure colorization**: (a) grayscale image and (b) pressure colormap. The Jet colormap shows pressure intensity from low (blue) to high (red), calculated by distributing load proportional to pixel values $[0, 255]$. This visualization is used only for evaluating generated images.

5 Experiment Results

5.1 Results of Embedding Strategy Exploration

Quantitative Results Table 1 compares the tire footprint image generation performance across different encoders. LLaMA3 achieves the best performance, with the lowest C-Avg of 8.0993, LPIPS of 0.1372, and an SSIM of 0.7614, demonstrating superior image quality and structural similarity. Trained on approximately 15 trillion tokens from academic papers, books, and Wikipedia, LLaMA3 effectively embeds tire variable meanings and descriptions, leading to excellent generation performance. The second-best performer, ManuBERT, achieves a C-Avg of 9.6576, LPIPS of 0.1761, and SSIM of 0.7596, effectively capturing physical properties through its manufacturing-domain training. These language model-based methods demonstrate superior performance through their ability to process variable names and textual descriptions. The performance metrics in Table 1 present each language model's results using its optimal configuration, selected from Eq.(1), Eq.(2), and their Dual-Stream Embedding variants, [Eq.(1)||Numeric] and [Eq.(2)||Numeric].

Among non-language model-based methods, TabResNet shows the best performance with a C-Avg of 9.7316, but its LPIPS (0.1858) and SSIM (0.7578) are relatively low, indicating limitations in image quality. Traditional tree-based models like CatBoost and XGBoost show limitations in

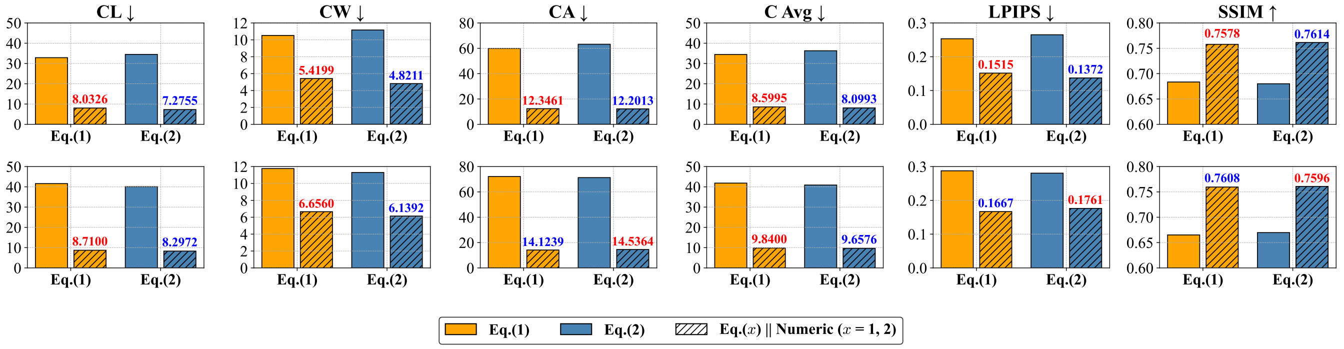


Figure 7: **Impact of Dual-Stream Embedding on Tire Specification Processing Performance** Results display LLaMA3 (first row) and ManuBERT (second row) performance metrics, with the best and second-best values for each metric highlighted in blue and red respectively.

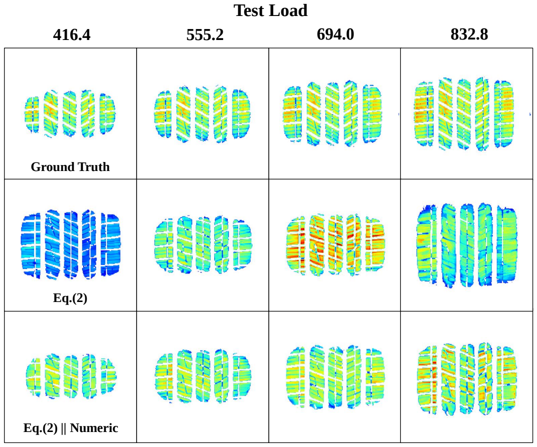


Figure 8: **Visualization of Tire Contact Area Under Varying Load Conditions**

high-dimensional feature learning, resulting in lower performance across metrics.

Qualitative Results Figure 1 shows the pressure distribution visualizations (detailed in Figure 6) of generated tire footprint images. Due to space constraints, we present results from the baseline Table Only model, the best-performing language models (LLaMA3 and ManuBERT), and the top non-language model (TabResNet) and XGBoost. Compared to the ground truth, LLaMA3 achieved the highest quality with detailed tread patterns and accurate pressure distributions, while ManuBERT showed similar performance but with slight limitations in pressure details. TabResNet captured overall patterns but struggled with fine details, and XGBoost produced blurred and distorted results. The Table Only approach struggled in both aspects.

5.2 Analysis of Dual-Stream Embedding Effects

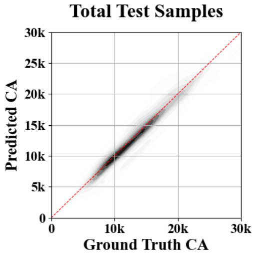
Building on the superior performance of language model-based approaches, we conducted an in-depth analysis of the Dual-Stream Embedding strategy. We compared two linearization embedding techniques ([Eq.(1)], [Eq.(2)]) with and without Dual-Stream Embedding strategy using both LLaMA3 and ManuBERT models.

Tire Custom Loss	Variable Description	CL ↓	CW ↓	CA ↓	C-Avg ↓	LPIPS ↓	SSIM ↑
✗	✗	8.4187	5.7318	13.0957	9.0821	0.1599	0.7574
✗	✓	7.2755	4.8211	12.2013	8.0993	0.1372	0.7714
✓	✗	7.4418	5.1339	12.5177	8.3645	0.1412	0.7981
✓	✓	**6.2437**	**4.0294**	**10.1817**	**6.8183**	**0.1241**	**0.8107**

Table 2: **Ablation studies on variable descriptions and Tire Custom Loss**

Our qualitative analysis showed that the two linearization techniques exhibited relatively similar performance before applying Dual-Stream Embedding. However, the application of Dual-Stream Embedding led to substantial performance improvements in both models. While LLaMA3's [Eq.2)||Numeric] embedding achieved the best performance (C-Avg: 8.0993, LPIPS: 0.1372, SSIM: 0.7614), the baseline [Eq.(2)] embedding showed significant degradation (C-Avg: 36.2806, LPIPS: 0.2650, SSIM: 0.6800). Similar patterns were observed in ManuBERT. These results demonstrate the limitations of language models in processing numerical tire specifications while validating the effectiveness of our simple yet intuitive dual-stream approach.

Additionally, we evaluated the Dual-Stream Embedding effects through load variations. Load is a key numerical variable that represents the vehicle weight applied to the tire and determines tire performance by reflecting vehicle-road interactions (Padula 2005). As shown in Figure 8 (Result of LLaMA3), Ground Truth demonstrated gradual expansion of tire contact area as the load increased from 416.4 to 832.8 units. The [Eq.(2)||Numeric] embedding precisely captured these contact area changes, showing high consistency. In contrast, [Eq.(2)] embedding failed to effectively capture the dynamic changes under load variations. This analysis demonstrates the contribution of Dual-Stream Embedding to reliable prediction of tire dynamic behavior under varying load conditions.



	Total Test Samples
N of samples	3,686
R^2	0.9311
Pearson Corr	0.9672
Mean MAPE	7.0934%

Figure 9: **Performance of the Tire-Reference-Free Predictor** (Left) Density plots of predicted vs. ground truth contact area, with the red line representing ideal predictions (Right) Performance metrics.

5.3 Ablation Studuies

To validate our framework, ablation studies were conducted on two key components: variable description embedding and custom tire loss function (Table 2).

Effect of Variable Description Inclusion Building upon our findings with the Dual-Stream Embedding strategy, we investigated the impact of variable descriptions on LLaMA3's embedding performance using its optimal [Eq.(2)||Numeric] configuration. When descriptions were excluded, inputs were simplified to the format "[**column name**] ([**column description**]) is [**cell value**]". Removing descriptions degraded performance across all metrics, with C-Avg increasing by 1.1432, LPIPS by 0.0227, and SSIM decreasing by 0.014 without Tire Custom Loss. These results demonstrate that variable descriptions play a crucial role in enhancing the model's semantic understanding of tire specifications, contributing to better embedding quality and generation performance.

Effect of Applying Tire Custom Loss We further evaluated our proposed custom tire loss function. With variable descriptions included, the improvements were more pronounced: C-Avg decreased by 2.0196, LPIPS by 0.0171, and SSIM increased by 0.0393. These improvements are reflected in Figure 1, where generated images show more refined tread patterns and precise pressure distributions, validating our custom loss design for tire manufacturing applications. The results demonstrate that our physics-informed loss function effectively guides the model to preserve critical tire characteristics during the generation process.

5.4 Analysis of Tire-Reference-Free Predictor

We analyzed the effectiveness of the Tire-Reference-Free Predictor by training a contact area prediction model $g_\Phi(\cdot)$ using TireEval tabular data for 300 epochs with Mean Squared Error (MSE) loss. The model's performance is summarized in fig:fig7. The $g_\Phi(\cdot)$ achieved an R^2 value of 0.9311 and a Pearson correlation coefficient of 0.9672, demonstrating a strong agreement between the predicted and ground truth contact areas. Furthermore, the model exhibited a low Mean Absolute Percentage Error (MAPE) of 7.0934%, indicating its effectiveness in predicting contact area across the entire test set. These results suggest that our Tire-Reference-Free Predictor can serve as a reliable quality assessment tool for tire manufacturers, potentially reducing the need for physical prototyping in the early design stages.

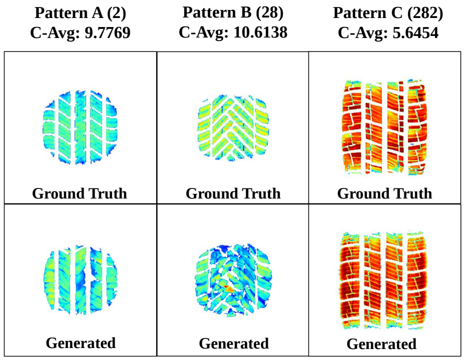


Figure 10: **Comparison of Generated and Ground Truth Images Across Minor and Major Patterns**

5.5 Effect of Pattern Sample Size on Generation Quality

Our analysis revealed variations in TireDiff generation performance across different pattern sample sizes in the training data. Given the median number of training samples per pattern (55), we categorized patterns into minority (< 55 samples) and majority (≥ 55 samples) groups. Across all test samples, while the overall C-Avg was 8.0412, minority patterns showed a higher C-Avg of 9.0914 compared to 7.5010 for majority patterns. Figure 10 demonstrates this variation through patterns A (2 samples), B (28 samples), and C (282 samples). Minority patterns A and B showed substantial deviations in pressure distribution and tread details from ground truth data, while majority pattern C maintained consistent generation quality. These results suggest limitations in generalization for minority patterns due to data imbalance, indicating the need for improved training strategies. We plan to address these limitations in future work.

6 Conclusion

We proposed TireDiff to address key challenges in tire manufacturing: high prototyping costs, environmental waste, and slower innovation cycles. Our framework generates tire footprint images directly from manufacturing specifications, eliminating the need for physical prototypes. Through Dual-Stream Embedding strategy, which effectively handles both textual and numerical data, and physics-informed Tire Custom Loss, TireDiff achieves accurate physical property preservation. The Tire-Reference-Free Predictor enables reliable quality assessment without ground truth data, solving a critical evaluation challenge in practical applications. our framework demonstrates the potential for significant environmental benefits where generative models can reduce material waste while maintaining competitive performance across diverse tire patterns. We believe these advancements will promote sustainability in the tire manufacturing industry and catalyze broader adoption of generative models in manufacturing sectors that demand efficient, prototype-free design validation.

7 Acknowledgments

This work was supported by Hankook Tire & Technology Co., Ltd, Republic of Korea, and some experiments were conducted using the instruments of Hankook Tire & Technology Co., Ltd. The authors extend their sincere thanks to Hankook Tire & Technology Co., Ltd for their help and support.

References

Agarwal, R.; Liang, C.; Schuurmans, D.; and Norouzi, M. 2019. Learning to generalize from sparse and underspecified rewards. In *International conference on machine learning*, 130–140. PMLR.

Agnese, J.; Herrera, J.; Tao, H.; and Zhu, X. 2020. A survey and taxonomy of adversarial neural networks for text-to-image synthesis. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(4): e1345.

Arik, S. Ö.; and Pfister, T. 2021. Tabnet: Attentive interpretable tabular learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, 6679–6687.

Balakina, E.; Zadvornov, V.; Sarbaev, D.; Sergienko, I.; and Kozlov, Y. N. 2019. The calculation method of the length of contact of car tires with the road surface. In *IOP Conference Series: Materials Science and Engineering*, volume 632, 012022. IOP Publishing.

Barbosa, B. H. G.; Xu, N.; Askari, H.; and Khajepour, A. 2021. Lateral force prediction using Gaussian process regression for intelligent tire systems. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 52(8): 5332–5343.

Chen, T.; and Guestrin, C. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785–794.

Chen, W.; Wang, H.; Chen, J.; Zhang, Y.; Wang, H.; Li, S.; Zhou, X.; and Wang, W. Y. 2019. Tabfact: A large-scale dataset for table-based fact verification. *arXiv preprint arXiv:1909.02164*.

Devlin, J. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Dhariwal, P.; and Nichol, A. 2021. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34: 8780–8794.

Dubey, A.; Jauhri, A.; Pandey, A.; Kadian, A.; Al-Dahle, A.; Letman, A.; Mathur, A.; Schelten, A.; Yang, A.; Fan, A.; et al. 2024. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.

Esser, P.; Rombach, R.; and Ommer, B. 2021. Taming transformers for high-resolution image synthesis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 12873–12883.

Ho, J.; Jain, A.; and Abbeel, P. 2020. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33: 6840–6851.

Howland, S.; Kassab, L.; Kappagantula, K.; Kvinge, H.; and Emerson, T. 2023. Parameters, Properties, and Process: Conditional Neural Generation of Realistic SEM Imagery Toward ML-Assisted Advanced Manufacturing. *Integrating Materials and Manufacturing Innovation*, 12(1): 1–10.

Huang, X.; Khetan, A.; Cvitkovic, M.; and Karnin, Z. 2020. Tabtransformer: Tabular data modeling using contextual embeddings. *arXiv preprint arXiv:2012.06678*.

Kanopoulos, N.; Vasanthavada, N.; and Baker, R. L. 1988. Design of an image edge detection filter using the Sobel operator. *IEEE Journal of solid-state circuits*, 23(2): 358–367.

Kingma, D. P. 2013. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.

Koleva, A.; Ringsquandl, M.; Joblin, M.; and Tresp, V. 2021. Generating table vector representations. *arXiv preprint arXiv:2110.15132*.

Kumar, A.; Starly, B.; and Lynch, C. 2023. ManuBERT: A pretrained Manufacturing science language representation model. *Available at SSRN 4375613*.

Kusiak, A. 2024. Generative artificial intelligence in smart manufacturing. *Journal of Intelligent Manufacturing*, 1–3.

Lee, Y.; Park, I.; and Kang, M. 2024. FLEUR: An Explainable Reference-Free Evaluation Metric for Image Captioning Using a Large Multimodal Model. *arXiv preprint arXiv:2406.06004*.

Liu, R.; Ma, B.; Zhang, W.; Hu, Z.; Fan, C.; Lv, T.; Ding, Y.; and Cheng, X. 2024. Towards a simultaneous and granular identity-expression control in personalized face generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2114–2123.

Loshchilov, I.; and Hutter, F. 2017. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*.

Mehri, S.; and Eskenazi, M. 2020. USR: An unsupervised and reference free evaluation metric for dialog generation. *arXiv preprint arXiv:2005.00456*.

Naseem, U.; Razzak, I.; Khan, S. K.; and Prasad, M. 2021. A comprehensive survey on word representation models: From classical to state-of-the-art word representation language models. *Transactions on Asian and Low-Resource Language Information Processing*, 20(5): 1–35.

Nichol, A.; Dhariwal, P.; Ramesh, A.; Shyam, P.; Mishkin, P.; McGrew, B.; Sutskever, I.; and Chen, M. 2021. Glide: Towards photorealistic image generation and editing with text-guided diffusion models. *arXiv preprint arXiv:2112.10741*.

Padula, S. M. 2005. Tire load capacity. *The Pneumatic Tire*, 186.

Pearson, M.; Blanco-Hague, O.; and Pawlowski, R. 2016. TameTire: Introduction to the model. *Tire Science and Technology*, 44(2): 102–119.

Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A. V.; and Gulin, A. 2018. CatBoost: unbiased boosting with categorical features. *Advances in neural information processing systems*, 31.

Radford, A.; Kim, J. W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.;

et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, 8748–8763. PMLR.

Ramesh, A.; Dhariwal, P.; Nichol, A.; Chu, C.; and Chen, M. 2022. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125*, 1(2): 3.

Rasmy, L.; Xiang, Y.; Xie, Z.; Tao, C.; and Zhi, D. 2021. Med-BERT: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. *NPJ digital medicine*, 4(1): 86.

Ribeiro, A. M.; Moutinho, A.; Fioravanti, A. R.; and de Paiva, E. C. 2020. Estimation of tire–road friction for road vehicles: a time delay neural network approach. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, 42(1): 4.

Ridha, R. A.; and Curtiss, W. W. 2018. Developments in tire technology. In *Rubber Products Manufacturing Technology*, 533–564. Routledge.

Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; and Ommer, B. 2022. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 10684–10695.

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, 234–241. Springer.

Saharia, C.; Chan, W.; Saxena, S.; Li, L.; Whang, J.; Denton, E. L.; Ghasemipour, K.; Gontijo Lopes, R.; Karagol Ayan, B.; Salimans, T.; et al. 2022. Photorealistic text-to-image diffusion models with deep language understanding. *Advances in neural information processing systems*, 35: 36479–36494.

Sennrich, R. 2015. Neural machine translation of rare words with subword units. *arXiv preprint arXiv:1508.07909*.

Simonyan, K. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Singh, R.; and Bedathur, S. 2023. Embeddings for Tabular Data: A Survey. *arXiv preprint arXiv:2302.11777*.

Somepalli, G.; Goldblum, M.; Schwarzschild, A.; Bruss, C. B.; and Goldstein, T. 2021. Saint: Improved neural networks for tabular data via row attention and contrastive pre-training. *arXiv preprint arXiv:2106.01342*.

Song, Y.-Y.; and Ying, L. 2015. Decision tree methods: applications for classification and prediction. *Shanghai archives of psychiatry*, 27(2): 130.

Thawani, A.; Pujara, J.; Szekely, P. A.; and Ilievski, F. 2021. Representing numbers in NLP: a survey and a vision. *arXiv preprint arXiv:2103.13136*.

Wang, Z.; Bovik, A. C.; Sheikh, H. R.; and Simoncelli, E. P. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4): 600–612.

Wolf, T. 2019. Huggingface's transformers: State-of-the-art natural language processing. *arXiv preprint arXiv:1910.03771*.

Wu, Y. 2016. Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144*.

Yin, P.; Neubig, G.; Yih, W.-t.; and Riedel, S. 2020. TaBERT: Pretraining for joint understanding of textual and tabular data. *arXiv preprint arXiv:2005.08314*.

Zaurin, J. R.; and Mulinka, P. 2023. pytorch-widedeep: A flexible package for multimodal deep learning. *Journal of Open Source Software*, 8(86): 5027.

Zhan, Z.; Chen, D.; Mei, J.-P.; Zhao, Z.; Chen, J.; Chen, C.; Lyu, S.; and Wang, C. 2024. Conditional Image Synthesis with Diffusion Models: A Survey. *arXiv preprint arXiv:2409.19365*.

Zhang, L.; Zhang, S.; and Balog, K. 2019. Table2vec: Neural word and entity embeddings for table population and retrieval. In *Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval*, 1029–1032.

Zhang, R.; Isola, P.; Efros, A. A.; Shechtman, E.; and Wang, O. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 586–595.