



# IJRASET

International Journal For Research in  
Applied Science and Engineering Technology



---

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

---

**Volume:** 11    **Issue:** XI    **Month of publication:** November 2023

**DOI:** <https://doi.org/10.22214/ijraset.2023.56847>

[www.ijraset.com](http://www.ijraset.com)

Call:  08813907089

E-mail ID: [ijraset@gmail.com](mailto:ijraset@gmail.com)

# A Study on an Approach to Driver Distraction Detection Based on Predicting the Distribution of Categorical Response Variable with Denoising Diffusion Probabilistic Models

Joshua Banda<sup>1</sup>, Yang Zhou<sup>2</sup>

School of Information and Electronic Engineering, Zhejiang University Of Science And Technology

**Abstract:** As per a World Health Organization (WHO) report, inattentive driving is the 8th greatest reason for traffic fatalities. A lot of work has been done to combat this problem including research into advanced driver-assistance systems (ADAS). This work includes the employment of distracted driver detection systems and algorithms to serve as a warning system. Recent work has shown how deep learning techniques can offer higher accuracy in classifying distracted driver. Earlier work has also shown how machine learning and later deep learning can be used for this task. At its core the task can be broken down to learning features in an image to predict the state. ML and DL algorithms were employed in this research to classify diverted driving using images achieving state of the art results in many occurrences. However, as the research has progressed and models have gotten bigger. How these predictions are made has not been explored. The black box problem is now an area of research and quantifying the uncertainty of the model may help shed light on my predictions were made. We suggest using a new approach to distracted driver detection using denoising diffusion probabilistic models for this task. Performance is enhanced by reconfiguring the CARD model for the task. Our motivation is not to achieve state-of-the-art performance in terms of mean accuracy on the State farm distracted driver dataset, which is strongly related to network architecture design. Our goal is to perform classification via a generative model emphasize the model's capability to improve the performance of a base classifier with deterministic outputs in terms of accuracy.

**Keywords:** distracted driver detection, denoising diffusion probabilistic models, uncertainty estimation.

## I. INTRODUCTION

Distracted Driving behaviour defined as any activity that a driver takes part in that takes attention away from the road. According to the Fatality Analysis Reporting System of the United States Department of Transportation, 33,244 fatal accidents occurred in the United States in 2019, with 36,096 people deaths [1]. According to the World Health Organization (WHO) survey, 1.3 million people worldwide die in traffic accidents each year, making them the eighth leading cause of death and an additional 20-50 millions are injured/ disabled.[2]

There are three main types of distraction. Visual: taking your eyes off the road. Manual: taking your hands off the wheel and Cognitive: taking your mind off driving. As a result, a distracted driver has a significantly increased probability of getting involved in a vehicle accident. There are many types of distractions that can lead to impaired driving examples of distractions that take drivers focus and attention away from driving are talking to other passengers, mobile phones, navigation systems, and complex air conditioning systems [3]. The distraction caused by mobile phones is a growing concern for road safety. Drivers using mobile phones are approximately 4 times more likely to be involved in a crash than drivers not using a mobile phone. Using a phone while driving slows reaction times (notably braking reaction time, but also reaction to traffic signals), and makes it difficult to keep in the correct lane, and to keep the correct following distances. Hands-free phones are not much safer than hand-held phone sets, and texting considerably increases the risk of a crash [1].

Most of the existing research on distracted driving behaviour classification has focused on traditional machine learning and deep learning approaches, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and support vector machines (SVMs). These approaches have shown state of the art results in detecting various forms of distracted driving behaviour, but they have limitations in terms of interpretability and generalizability. They are often criticized for their "black-box" nature, which makes it challenging to understand why they make certain predictions.

The "blackbox problem" in the context of deep neural networks refers to the difficulty in understanding the internal workings of the model, including the relationships between input and output, and the specific features that the model considers important. This lack of interpretability can make it difficult to trust the model's predictions, especially in critical applications like healthcare or autonomous driving [4].

One way to tackle the blackbox problem is through uncertainty estimation. Uncertainty estimation in deep learning refers to the process of quantifying the level of confidence that the model has in its predictions. This can be achieved by training the model to output a distribution over predictions, rather than a single point estimate. This distribution can provide insights into the model's confidence in its predictions, which can be used to understand the model's uncertainty. For example, in Bayesian neural networks, the weights of the network are treated as random variables that follow a certain distribution. This allows the model to output a distribution over predictions for each input, which can be used to estimate the uncertainty of the predictions.

Diffusion Probabilistic Predictive Models (DPPMs) offer a unique approach to tackle the "blackbox" problem in deep learning models. Unlike traditional deep learning models that output a single point estimate, DPPMs are probabilistic models that output a distribution over predictions. This allows for a more nuanced understanding of the model's uncertainty and can provide insights into the model's decision. By modelling the way in which data points diffuse through the latent space, diffusion models can capture the underlying structure of the data and provide insights into the factors that contribute to distracted driving behaviour. However, there is limited research on the use of diffusion models for this task, and more research is needed to explore their potential in this domain. There is a gap in the literature regarding the use of the diffusion model in classifying distracted driving images, as this is a relatively new area of research that has not yet been extensively explored. While diffusion models have been used for various computer vision tasks, such as image denoising, inpainting, and generation, there is limited research on their use in classifying distracted driving behaviour.

## II. RELATED WORK

According to [5] technology solutions have also been used to reduce distracted driving incidents. There are applications and devices that can help prevent distracted driving by blocking calls, texts, or internet access while the vehicle is in motion. Some advanced systems can also monitor driving behaviour and send notifications to parents of teen drivers. Naturalistic driving studies that use onboard sensors and cameras to capture data right before crashes as well as during normal driving situations are beginning to shed light on the risks posed by specific distracted driving behaviours [6]. For example, a change in risk greater than 1 represents increases in crash risk due to the secondary task, while a change in risk less than 1 represents a decrease in crash risk. Interacting with a handheld cell phone increases the risk of a crash 3.6 times compared to baseline driving without a phone in hand.

Various approaches have been used to detect and classify distracted driving behaviour, including human observation and machine learning models. According to authors [7], machine learning approaches, including feature generation, can be used to determine the best-performing algorithms for detecting driver distraction and predicting the source of distraction. In this study [7], 21 algorithms were trained to identify when drivers were distracted by secondary cognitive and texting tasks, and the highest-performing algorithm for accurately classifying driver distraction was a Random Forest algorithm, trained using only driving behaviour measures and excluding driver physiological data. The most important input measures identified were lane offset, speed, and steering, whereas the most important feature types were standard deviation, quantiles, and nonlinear transforms. This work suggests that distraction detection algorithms may be improved by considering ensemble machine learning algorithms that are trained with driving behaviour measures and nonstandard features. In a study [8] authors used a combination of head and facial associated features to target distracted driving. Monitoring the state of the eye and position of the head patterns were tracked overtime to classify between alert versus non-alert. This work highlighted the very rich information that can be extracted from the head and eye regions and showed its great potential towards understanding distracted behaviour. The method proposed by [9] authors tried to address the problem by acknowledging and targeting a common issue across many machine learning applications; the lack of labelled data. The research team proposed a semi-supervised method that, similar to works of the past, utilized eye and head movements to detected distractions based on both labelled and unlabelled data.

In more recent works, deep-learning methods have been evaluated on similar experimental setups. The works proposed by [10] and [11], authors utilized convolutional neural networks to classify video segments into 10 target classes using the dataset proposed by [12]. These two papers were likely the first to go beyond distraction detection to distraction recognition. However, their methods were highly dependent on discriminating physical distractors by targeting labels such as "reaching behind" or "talking on phone with the right hand", thus being very limited to other kinds of passive distractors that relate to anxiety, frustration or even verbal interaction.

An approach that is increasingly gaining the attention of related research as modern cars are being equipped with more advanced sensors is physiological based driver modelling [13, 14]. The review provided by [15] offered a detailed overview of the early approaches on distracted driving detection using physiological data. Since then, things have not drastically changed as the community keeps addressing the topic based on signals related to respiration, heart rate, muscle activity and visual clues. However, research has slowly shifted from understanding statistical correlations to building driver-centric behaviour models based on the aforementioned signals.

The work in [16] used deep learning to detect driver inattentive and aggressive behaviour. They classified inattentive driver behaviour into driver fatigue, downiness, driver distraction, and other risky driver behaviour such as driving aggressiveness. All these risky driving behaviours are associated with various factors that include driving age, experience, illness, and gender. The authors used CNNs, RNNs and LSTMs. They showed that the CNNs achieved the best performance.

The algorithm in [17] detects driver manual distraction using two modules; in the first module, the bounding boxes of the driver's right ear and right hand are detected from RGB images through YOLO, a deep learning object detection model. Then, the bounding boxes are taken as an input by the second module, a multi-layer perceptron, to predict the distraction type. The dataset consisted of 106,677 frames extracted from a video obtained from 20 participants in a driving simulator. The proposed algorithm achieved comparable results with other models in the same field. There is a research gap in the literature in using Bayesian neural networks and denoising diffusion-based conditional generative models for the task of distracted driver detection but we explore them in the context of image classification to establish an opportunity for novel approaches to the task of distracted driving classification.

Diffusion-based generative models have received significant attention recently due to not only their ability to generate high-dimensional data, such as high-resolution photo-realistic images, but also their training stability. The scholarly literature on diffusion-based generative models has shown that these models have achieved a rapid paradigm shift in deep generative models by showing ground-breaking performance across various applications. These models have been applied to various domains, including visual computing, structured data, and natural language generation [18].

The paper [19] the authors provide a comprehensive overview of the current state of diffusion models in the field of visual computing. The authors report on diffusion models for visual computing introduces the basic mathematical concepts of diffusion models, implementation details and design choices of the popular Stable Diffusion model, and overviews important aspects of these generative AI tools, including personalization, conditioning, inversion, among others. It provides a comprehensive overview of the rapidly growing literature on diffusion-based generation and editing, categorized by the type of generated medium, including 2D images, videos, 3D objects, locomotion, and 4D scenes. In addition to the basic understanding of diffusion models, the literature also explores how to learn a conditional distribution using diffusion models via guidance. It also discusses the potential for combining diffusion models with other generative models for enhanced results.

Diffusion-based models can be understood from the perspective of score matching. In the paper [20] presents a novel approach to generative modelling using denoising autoencoders. The authors, show that a simple denoising autoencoder training criterion is equivalent to matching the score (with respect to the data) of a specific energy-based model to that of a nonparametric Parzen density estimator of the data. The authors argue that this equivalence defines a proper probabilistic model for the denoising autoencoder technique, which makes it in principle possible to sample from them or rank examples by their energy. This approach suggests a different way to apply score matching that is related to learning to denoise and does not require computing second derivatives. The authors also justify the use of tied weights between the encoder and decoder and suggest ways to extend the success of denoising autoencoders to a larger family of energy-based models.

In the paper [21], the authors presents a novel approach to predicting the distribution of a continuous or categorical response variable given its covariates. The authors, introduce Classification and Regression Diffusion (CARD) models, which combine a denoising diffusion-based conditional generative model and a pre-trained conditional mean estimator. The authors argue that deep neural network-based supervised learning algorithms have made significant progress in predicting the mean of the response variable given the covariates, but they often struggle to accurately capture the uncertainty of their predictions. CARD models aim to address this issue by predicting the distribution of the response variable, rather than just its mean.

The authors demonstrate the effectiveness of CARD models in predicting conditional distributions with both toy examples and real-world datasets. They show that CARD models generally outperform state-of-the-art methods, including Bayesian neural network-based ones designed for uncertainty estimation. This is especially true when the conditional distribution of the response variable given the covariates is multi-modal. The authors also utilize the stochastic nature of the generative model outputs to obtain a finer granularity in model confidence assessment at the instance level for classification tasks. This allows for a more nuanced understanding of the model's predictions and their associated uncertainties.

### III.DATE SET DESCRIPTION

For the purpose of this task, we made use of the State Farm Distracted Driver Detection (SFD3) dataset [1], which is the public database. This dataset offered a wide variety of classes showing different scenarios of a distracted driver behaviour. The dataset contains 10 classes of distracted driver behaviour consisting of safe driving, texting on mobile phone using left, talking on the phone left, texting on mobile phone right, talking on phone right, operating the radio, drinking, reaching behind, doing hair and makeup, and talking to passenger respectively. These classes are shown in the figure (insert distracted driver classes)

TABLE I

The classes in the dataset, the types of images it contains, and the number of images by classes

Class	Distraction Type	Number Of Images
C0	Safe Driving	2489
C1	Texting-Right	2267
C2	Talking on Phone Right	2317
C3	Texting Left	2346
C4	Talking on Phone Left	2326
C5	Operating Radio	2312
C6	Drinking	2325
C7	Reaching Behind	2002
C8	Hair, Make Up	1911
C9	Talking To Passenger	2129

State Farm’s Distracted Driver Detection competition on Kaggle was the first publicly available dataset for posture classification. In the competition. State Farm defined ten postures to be detected: safe driving, texting using right hand, talking on the phone using right hand, texting using left hand, talking on the phone using left hand, operating the radio, drinking, reaching behind, doing hair and makeup, and talking to passenger. Our work, in this paper, is mainly inspired by State Farm’s Distracted Driver’s competition. In 2017, Abouelnaga [22] created a new dataset similar to State Farm’s dataset for distracted driver detection. Authors pre-processed the images by applying skin, face and hand segmentation and proposed the solution using weighted ensemble of five different Convolutional Neural Networks. The system achieved good classification accuracy but is computationally too complex to be real time which is utmost important in autonomous driving.

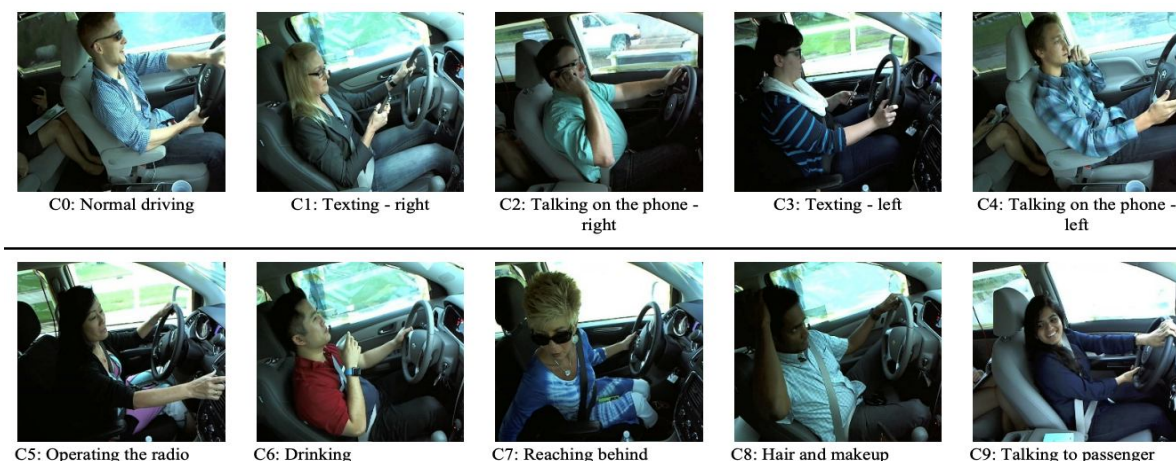


Fig. 1 Classes of driver behaviour

#### IV. PROPOSED METHODOLOGY

Our investigation primarily relies on the State Farm Distracted Driver dataset, which is well-documented and publicly accessible. This dataset comprises images captured from 2D dashboard cameras, aiming to detect various distracted driving behaviours. It encompasses ten classes, including instances of both distracted driving behaviours and alert driving. To ensure comprehensive evaluation, the dataset is divided into training and testing images, forming the foundation for model training.

The methodology delves into the intricate details of the CARD architecture, a distinct approach that combines denoising diffusion probabilistic models with a pre-trained conditional mean estimator. This architecture is a cornerstone of our work, designed to provide not only point estimates but also uncertainty measures in its predictions. By employing forward and reverse diffusion chains, CARD offers a means to model the distribution of response variables given their covariates.

Moreover, our study explores the Bayesian Neural Networks (BNNs) to harness their unique capabilities for capturing and quantifying uncertainty in neural network predictions. We delve into the architectural details of the Linear Neural Network, Convolutional Neural Network, Resnet20, and VGG, which are all equipped with Bayesian elements to provide probabilistic assessments of their predictions. By comparing the CARD model with BNNs and employing an array of neural network architectures, our methodology is designed to provide a comprehensive analysis of the predictive distribution of response variables, particularly within the context of distracted driver detection. The subsequent sections will delve deeper into the specific methodologies employed for training, experiments, and the comparative analysis of these techniques.

##### A. CARD model

The overall architecture of the CARD model is based on the idea of denoising diffusion probabilistic models (DDPM), which use a diffusion process to generate samples from a target distribution. The CARD model consists of two diffusion chains: a forward chain and a reverse chain. The forward chain generates samples from the prior distribution of the response variable, while the reverse chain generates samples from the posterior distribution of the response variable given its covariates.

The forward chain is a diffusion process that generates samples from the prior distribution of the response variable. The diffusion process is modelled using a series of  $T$  steps, where each step involves adding Gaussian noise to the current sample and then applying a neural network to denoise the sample. The neural network is trained to remove the added noise and recover the original sample. The reverse chain is a diffusion process that generates samples from the posterior distribution of the response variable given its covariates. The diffusion process is modelled using a series of  $T$  steps, where each step involves adding Gaussian noise to the current sample and then applying a neural network to denoise the sample. The neural network is trained to remove the added noise and recover the original sample, conditioned on the covariates.

The pre-trained conditional mean estimator is a neural network that estimates the mean of the conditional distribution of the response variable given its covariates. The conditional mean estimator is trained separately from the diffusion chains and is used to provide an initial estimate of the mean of the conditional distribution. During training, the CARD model is optimized to minimize the difference between the estimated conditional distribution and the true conditional distribution of the response variable given its covariates. The model is trained using a maximum likelihood objective, which maximizes the likelihood of the observed data given the model parameters. In CARD model training, the size of the input image is adjusted to  $128 \times 128$ . We normalize the dataset with the mean and standard deviation of the training set both set to  $(0.5, 0.5, 0.5)$ . With number of workers set to 2, and a batch size set to 48. The training cycle is 600 epochs and use Adam Optimizer and adopt a cosine learning rate decay [23] for all tasks. we set the number of timesteps to 1000, and adopt a linear  $\beta_t$  schedule same as [24]. We use exponentially weighted moving averages [25] on model parameters with a decay factor of 0.9999. We adopt antithetic sampling [26] to draw correlated timesteps during training.

For the diffusion model, we follow a similar network architecture of [27] which also followed a similar architecture of [28] and [29] by first changing the Transformer sinusoidal position embedding to linear embedding for the timestep. As the network has three other inputs besides the timesteps, we integrate them by first applying an encoder on the flattened input image ( $128 \times 128 \times 3$ ) to obtain a representation with 4096 dimensions. The encoder consists of three fully-connected layers with an output dimension of 4096. Meanwhile we concatenate  $y_t$  and  $f\phi(x)$ , and apply a fully-connected layer to obtain an output vector of 4096 dimensions. We perform a Hadamard product between such vector and a timestep embedding to obtain a response embedding conditioned on the timestep. We then perform Hadamard product between image embedding and response embedding to integrate these variables, and send the resulting vector through two more fully-connected layers with 4096 output dimensions, each would first follow by a Hadamard product with a timestep embedding, and lastly a fully-connected layer with an output dimension of 1 as the noise prediction. Note that all fully-connected layers are also followed by a batch normalization layer and a Softplus non-linearity, except the output layer.

For classification tasks, the CARD model works by first training a denoising diffusion-based generative model to learn the joint distribution of the response variable and the covariates. The generative model is then used to generate samples of the response variable given new covariates. Next, the pre-trained mean estimator is used to estimate the mean of the response variable given the generated samples. The resulting distribution is then used to make predictions about the response variable. One of the key features of the CARD model for classification tasks is its ability to provide calibrated confidence in its predictions. The stochastic nature of the generative model outputs is used to obtain a finer granularity in model confidence assessment at the instance level. This allows the model to identify cases where it is less assertive and pass them to humans for further evaluation.

**B. Comparison with Bayesian Neural Network Experiment Details**

Bayesian Neural Network (BNN) is a neural network that is augmented with prior distributions on its weights and biases. During training, the goal is to learn the posterior distribution over the weights and biases given the observed data [29]. This is done using variational inference, which involves approximating the posterior distribution with a simpler distribution that is easier to work with. The simpler distribution is typically chosen to be a factorized Gaussian distribution, and the parameters of this distribution are learned during training. Once the posterior distribution is learned, it can be used to make predictions on new data points, and the uncertainty in the predictions can be quantified using the variance of the posterior distribution.

For classification tasks, a common architecture for a BNN is a feedforward neural network with one or more hidden layers. The input to the network is a vector of features, and the output is a probability distribution over the classes. The hidden layers of the network consist of a set of neurons, each of which computes a weighted sum of its inputs and applies a non-linear activation function to the result. The weights and biases of the neurons are drawn from the prior distribution, and are updated during training to learn the posterior distribution.

During training, the network is trained to minimize the negative log-likelihood of the observed data given the posterior distribution over the weights and biases. This is done using stochastic gradient descent, where the gradients of the loss function with respect to the weights and biases are estimated using a mini-batch of data points. The gradients are then used to update the parameters of the posterior distribution, which in turn are used to update the weights and biases of the network.

Once the network is trained, it can be used to make predictions on new data points by computing the posterior distribution over the weights and biases given the observed data, and using this distribution to compute the probability distribution over the classes. The uncertainty in the predictions can be quantified using the variance of the posterior distribution.

TABLE 2  
BNN ARCHITECTURE SUMMARY AND KEY FEATURES

Architecture	Layers	Key Features
LinearBnn	Fully connected (linear)	Uncertainty modeling, probabilistic predictions, ReLU activation, KL divergence
ConvBnn	Convolutional	Uncertainty modeling, probabilistic predictions, ReLU activation, KL divergence, convolutional operations
Resnet20	Residual blocks	Uncertainty modeling, probabilistic predictions, skip connections, max-pooling, ReLU activation, KL divergence
VGG	Convolutional blocks	Uncertainty modeling, probabilistic predictions, max-pooling, ReLU activation, KL divergence

**V. RESULTS AND DISCUSSION**

**A. CARD Experiment Results**

The results of classification using the CARD model are presented below:

In the conducted analysis of card classification, the mean accuracy of all test instances at the test instance level was found to be 90.5611%. To further evaluate the model, metrics related to predicted probability quantiles for all classes were computed. The model's accuracy was assessed based on class-specific narrowest credible interval (CI) widths. Specifically, the label prediction by the class with the narrowest CI width (1-th narrowest) resulted in a test accuracy of 90.6510%. However, as the CI width increased for the classes with lower narrowness, the accuracy decreased considerably

TABLE 3  
COMPARISON OF MODEL ACCURACY WITH ASSOCIATED HYPER PARAMETERS

Metric	Value	Metric
Mean Accuracy	90.5611%	Mean Accuracy
Majority-Voted Accuracy	90.5470%	Majority-Voted Accuracy
Total Test Instances	6728	Total Test Instances
Correct Predictions	6092	Correct Predictions
Incorrect Predictions	636	Incorrect Predictions
Mean CI Width (Correct)	0.0034	Mean CI Width (Correct)
Mean CI Width (Incorrect)	0.0666	Mean CI Width (Incorrect)

A majority-voted class label approach was also applied, yielding an overall test accuracy of 90.5470%. Out of 6728 test instances, 6092 were classified correctly, while 636 instances were classified incorrectly. The mean credible interval width of the predicted probability for the true class was 0.0034 for correct predictions and 0.0666 for incorrect predictions. The performance of the model was further analyzed on a per-class basis. Class-specific accuracy and mean CI widths were calculated for each of the ten classes (Class 0 to Class 9). Accuracy varied across classes, with Class 7 having the highest accuracy at 99.6569% and Class 0 having the lowest at 71.6016%. Mean CI widths also varied, indicating differences in prediction confidence among classes.

TABLE 4  
CARD MODEL CLASS-SPECIFIC RESULTS

Class	Accuracy	Total Images	Correct Predictions	Incorrect Predictions
C0	71.60%	743	532	211
C1	98.56%	694	684	10
C2	97.46%	710	692	18
C3	97.42%	697	679	18
C4	96.36%	714	688	26
C5	87.79%	688	604	84
C6	94.48%	724	684	40
C7	99.66%	583	581	2
C8	88.35%	558	493	65

Paired two-sample t-tests were conducted to evaluate the statistical significance of the model's predictions. The majority-voted class label approach was used, resulting in an overall test accuracy of 90.5470%. Among all test instances, 6723 t-tests were rejected with a mean accuracy of 90.5994%, while 5 t-tests were not rejected with a mean accuracy of 20.0000%.

Performance on a per-class basis in terms of t-test rejection rates and associated mean accuracies was also assessed. In general, most t-tests were rejected with high accuracy, though there were a few exceptions with very low rejection rates and accuracies. The Probability of Accurate and Certain Predictions Under Uncertainty (PAvPU) was calculated to assess model reliability. It was found to be 0.90591558 with an alpha level of 0.0500 for the test set with a size of 6728.

Additionally, the analysis included the computation of Negative Log-Likelihood (NLL) and Expected Calibration Error (ECE) for the test set. The NLL was found to be 0.40194750, and the ECE was 0.02213415. The testing procedure, including all analyses and computations, was completed in a total time of 1143.0076 minutes.



**B. Comparison with Bayesian Neural Network Experiment Results**

The outcomes of this study offer valuable insights into the efficacy of diverse deep learning models for image classification, with a particular focus on the State farm dataset. The examination encompassed models with varying architectures, including Resnet, Linear, Convolutional (Conv), VGG16, and VGG19. Resnet Model: After 300 epochs of training, the Resnet model achieved a test accuracy of 97.71% and a corresponding test Expected Calibration Error (ECE) of 0.0218. In addition to accuracy, the Negative Log-Likelihood (NLL) score, a measure of model uncertainty, provided further insights. The test NLL was 0.0861. During training, the model achieved an accuracy of 100.0% and an ECE of 0.00431. The validation set echoed the high performance, with an accuracy of 97.74 and a validation ECE of 0.02186, along with a validation NLL of 0.0861.

Linear Model: After 300 epochs of training for the Linear model resulted in a test accuracy of 99.06% and a test ECE of 0.00976. The test NLL of 0.0533. The training and validation results were consistent with the test set outcomes. Convolutional Model: After 300 epochs of training, demonstrated a test accuracy of 99.18% and a test ECE of 0.0087. The test NLL was 0.0376. Both the training and validation accuracies aligned with the test results. VGG16 and VGG19 Models: The VGG16 model exhibited a test accuracy of 11.01% and a test ECE of 0.02209. The test NLL score was 2.2967. In contrast, the VGG19 model achieved a test accuracy of 10.89% and a test ECE of 0.02077, with a test NLL of 2.2978. The training and validation result closely mirrored the test set outcomes.

TABLE 5  
CARD MODEL CLASS-SPECIFIC RESULTS

Model	Accuracy	NLL	ECE
Resnet	97.71%	0.08608957	0.02184259
CARD	90.56%	0.4019475	0.02213415
Linear	99.06%	0.05334118	0.00975955
Conv	99.18%	0.03762113	0.00870287
VGG16	11.01%	2.296706	0.022087
VGG19	10.89%	2.29782867	0.02077414

**C. Discussion on CARD Model Results**

The results from the Classification and Regression Diffusion (CARD) model's performance provide crucial insights into its effectiveness for image classification. With a mean accuracy of 90.5611%, the CARD model demonstrates a strong capability to classify distracted driver behaviours effectively. Its predictive accuracy is promising, considering the complexity and variety of real-world distracted driving scenarios captured in the dataset. Notably, the majority-voted class label approach resulted in an overall test accuracy of 90.5470%. This approach indicates that the CARD model is capable of making informed decisions at the instance level. Out of 6728 test instances, 6092 were classified correctly, emphasizing the model's reliability. The assessment of class-specific narrowest credible interval (CI) widths provided an intriguing dimension to the evaluation. It revealed that the model's accuracy is closely related to the narrowness of the CI for each class. This analysis underscores the model's proficiency in making more accurate predictions for classes with narrow CIs, where the uncertainty is relatively low. Class-specific performance varied, with Class 7 exhibiting the highest accuracy at 99.6569%, indicating the model's aptitude to effectively distinguish instances of texting-right. Class 0, with an accuracy of 71.6016%, demonstrated the most challenging behaviour to classify. The evaluation of statistical significance through paired two-sample t-tests reinforced the model's credibility. The majority of t-tests were rejected with high accuracy, signifying the model's effectiveness in making predictions that align with ground truth. However, there were a few exceptions with very low rejection rates and accuracies, highlighting potential areas for further model improvement. The Probability of Accurate and Certain Predictions Under Uncertainty (PAvPU) was computed to assess the model's reliability. It yielded a PAVPU value of 0.90591558, indicating that the model provides reliable predictions with an alpha level of 0.0500. This reinforces the CARD model's effectiveness in providing both accurate and certain predictions. Additionally, the computation of Negative Log-Likelihood (NLL) and Expected Calibration Error (ECE) showcased the model's calibration and uncertainty estimation capabilities. The NLL value of 0.40194750 and ECE value of 0.02213415 affirm the model's potential for providing well-calibrated probabilistic predictions. The CARD model's performance is a significant contribution to the field of distracted driver behaviour detection. Its robustness, reliability, and ability to handle uncertainty make it a valuable asset in real-world applications.

#### D. Discussion on Comparison with Bayesian Neural Networks (BNNs)

The comparative analysis of the CARD model with various Bayesian Neural Network (BNN) architectures further enriches the discussion. This comparison sheds light on the strengths and weaknesses of the CARD model concerning established deep learning models. Resnet Model: Resnet stands out with a test accuracy of 97.71%, a test Expected Calibration Error (ECE) of 0.0218, and a test Negative Log-Likelihood (NLL) of 0.0861. These results indicate the robustness and calibration of the Resnet model, making it a strong contender.

Linear Model: The Linear model achieved an impressive test accuracy of 99.06% and exhibited a low test ECE of 0.00976 and test NLL of 0.0533. These findings highlight the efficacy of a linear model in capturing complex patterns. Convolutional Model: The Convolutional model delivered a test accuracy of 99.18% and showcased a test ECE of 0.0087 and a test NLL of 0.0376. Its performance aligns closely with the Linear model, demonstrating the strength of convolutional architectures. VGG16 and VGG19 Models: VGG models performed considerably less effectively, with test accuracies below 11%. These models exhibited significantly higher ECE values and NLL scores. Their results imply challenges in handling complex image classification tasks. The comparative analysis emphasizes that the CARD model, with a mean accuracy of 90.56%, competes effectively with other models. While it may not outperform models like Resnet and the Linear model in terms of accuracy, it excels in providing calibrated probabilistic predictions and reliable results. Overall, the comparison underscores the significance of the CARD model in the realm of distracted driver behaviour detection. Its balance between accuracy and reliability positions it as a valuable tool for real-world applications, especially when considering the importance of understanding model uncertainty and prediction reliability.

### VI. CONCLUSION AND FUTURE WORK

The results presented in this chapter demonstrate the effectiveness of the CARD model for image classification, especially in the context of distracted driver behaviour detection. Additionally, the comparative analysis with BNNs highlights the model's balance between accuracy and reliability. The CARD model's reliability, calibration capabilities, and provision of probabilistic predictions make it a valuable asset in safety-critical applications. These findings have implications for both research and practical applications, emphasizing the importance of understanding model uncertainty and prediction reliability in deep learning models. This chapter concludes the discussion of the results and their broader implications, setting the stage for future research and applications in the field of image classification and beyond.

The findings of this study hold several implications for both research and practical applications. The CARD model's ability to provide calibrated probabilistic predictions is invaluable in safety-critical applications such as distracted driver detection. It offers not only accurate predictions but also insights into the model's confidence in its predictions, enabling more informed decisions.

Future research directions may involve further enhancing the CARD model's performance, particularly for challenging classes like Class 5, which demonstrated lower accuracy. Investigating model interpretability and expanding the dataset to encompass more diverse scenarios could lead to improved results. In practical applications, the CARD model can be instrumental in advanced driver assistance systems (ADAS) and autonomous vehicles. Understanding driver behavior with reliable confidence levels can enhance road safety measures. This study contributes to the growing body of knowledge on deep learning for image classification, emphasizing the significance of robust, calibrated, and reliable models. It reinforces the importance of model uncertainty assessment in real-world applications.

### VII. ACKNOWLEDGMENT

I would like to thank and acknowledge first and foremost Theresa Mwila my dear mother whose tireless effort and encouragement have been crucial for my studies and advancements. I would also like to acknowledge Daniel Banda who was also instrumental in me being able to continue on my journey with the vital help he offered. I would also like to thank and appreciate the faculty of Zhejiang University of Science and technology. For fulfilling their duties diligently and providing me the resources needed for my higher education.

### REFERENCES

- [1] S. National Highway Traffic Safety Administration. (2010). Fatality analysis reporting system (FARS) encyclopedia. FARS Data Table . 2016.
- [2] "Road traffic Injuries", WHO fact sheet, June 2021.
- [3] Wahlstrom, E., O. Masoud, and N. Papanikolopoulos. Visionbased methods for driver monitoring. in Proceedings of the 2003 IEEE International Conference on Intelligent Transportation Systems. 2003. IEEE.
- [4] Gawlikowski J, Tassi C R N, Ali M. A Survey of Uncertainty in Deep Neural Networks, 2022.
- [5] AAA Exchange. Tips for Preventing Distracted Driving . 2022.

- [6] National Safety Council. Distracted Driving Technology Solutions. /2023-10-15. <https://www.nsc.org/road/safety-topics/distracted-driving/technology-solutions>.
- [7] Ahmed M M, Khan M N, Das A. Global lessons learned from naturalistic driving studies to advance traffic safety and operation research: A systematic review. *Accident Analysis & Prevention*, 2022.
- [8] Wang Rongben, Guo Lie, Tong Bingliang. Monitoring mouth movement for driver fatigue or distraction with one camera. *Proceedings. The 7th International IEEE Conference on Intelligent Transportation Systems (IEEE Cat. No.04TH8749)[C]*. Washington, WA, USA: IEEE, 2004: 314–319.
- [9] Oyini Mbouna R, Kong S G, Chun M-G. Visual Analysis of Eye State and Head Pose for Driver Alertness Monitoring. *IEEE Transactions on Intelligent Transportation Systems*, 2013, 14(3): 1462–1469.
- [10] Liu T, Yang Y, Huang G-B. Driver Distraction Detection Using Semi-Supervised Machine Learning. *IEEE Transactions on Intelligent Transportation Systems*, 2016, 17(4): 1108–1120.
- [11] Kose N, Kopuklu O, Unnervik A. Real-Time Driver State Monitoring Using a CNN Based Spatio-Temporal Approach. *arXiv*, 2019.
- [12] Rao X, Lin F, Chen Z. Distracted driving recognition method based on deep convolutional neural network. *Journal of Ambient Intelligence and Humanized Computing*, 2021, 12(1): 193–200.
- [13] Abouelnaga Y, Eraqi H M, Moustafa M N. Real-time Distracted Driver Posture Classification. *arXiv*, 2018.
- [14] Muhlbacher-Karrer S, Mosa A H, Faller L-M. A Driver State Detection System—Combining a Capacitive Hand Detection Sensor With Physiological Sensors. *IEEE Transactions on Instrumentation and Measurement*, 2017, 66(4): 624–636.
- [15] Heung-Sub Shin, Sang-Joong Jung, Jong-Jin Kim. Real time car driver's condition monitoring system. *2010 IEEE Sensors*. Kona, HI: IEEE, 2010: 951–954.
- [16] Alkinani M H, Khan W Z, Arshad Q. Detecting Human Driver Inattentive and Aggressive Driving Behavior Using Deep Learning: Recent Advances, Requirements and Open Challenges. *IEEE Access*, 2020, 8: 105008–105030.
- [17] Li L, Zhong B, Hutmacher C. Detection of driver manual distraction via image-based hand and ear recognition. *Accident Analysis & Prevention*, 2020, 137: 105432.
- [18] Barber D, Bishop C M. Ensemble learning in Bayesian neural networks. 1998.
- [19] Koo H, Kim T E. A Comprehensive Survey on Generative Diffusion Models for Structured Data[J]. *arXiv*, 2023.
- [20] Po R, Yifan W, Golyanik V. State of the Art on Diffusion Models for Visual Computing[J]. *arXiv*, 2023
- [21] Han X, Zheng H, Zhou M. CARD: Classification and Regression Diffusion Models[J]. *arXiv*, 2022.
- [22] N. Das, E. Ohn-Bar, and M. M. Trivedi. On performance evaluation of driver hand detection algorithms: Challenges, dataset, and metrics. In *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, pages 2953–2958, Sept 2015.
- [23] State Farm Distracted Driver Detection [EB/OL]. /2023-10-16. <https://kaggle.com/competitions/state-farm-distracted-driver-detection>.
- [24] Sohl-Dickstein J, Weiss E A, Maheswaranathan N. Deep Unsupervised Learning using Nonequilibrium Thermodynamics. *arXiv*, 2015.
- [25] Loshchilov I, Hutter F. SGDR: Stochastic Gradient Descent with Warm Restarts. *arXiv*, 2017.
- [26] Cox D R. Prediction by Exponentially Weighted Moving Averages and Related Methods. *Journal of the Royal Statistical Society: Series B (Methodological)*, 1961, 23(2): 414–422.
- [27] Ramesh A, Dhariwal P, Nichol A. Hierarchical Text-Conditional Image Generation with CLIP Latents. *arXiv*, 2022.
- [28] Ren H, Zhao S, Ermon S. Adaptive Antithetic Sampling for Variance Reduction[A]. *Proceedings of the 36th International Conference on Machine Learning[C]*. PMLR, 2019: 5420–5428.
- [29] Xiao Z, Kreis K, Vahdat A. Tackling the Generative Learning Trilemma with Denoising Diffusion GANs. *arXiv*, 2022.
- [30] Zheng H, He P, Chen W. Truncated Diffusion Probabilistic Models and Diffusion-based Adversarial Auto-Encoders. *arXiv*, 2023.
- [31] Ho J, Jain A, Abbeel P. Denoising Diffusion Probabilistic Models. *arXiv*, 2020.



10.22214/IJRASET



45.98



IMPACT FACTOR:  
7.129



IMPACT FACTOR:  
7.429



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24\*7 Support on Whatsapp)