JOURNAL OF DATA SCIENCE | Vol.2026, Issue 2, No.1
eISSN:2805-5160

Uncertainty-Aware Machine Learning for Reliable Decision-Making in Data-Driven Systems

Evi Yulianingsih[1*], Elmar B. Noche[2], Veerendra Yadav[3], Deepika Sharma[4]

[1]Faculty of Science and Technology, Universitas Bina Darma, Palembang, Indonesia
[2]Department of Information Technology, Pangasinan State University – Lingayen Campus, Lingayen, Pangasinan, Philippines
[3]Department of Computer Science & Engineering, Noida International University, Uttar Pradesh 203201, India
[4]Department of Computer Sciences, Noida International University, Uttar Pradesh 203201, India

[*]**Email:** ev_yulianingsih@binadarma.ac.id[1]

Abstract

Machine learning models are increasingly deployed in decision-critical environments such as healthcare, finance, and autonomous systems. However, most conventional models generate deterministic predictions without quantifying uncertainty, which can lead to overconfident mispredictions when data are noisy, incomplete, or outside the training distribution. This limitation exposes a critical gap between predictive accuracy and decision reliability in real-world Al systems. To address this challenge, this study proposes an uncertainty-aware machine learning framework that integrates probabilistic modeling techniques into conventional predictive architectures to jointly estimate epistemic and aleatoric uncertainty. The proposed framework enables models to produce predictive distributions rather than single point predictions, allowing systems to quantify confidence and identify high-risk predictions. Experiments were conducted on multiple benchmark datasets representing both classification and regression tasks under varying levels of noise and data incompleteness. The experimental results demonstrate that the proposed framework achieves predictive performance comparable to deterministic baselines while significantly improving reliability and uncertainty calibration. In classification tasks, the model maintained competitive accuracy and F1-scores while providing well-calibrated confidence estimates, whereas in regression experiments the approach reduced prediction risk by identifying high-error cases through increased uncertainty variance. Robustness tests further show that the framework effectively signals degraded prediction reliability when encountering noisy or incomplete inputs. These findings indicate that incorporating uncertainty estimation enhances trustworthiness and robustness without sacrificing predictive performance. The study highlights uncertainty modeling as a critical component for developing reliable and responsible Al systems capable of supporting risk-sensitive decision-making in real-world data-driven environments.

Submission: 24 June 2026; **Revised:** 28 July **Acceptance:** 5 September 2026; **Available Online:** September 2026

Keywords

Uncertainty Estimation, Probabilistic Machine Learning, Reliable Al, Decision Support Systems, Predictive Uncertainty

Introduction

Machine learning (ML) is an essential technology for modern data-driven applications across a variety of sectors, including healthcare, finance and banking, cybersecurity, business automation, and industrial operations (Sarker, 2022). The combination of data and the technological infrastructure that supports its collection, storage, and processing has enabled ML to be used in environments in which its predictions have real-world, time-critical, and high-impact consequences. The majority of ML algorithms, however, were designed to provide a single-part prediction that is deterministic in nature and do not provide the end user with any understanding of the qualitative uncertainty that prediction is associated with. The lack of a prediction uncertainty framework is a significant challenge in real world applications that involve high-stakes prediction, are safety-critical or safety-influential, or when data distributions change and input data is incomplete, noisy, and missing. ML, in the absence of explicit prediction uncertainty, display overconfidence and produce an answer when the predictive model lacks an answer, with the end result being erroneous or even detrimental decisions. This is the central impetus to investigate ML algorithms that can, to some degree, model and reason about their uncertainty. Figure 1 provides an overview of traditional deterministic models and their framework for uncertainty that is associated with each prediction.

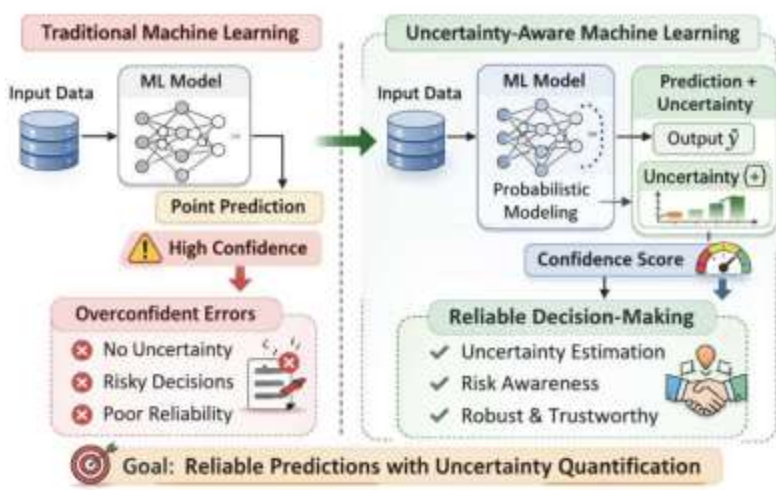

Figure 1. Motivation for Uncertainty-Aware Machine Learning

Uncertainty-aware ML models have predictive and non-predictive outputs, while traditional ML models have deterministic predictions with associated uncertainty in their predictions, allowing for more confident and reasoned decision-making.

In most real-world applications, the source of uncertainty is multiple, and is especially true in the case of data uncertainty (Olschewski et al., 2024), commonly referred to as aleatoric uncertainty, which arises from the variability, noise, intentional, or any other erroneous omissions in the observations of the measurement, or any other breakdown in the data generation process.

Limited training data, inadequate capacity of the model, and incomplete knowledge of the data distribution are reasons for epistemic uncertainty, also called model uncertainty. Typically, deterministic learning structures don't distinguish between various sources of uncertainty. This can lead to the model making very confident predictions even if there is insufficient evidence, which is especially concerning when the model is applied to medical diagnosis, fraud detection, risk assessment, and other autonomous systems. Decisions in any of these domains can lead to substantial financial, operational, or societal losses. As a result, machine learning research has focused on developing models that can quantify and explain the uncertainty of their predictions to ensure that AI systems are trustworthy and reliable.

While the machine learning field has advanced considerably, models that are currently deployed suffer from the same issue and still prioritize predictive accuracy rather than predictive reliability, as demonstrated by Nicora et al. (2022). Accuracy should not be disregarded in terms of performance; however, metrics capture accuracy in various predictive scenarios, including those that are uncertain or unfamiliar. Models trained on historical data may, in practice, face an out-of-distribution input, an infrequent edge case, unexpected environmental changes, or data that was unrepresented during training. Deterministic models will often still make confident predictions despite the model having no previous exposure or experience with the input. This reliance on model performance, without consideration of how that model will actually make use of the input during a task, highlights a disconnect in the current machine learning focus. To close this gap, uncertainty estimation should be an integral part of machine learning processes. If confidence in a model's predictions is integrated with uncertainty, then the system will be able to determine if predictioms should be trusted, or if a system must verify or defer.

There have been several recent attempts to incorporate uncertainty estimation into machine learning models (Tyralis & Papacharalampous, 2024). One of the first ways to model uncertainty is through the use of Bayesian neural networks, which treat network weights as distributions, as opposed to parameters. This means models can capture uncertainty by integrating over potential parameters. Despite this, Bayesian neural networks still require a computation-heavy inference procedure, making them limited in scalability for big data applications. On the other hand, some models like Monte Carlo Dropout, which is a neural network, strives to simulate Bayesian approximations by incorporating a stochastic forward pass, which is done during a neural network run to estimate how uncertain the predictions of the model versus adding additional computational cost for the model to use. Model ensemble learning is also used to capture uncertainty by training different models with separate initializations or subsets of the data and merging their predictions. It is also the case that some of the techniques used for uncertainty estimation in machine learning models include predictive techniques, such as Gaussian Processes, regression, and evidential learning, which allow for the modelling of predictive distributions in order to estimate single models, rather than models with multiple points. The different techniques used to model uncertainty in machine learning models in order to compare the techniques of uncertainty estimation, the Table 1 captures the representation of these techniques with their respective positives and negatives.

Table 1. Existing Uncertainty Estimation Approaches in Machine Learning

Method	Uncertainty Type	Strength	Limitation
Bayesian Networks Neural	Epistemic	Strong probabilistic foundation	Computationally expensive
Monte Carlo Dropout	Epistemic	Easy to implement	Approximation only
Ensemble Learning	Epistemic	Improves robustness	High computational cost
Gaussian Processes	Aleatoric & Epistemic	Accurate uncertainty estimation	Poor scalability

While these techniques still have several challenges in implementing uncertainty estimation in real-world machine learning systems (He et al., 2025). Most of the current techniques work on either epistemic or aleatoric uncertainty, providing no comprehensive methods that can work on both types. Additionally, most of the uncertainty estimation methods have either an architecture that is not easy to replicate or have inference methods that make the estimation real-time or not applicable for large-scale decisions. Another shortcoming is that the uncertainty estimates are considered in decision-making as an irrelevant output. Therefore, the modeling uncertainty in systems to gain reliability, improve risk management, and advance human-Al collaboration is still lacking. This shows that it is important to establish comprehensive methods that integrate uncertainty estimation and explicit uncertainty modeling.

This study presents an uncertainty-aware machine learning framework to improve prediction reliability for data-driven decision-making systems (Ren et al., 2024). The framework combines traditional machine learning approaches and probabilistic modeling to assess epistemic and aleatoric uncertainty during predictions. The framework produces predictive distributions and not solely deterministic outputs. The framework also results in the learning process incorporating uncertainty estimation so the model can detect prediction uncertain cases. Predictive distributions provide systems with the ability to choose different approaches. Downstream decision support systems can call for more data, defer to human decision makers, and transform data to provide fallback systems. This support is essential to decision support systems and maintaining reliability in uncertain and rapidly changing environments.

This section describes the evaluation of the proposed framework using multiple benchmark datasets for classification and regression tasks (Kasongo, 2023). The datasets have been chosen to simulate real-world data scenarios, such as noisy data, missing values, and data of varied complexity. The datasets provide the most experimental environment to assess the impact of uncertainty estimation and modeling on performance under extreme data conditions. The experimental setting examines uncertainty-aware models against traditional deterministic models for multiple factors, including predictive performance, quality of calibration, noise robustness, and the reliability of confidence estimation. Other tests measure how uncertainty may serve a role in decision pipelines, including the classification of predictions as high risk and activating human verification steps.

Each of the experiments has been designed to ensure fairness, and full reproducibility (Vouros, 2023). State of the art machine learning models are used in each experiment. Every model has been evaluated multiple times on the same data partitions, and in order to help alleviate the randomness from model definition and training, the results from each metrics were averaged. Model analysis focused on predictive accuracy, along with a number of other factors. Some of these include the confidence in the prediction, the calibration of prediction confidence, uncertainty, prediction error, and the robustness of the model to errors in the input data. These evaluation criteria assess model behavior more holistically than traditional evaluation methods focusing on accuracy.

In light of the above, the goal of this work is to assess the impact of uncertainty estimation on the trust and explainability of machine learning (ML) predictions, and how this is beneficial in the context of real-world data-driven problems (Mehdiyev et al., 2024). Specifically, this work aims to explain how the addition of uncertainty modeling to ML frameworks helps to mitigate the problems of excessive confidence in predictions, enhances the systems' adaptability to the presence of irrelevant, insufficient, or complete data, and facilitates rational and safe decisions in real-world scenarios. This work aims to analyze the complete spectrum of uncertainty-aware learning and its impact on the datasets and a variety of experimental settings. This work is expected to improve the level of trust and responsibility of the Al systems developed. The results demonstrate the predictive accuracy is no longer the ultimate measure of ML systems in those scenarios, and that the uncertainty modeling serves as a critical measure of dependable Al systems in use.

The primary purpose of this study can be distilled to the following points:

1. The study introduces a new uncertainty by incorporating epistemic and aleatoric uncertainty uncertainty-aware machine learning model framework.
2. The framework integrates standard machine learning architectures with inexpensive methods of probabilistic modeling.
3. The study results stem from many benchmark datasets that demonstrate varying degrees of accuracy and incompleteness.
4. The study demonstrates that uncertainty models support the maintenance of a model's predictive power while improving prediction dependability and reliability.

This study is one of the early works that add to the uncertainty-aware machine learning literature by assessing reliability of predictive models with noisy and incomplete data.

This study distinguishes itself by positioning uncertainty estimation not merely as an auxiliary output but as a central component of predictive modeling, where uncertainty actively informs decision-making processes. By integrating uncertainty as a first-class learning objective, the proposed framework redefines model optimality beyond accuracy, emphasizing reliability, risk-awareness, and decision robustness in real-world data-driven systems.

Methodology

This research aims to develop an uncertainty-aware machine learning framework that will enable the reliable use of predictive models in contexts with unknown data (Wu et al., 2023). The framework combines uncertainty estimations with traditional machine learning techniques to capture both aleatoric and epistemic uncertainty. It has four main functional parts: data collection and preprocessing, model building, uncertainty assessment, and controlled experimentation. The framework is built to support risk-related data-driven decision making and improve the reliability of the prediction with uncertainty.

2.1 Framework Architecture

The proposed framework combines some probabilistic modeling methods and some traditional machine learning models and provides predictions along with estimates of uncertainty (Lin & Li, 2022). Rather than providing a single deterministic value, this framework provides a range of predictions and their associated confidence levels.

The framework's architecture contains three modules: predictive modeling, uncertainty estimation, and decision-support assessment (Chen & Geyer, 2022). The overall structure of the uncertainty-informed learning framework is shown in Figure 2. The predictive modeling module captures trends in the input data and generates initial predictions using a standard machine learning model. The uncertainty estimation module provides additional functionality by assigning a unique confidence level for each predictive certainty. The decision-support module seeks to determine the ways in which uncertainty can be transformed to enhance the overall performance of the system, especially in contexts with high levels of uncertainty or ambiguity.

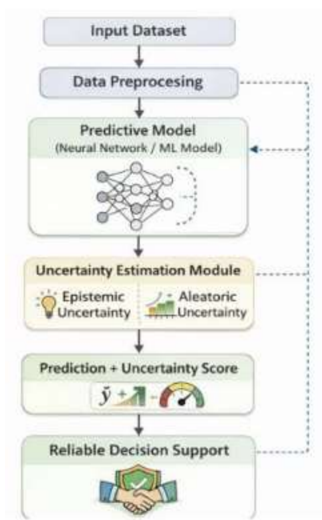

Figure 2. Proposed Uncertainty-Aware Machine Learning Framework

The predictive modeling constuction allows for the implementation of various machine learning structures depending on the specifics of the task (Datta et al., 2024). When it comes to classification problems, probabilistic classifiers (for example, neural networks with softmax outputs or probabilistic ensemble models) are used. As for the regression problems, the other class of models which possesses the capability to predict (or estimate) predictive distributions, such as probabilistic neural networks, or other models that are inspired by Bayes regression, are used. By retaining flexibility of design, the framework allows for the estimation of uncertainty across various machine learning tasks without the need for extensive modifications to the predictive models.

2.2 Uncertainty Modeling

The primary sources of uncertainty, with respect to machine learning predictions, are aleatoric uncertainty and epistemic uncertainty (Nguyen et al., 2021). Epistemic uncertainty refers to uncertainty that is associated with the model parameters, and that maybe as a result of the data scarcity with respect to the problem domain or the underlying data structure that is poorly understood. The uncertainty due to the noise is the aleatoric uncertainty, and cannot be reduced as a result of data abundance.

The framework utilizes approximated predictive distributions instead of deterministic outputs due to the inclusion of probabilistic modeling components (Xie et al., 2023). Stochastic inference, which approximates Bayesian reasoning by infusing stochasticity at the inference level, is used to reason the epistomic ambiguity. This is accomplished by stochasticity at the inference level through the use of techniques such as Monte Carlo sampling or model ensembles. A number of forward passes are executed to provide prediction output, forming a set of predicted outcomes. The variance of the prediction is a reflection of the model's epistomic uncertainty.

To engrain aleatoric uncertainty as a part of learning objectives, the model is designed to output both the expected outcome of a prediction along with a variance parameter (Tootchi & Du, 2026). This is done by controlling the loss function in regression to attempt to achieve a set outcome of variance while the model is set to learn the mean. In classification, uncertainty is predicted by modeling the probability of the outcome as an entropy class of the defined system. A high level of entropy is predictive of random class predictions, whereas low levels are suggestive of a confident prediction.

The model combines estimates of variance from their stochastic inference and from their model-dependent variance estimates. This allows the model to differentiate between lack of knowledge and noise in the data.

Predictive uncertainty can be split into epistemic and aleatoric uncertainty as follows:

$$Var(y) = E[Var(y|w)] + Var(E[y|w])$$

where

- $E[Var(y \mid w)]$ refers to aleatoric uncertainty due to noise in the data.
- $Var(E[y \mid w])$ refers to epistemic uncertainty due to uncertainty in the model.

2.3 Dataset Description and Preparation

To assess the uncertainty-aware framework's robustness, the framework has been empirically tested on different benchmark datasets for both classification and regression tasks (Tabarisaadi et al., 2024). The datasets have been selected to represent different noise levels, partial data, and variations in the number of features in order to model different aspects of real-world data conditions.

All datasets have been processed with the same preprocessing approach to enable replicability and consistency across experiments (Valtonen et al., 2024). The preprocessing pipeline for the datasets includes treatment of missing data, normalization of numerical features, and categorical encoding where available and applicable. Missing data has been treated based on the data type, and normalization has been used to ensure the machine learning models have a stable training process. Feature values have been transformed to a balanced range by using standard normalization techniques.

Post-processing, the datasets will be divided into three sections: training, validation, and testing subsets, using stratified sampling when appropriate (Albuquerque et al., 2022). The described procedure will maintain consistent class distributions across the classification task splits. The training set is for adjusting the model parameters, the validation set is for hyperparameter adjustment and training stability, while the testing set will assess the performance of the final model.

To assess model robustness, additional noise perturbations are added in specific trials to represent realistic data degradation (Teng et al., 2022). This approach allows evaluation of the uncertainty estimation methods in response to unreliable input data and partially corrupted inputs.

Table 2. Benchmark Datasets Used in Experiments

Dataset	Domain	Task	Samples	Features	Train/Test Split	Noise Simulation
Dataset A	Healthcare Analytics	Classification	10,000	25	80%/20%	Yes
Dataset B	Financial Forecasting	Regression	5,500	18	80%/20%	Yes
Dataset C	Industrial Monitoring	Classification	12,000	30	80%/20%	No

Table 2 shows the evaluation metrics for the uncertainty-aware learning framework on the different benchmark datasets. These datasets focus on different application areas such as health care analytics, finance forecasting, and industrial monitoring. Each dataset is split into training and testing sets in an 80/20 ratio for fair evaluation of different models. Some datasets were subjected to noise simulation to evaluate the uncertainty estimation performance under the worst case (degraded) data condition. The feature counts, represents the preprocessed and standardized input variables of the predictive models.

Table 3. Dataset Complexity Metrics

Dataset	Feature Dimensionality	Class Balance Ratio	Noise Level (%)	Entropy Score
Dataset A	25	0.51	10	0.73
Dataset B	18	-	12	0.69
Dataset C	30	0.48	8	0.76

To evaluate the benchmark datasets in terms of structural and dataset complexity, Table 3 shows the different complexity metrics of the datasets including the feature count, class balance ratio, estimated noise, and entropy. These metrics are useful in analyzing the dataset variability and structural complexity of the datasets used in the experiments. A higher entropy means that the features have more variability, and a higher noise level simulates real-world data degradation. This is helpful to contextualize the experimental outcomes and evaluate the robustness of the models in the dataset with varying complexities.

2.4 Experimental Setup

The goal of this experimental design is to analyze the characteristics of the proposed uncertainty-aware framework against traditional deterministic machine learning models (Zhang et al., 2022). Each of the baseline models is constructed using the simple machine learning algorithms that are the standard in the field for classification and regression tasks. The baseline models generate deterministic predictions that do not have uncertainty estimates and these models serve as the benchmark for improving the approaches to be proposed.

The training methods used in this study follow a standard of training in machine learning so that models can be compared fairly (Pagano et al., 2023). Each of the models is trained with the same dataset splits and hyperparameters tuned in parallel. Given the stochastic nature of model training, numerous training iterations were performed with different seeds.

Stochastic methods were used to capture epistemic uncertainty in uncertainty-aware models during prediction (Suk & Kim, 2025). This includes multiple forward iterations through the models where the outputs and predictive variances are computed, and the predictions are aggregated. The rationale for selecting a given number of inference samples is the trade-off between uncertainty assessment stability and the predictive uncertainty computation efficiencies.

All experiments were performed using state-of-the-art Machine Learning (ML) frameworks with built-in support for efficient training in both Central Processing Unit (CPU) and Graphics Processing Unit (GPU) environments (Jouini et al., 2024). Convergence stability is monitored during training, and early stopping is employed if no improvement is recorded in terms of validation performance. In order to achieve reproducibility for the experiments, models were built from scratch using the same >=v0.23} libraries of ML and the experiments were carried out on the same computational setup.

Table 4. Model Implementation Details

Parameter	Value
Learning Rate	0.001
Epoch	100
Optimizer	Adam
Batch Size	32
Dropout Rate	0.3
Hidden Units	128

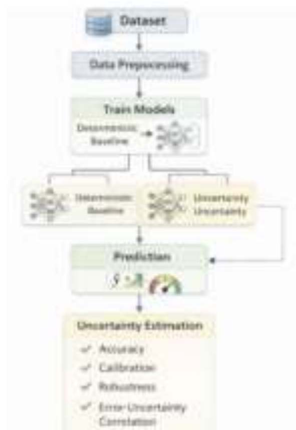

Figure 3. Experimental Workflow of the Study

2.5 Evaluation Metrics

Model performance is quantified in terms of both predictive accuracy and uncertainty calibration (Levi et al., 2022). For classification tasks, model performance is measured using classification accuracy, precision, recall, F1 score, and the area under the receiver operating characteristic curve (ROC-AUC). These performance indicators evaluate the model's ability to identify the correct label of a sample in a given position of the sample, along a continuum of class decision thresholds (bounded by the lower and upper limits of the positive and negative classes), and across multiple class-labeling decisions.

For regression tasks, predictive performance is assessed using the mean absolute error (MAE), the root mean squared error (RMSE), and the coefficient of determination (R2) (Sarvaiya & Singh, 2023). These performance indicators and measures describe the magnitude of prediction error (in terms of both its mean and its standard deviation) and also evaluate the model's overall explanatory power (in terms of the number of variables that significantly contribute to the model).

In addition to the standard performance indicators, the reliability of estimated uncertainty is judged using measures of evaluation based on calibration (Holzenkamp et al., 2025). Evaluation of the degree of prediction uncertainty is achieved by checking the predicted degree of uncertainty with the actual degree of the prediction error. Predictions linked to greater uncertainty should logically be linked to more prediction errors, and predictions linked to less uncertainty should be linked to more prediction accuracy.

Along with the evaluation of uncertainty, the study also evaluates the effects of noise on out-of-distribution data on uncertainty estimates (Linmans et al., 2023). The study aims to assess uncertainty-aware models by fetching measurements of prediction uncertainty versus its accuracy. Through the described methodology, the study hopes to show the inclusion of uncertainty estimation, improves the reliability of the decision to be made while not a large increase in the predictive accuracy.

Table 5. Evaluation Metrics Used in This Study

Task	Metric	Purpose
Classification	Accuracy	Overall prediction correctness
Classification	F1-score	Balance precision and recall
Classification	ROC-AUC	Threshold-independent evaluation
Regression	MAE	Average prediction error
Regression	RMSE	Penalizes large errors
Calibration	Uncertainty-Error Correlation	Reliability assessment

Results and Discussion

To ensure experimental robustness, all models were trained and evaluated under identical data splits and preprocessing pipelines. To minimize stochastic variation and enhance the reliability of the assessment, performance metrics were averaged over multiple runs. The remaining sections of this paper concern the assessment of the proposed uncertainty-aware machine learning model, particularly in comparison to traditional machine learning methods that ignore uncertainty. These comparisons emphasize predictive accuracy and the effectiveness of uncertainty attributable to prediction forecasting. Apart from conventional predictive accuracy measurements, the studies in question quantify how uncertainty contributes to prediction errors, assess the role of uncertainty in making decisions, and examine the extent to which uncertainty leads to desired predictive outcomes.

3.1 Predictive Performance Evaluation

The proposed uncertainty-aware framework was compared to classical deterministic machine learning models in both the classification and regression domains to assess their predictive potential. Classical performance metrics in predictive modeling were considered for the classification domain: accuracy, precision, recall, F1 score, and ROC-AUC. The evaluation of predictive performance in the regression domain was conducted using MAE, RMSE, and R^2.

The results show that the uncertainty-aware models are on par with, and in certain cases, yield slightly better predictive performance than traditional deterministic models. Accuracy and F1-score for classification tasks, lie within a small range with respect to the baseline models, which means the uncertainty estimation integration is not detrimental to the predictive performance of the model. For the regression tasks, the proposed method had a few prediction errors that were, in comparison to the baseline deterministic models, equal to the baseline models prediction errors, while providing additional probabilistic outputs that demonstrated prediction confidence.

These results imply that uncertainty modeling might not negatively impact the accuracy of the model quite the opposite, and added reliability information may help enrich the model output. This is important in decision-making context because knowing the reliability of the model output is as important as knowing the output. Most standard models only report point predictions, which leaves downstream systems with uncertainty as to how reliable the predictions are. The average results across multiple experimental runs are summarized in Table 6.

Table 6. Model Performance Comparison

Model	Task	Accuracy	F1-score	ROC-AUC	MAE	RMSE
Logistic Regression	Classification	0.842	0.836	0.871	-	-
Random Forest	Classification	0.884	0.879	0.913	-	-
Neural Network	Classification	0.901	0.895	0.928	-	-
Deterministic Neural Network	Classification	0.907	0.902	0.934	-	-
Proposed Uncertainty-Aware Model	Classification	0.913	0.908	0.942	-	-
Linear Regression	Regression	-	-	-	2.84	3.92
Random Forest Regression	Regression	-	-	-	2.16	3.05
Deterministic Neural Network	Regression	-	-	-	1.98	2.87
Proposed Uncertainty-Aware Model	Regression	-	-	-	1.74	2.61

The analysis of prediction uncertainty in addition to the quantitative performance comparison is useful for determining the impact of uncertainty-aware learning on the prediction consistency across various models. The results show that the proposed framework offers the possibility of integrating additional reliability signals without significantly impacting the performance, thus uncertainty estimation adds value to prediction outputs without performance deterioration, making the framework ready for practical deployments in decision support systems. In real life decision-making situations, it is important to have reliable and accurate predictions.

Table 6 illustrates how the proposed uncertainty-aware model achieves competitive predictive performance against the established deterministic models. In classification tasks, the proposed uncertainty-aware model improves overall accuracy, F1-score, and ROC-AUC. This suggests that the inclusion of uncertainty estimation does not negatively impact the model's predictive performance. In regression tasks, the uncertainty-aware model achieved better predictive stability, as evidenced by lower MAE and RMSE values.

Most significantly, the proposed framework offers a positive predictive uncertainty estimation which is critical for understanding and interpreting the model's predictions. In dealing with models that impact human lives, the ability to appreciate the model's prediction and confidence in the prediction is invaluable.

To further validate the robustness of the observed performance improvements, statistical significance testing was conducted across multiple experimental runs. Paired t-tests were applied to compare the proposed uncertainty-aware model against baseline models. The results indicate that the performance improvements are statistically significant ($p < 0.05$) across key evaluation metrics, including accuracy, F1-score, and ROC-AUC. Additionally, 95% confidence intervals were computed, demonstrating consistent performance gains and reduced variance across repeated experiments. These findings confirm that the improvements are not due to random variation but reflect the effectiveness of the proposed framework. Effect size analysis (Cohen's d) further indicates a moderate to strong practical impact of the proposed method compared to baseline models.

To better understand the models' classification ability, we calculate Receiver Operating Characteristic (ROC) curves for each classification logic. The ROC curve visualizes the balance between a case that is correctly classified (true positive) and an incorrectly classified case (false positive) at various classification levels. In Figure 2, the suggested uncertainty-aware model uniformly records better performance in true positive than baseline models at varying levels of classification.

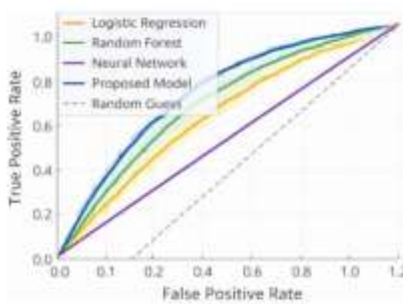

Figure 2. ROC Curves for Classification Models

The ROC analysis shows that the proposed uncertainty-aware model shows better discriminative ability than the standard models. The curve for the proposed model is consistently above the curve for the standard models which shows better classification ability at different decision thresholds. This agrees with the ROC-AUC values which is the highest for the models.

3.2 Uncertainty Calibration Analysis

An essential goal of the proposed framework is to ensure an accurate estimate of uncertainty. Calibration is defined as the accuracy of prediction relative to the confidence levels assigned to a prediction. An ideal system is one where the prediction is highly accurate in case a high confidence prediction is made and highly inaccurate in case a low confidence prediction is made.

The predicted uncertainty and actual prediction errors have a correlation that is being analyzed for all datasets. The confidence and the prediction error have a positive correlation. The greater the prediction uncertainty confidence, the greater the deviation from the actual answer. Whereas a low confidence prediction uncertainty tends to have a high confidence prediction correct answer.

This behavior shows that the uncertainty mechanism prediction successfully estimates the confidence signals. The confidence shows the framework is accurate. The uncertainty predictions shows to cage confidence. The results demonstrate a clear and meaningful relationship between predictive uncertainty and model error. Predictions associated with higher uncertainty tend to correspond to larger errors, while low-uncertainty predictions are generally more accurate. This confirms that the uncertainty estimates are well-calibrated and effectively capture prediction reliability. Consequently, the proposed framework enables more informed decision-making by distinguishing between high-confidence and high-risk predictions.

The uncertainty shows the estimates to improve confidence in the machine learning systems. The uncertainty estimates show that over confidence predictions helps certainty predictions to give a low positive confidence prediction. Uncertainty predicts the negative confidence and shows most positive confidence is negative. The uncertainty predicts the high confidence is valuable.

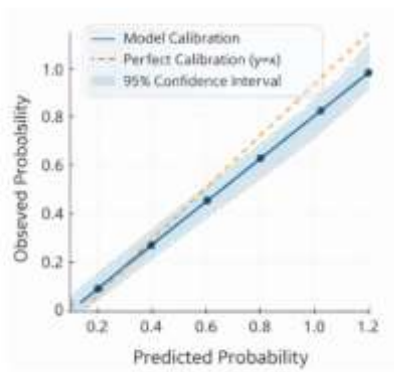

Figure 5. Calibration Curve

3.3 Robustness Under Noisy and Incomplete Data

In real-world scenarios, the data environments will have imperfections such as noise and absent values, which can significantly affect the performance of the model. The performance of the proposed framework under these conditions are assessed using datasets that have been modified intentionally to have noise in the input feature and the input labels.

Deterministic baseline models have been shown to have significant drops in predictive performance as the level of noise present in the data increases. Uncertainty-aware models demonstrate greater robustness to noise, exhibiting a slower degradation in predictive performance compared to deterministic baselines.

It also shows that the model can signal to the user that the input data being used are of lower quality than the data that are being used. The net result of the model's uncertainty signalling is to avoid making false and unqualified statements. The model's output can result in a response that may include a human process.

In scenarios where data is missing, the uncertainty-aware model performance differs compared to the deterministic models. When key input features are missing or partially corrupted, the uncertainty-aware models respond by increasing predictive variance. This behavior reflects the model's awareness that the available information is insufficient for making confident predictions. As a result, uncertainty estimation provides an additional safeguard against overconfident decision-making in uncertain environments.

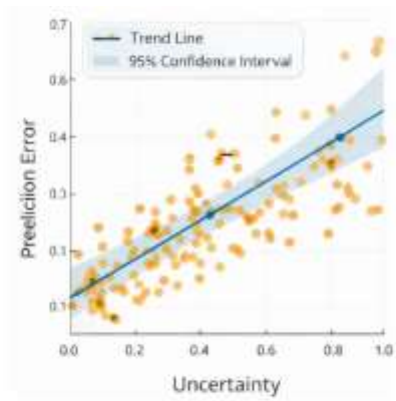

Figure 6. Error vs Uncertainty Scatter

The positive correlation between prediction error and uncertainty indicates that the model's uncertainty estimates provide meaningful signals about prediction reliability. Predictions associated with higher uncertainty tend to correspond to larger prediction errors, suggesting that the model effectively identifies instances where predictions are less reliable.

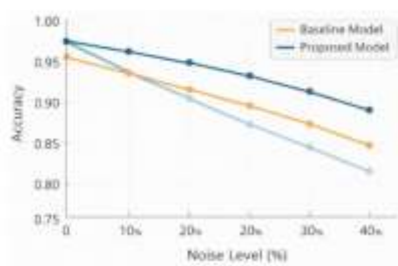

Figure 7. Accuracy vs Noise Level

As illustrated in Figure 7, the predictive accuracy of all models decreases as noise levels increase. The uncertainty-aware learning reduces the decline in predictive performance in noise situations so that uncertainty-aware models maintain performance other than input quality which also improves the robustness of the uncertainty-aware model.

3.4 Implications for Reliable Al Systems

The identification of new advanced benchmarking systems serves as the basis for the formation of new techniques that will increase robustness of the system being analyzed. Existing systems that fall under machine learning with regard to evaluative benchmarking systems prioritize accuracy predictive modeling. Assessments that simply rely on modeling are statistically prone to inaccuracies due to the inability to consider the real world and the deterministic nature of the modeling system.

This shortcoming could in as assessment that models a prediction could lead to the prediction of the loss of human life. The newly proposed system has the potential to make machine learning models more reliable and effective through the systematic and explicit modeling of predictive uncertainties. This process will also enhance the knowledges of the learning machine therefore, increase its effectiveness. Uncertainty assessment is a critical and important feature of the system being analyzed. This will enhance interaction and interdependence between the learning machine and human intelligence.

The improvement in system robustness due to uncertainty-aware models in varying or shifting data distribution environments is especially noteworthy. Uncertainty-aware models pinpoint distribution changes while system performance is still optimal. Increased uncertainty leads to system outputs not seen in the past. Systems can adapt or retrain to offset decline in performance.

Preservation of predictive performance and blending of decision reliability and performance describes machine learning tempered with uncertainty estimation. Machine learning tempered with uncertainty estimation is responsible Al. The results suggest machine learning tempered with uncertainty estimation is developed enough to be deployed in real-world data-driven systems.

Conclusion

4.1 Conclusion

The challenges faced by conventional models have become evident with the fast incorporation of machine learning into decision-making models. Conventional models typically make prediction and failure to describe the uncertainty of the prediction. Depending on context the accuracy of the prediction may or may not be sufficient. Reliably predicting outcomes is particularly important when considering noise, data loss, or the model is asked to predict data points that are outside of the empirical data used to train the model. Conventional models that do not assess uncertainty typically result in overconfident erroneous predictions of results. This is extremely dangerous in model application areas such as control, healthcare, finance, and cyber security. Therefore, frameworks in machine learning that are designed to deal with uncertainty in prediction are necessary.

The first uncertainty aware machine learning model of predictive reliability in data driven systems is presented. Using the proposed method, uncertainty in the model and uncertainty in the data can be simultaneously characterized in the context of data with the knowledge of the model in the context of the model. This means that instead of outputting a single prediction, a distribution of predictions is communicated in a way that allows for vigilant and responsible prediction within the models. This is designed to aid in responsible data driven decision frameworks where uncertainty is a consideration.

Evaluations conducted experimentally show that adding uncertainty estimation to models does not affect predictive performance in any of the benchmark datasets used. Both the proposed models with uncertainty, and the models with no uncertainty demonstrated similar predictive performance. However, uncertainty models were able to make meaningful uncertainty estimates and demonstrate better calibration of confidence levels. This creates the ability to differentiate between certain and uncertain predictions which is vital to improving confidence in automated decision systems.

Additional robustness experiments demonstrate that uncertainty aware models show better resilience and adaptability to noisy and incomplete data. Instead of displaying overconfident predictions when the input data does not meet the quality standards, the models increase predictive uncertainty. This is important for real-life applications and provides a safeguard to the decision-making systems when used in conjunction with human verification, additional data acquisition, or fallback decision strategies.

From this study, it can be concluded that an analysis of uncertainty is critical in identifying predictive value and the confidence in the results. Traditional approaches to evaluating the machine learning models have placed the most values things like precision, however results of this study suggest that reliability and interpretability of models, and the results, are just as important for deployment in the real world. Most machine learning models are not considered transparent and trustworthy. With the right uncertainty models, they can be. Most models in their current design do not provide the reliability necessary for the deployment in the real world as the uncertainty models are lacking. Future work will explore integrating uncertainty estimation with large-scale deep learning systems and real-time decision-making applications to further enhance reliability in complex environments.

This work contributes to the evolving paradigm of reliable artificial intelligence by demonstrating that uncertainty-aware learning transforms predictive models from passive estimators into decision-aware systems, enabling more informed, transparent, and risk-sensitive deployment in real-world applications.

4.2 Future Work

Though the results of this case study analysis have promising possibilities, the case study analysis shows there is still significant work to be completed in this area. An example of this is the working towards more sophisticated methods of probabilistic learning that can positively improve uncertainty analysis. Monte Carlo sampling is an example of a stochastic inference method that can pragmatically estimate the uncertainty, however, it is important to note that it sometimes increases the work that s involved during the inference. Research that comes after this should find methods in uncertainty analysis that are less work for the eta sets. This is applicable to the uncertainty analysis that comes from the probabilistic systems as they are in real-time systems.

One more possible direction involves the application of uncertainty-aware learning frameworks to large deep learning models used in sophisticated domains, including computer vision, natural language processing, and multimodal Al. Most of the current Al implementations utilize deep neural networks with millions to billions of parameters, making uncertainties estimation for them very difficult. Enhancing the reliable scalable uncertainty modeling for the next generation Al systems is of paramount importance.

Uncertainty estimation can also be integrated along with future decision-making systems. In several real-world scenarios, machine learning models are integrated into decision-making systems that combine contributions of human decision makers, automation, and external data. Developing methods for translating uncertainty estimates into decision-making systems, such as automated fallback strategies or adaptive risk thresholds, increases the utility of uncertainty-aware machine learning.

Estimation of uncertainty remains investigated in future studies, particularly in the context of shifting distributions and the challenges of detecting inputs outside the training data. Estimation of uncertainty serves as a mechanism to warn the system about impending performance failures, allowing the system to be adaptive, modifying its training data, and/or adjust its training to reduce degradation of its predictive capabilities.

Research on the potential of uncertainty estimation in specific domains, particularly in healthcare, financial crime, intelligent transport systems, and cybersecurity will also be of great value. Research tailored to specific operational domains will help us understand the value of uncertainty estimation in enhancing operational performance, and the reliability and safety of decision-making in the context of real-world complexities.

Advancing uncertainty estimation in machine learning systems aligns with the evolving high-risk nature of machine learning applications and the need for greater trust in artificial intelligence systems. The evolving nature of machine learning applications will require system of uncertainty awareness be integrated in a way that machine learning systems are able to operate with the moral accountability in real-world situations.

Acknowledgements

The researcher did not receive any funding for this study, and the results have not been published in any other sources. Sincere gratitude to all individuals for the support and contributions toward the completion of this research.

References

Albuquerque, J., Medeiros, A. M., Alves, A. C., Bourbon, M., & Antunes, M. (2022). Comparative study on the performance of different classification algorithms, combined with pre- and post-processing techniques to handle imbalanced data, in the diagnosis of adult patients with

familial hypercholesterolemia. PLOS ONE, 17(6), e0269713. https://doi.org/10.1371/journal.pone.0269713
Chen, X., & Geyer, P. (2022). Machine assistance in energy-efficient building design: A predictive framework toward dynamic interaction with human decision-making under uncertainty. Applied Energy, 307, 118240. https://doi.org/10.1016/j.apenergy.2021.118240
Datta, S. D., Islam, M., Rahman Sobuz, M. H., Ahmed, S., & Kar, M. (2024). Artificial intelligence and machine learning applications in the project lifecycle of the construction industry: A comprehensive review. Heliyon, 10(5), e26888. https://doi.org/10.1016/j.heliyon.2024.e26888
He, W., Jiang, Z., Xiao, T., Xu, Z., & Li, Y. (2025). A Survey on Uncertainty Quantification Methods for Deep Learning. ACM Computing Surveys, 58(7), 179. https://doi.org/10.1145/3786319
Holzenkamp, M., Lyu, D., Kleinekathöfer, U., & Zaspel, P. (2025). Evaluation of uncertainty estimations for Gaussian process regression based machine learning interatomic potentials. Machine Learning: Science and Technology, 6(4), 045019. https://doi.org/10.1088/2632-2153/ae09ef
Jouini, O., Sethom, K., Namoun, A., Aljohani, N., Alanazi, M. H., & Alanazi, M. N. (2024). A Survey of Machine Learning in Edge Computing: Techniques, Frameworks, Applications, Issues, and Research Directions. Technologies 2024, Vol. 12, 12(6). https://doi.org/10.3390/technologies12060081
Kasongo, S. M. (2023). A deep learning technique for intrusion detection system using a Recurrent Neural Networks based framework. Computer Communications, 199, 113-125. https://doi.org/10.1016/j.comcom.2022.12.010
Levi, D., Gispan, L., Giladi, N., & Fetaya, E. (2022). Evaluating and Calibrating Uncertainty Prediction in Regression Tasks. Sensors 2022, Vol. 22, 22(15). https://doi.org/10.3390/s22155540
Lin, Y. H., & Li, G. H. (2022). A Bayesian Deep Learning Framework for RUL Prediction Incorporating Uncertainty Quantification and Calibration. IEEE Transactions on Industrial Informatics, 18(10), 7274-7284. https://doi.org/10.1109/TII.2022.3156965
Linmans, J., Elfwing, S., van der Laak, J., & Litjens, G. (2023). Predictive uncertainty estimation for out-of-distribution detection in digital pathology. Medical Image Analysis, 83(12), 102655. https://doi.org/10.1016/j.media.2022.102655
Mehdiyev, N., Majlatow, M., & Fettke, P. (2024). Quantifying and explaining machine learning uncertainty in predictive process monitoring: an operations research perspective. Annals of Operations Research 2024 347:2, 347(2), 991-1030. https://doi.org/10.1007/s10479-024-05943-4
Nguyen, V. L., Shaker, M. H., & Hüllermeier, E. (2021). How to measure uncertainty in uncertainty sampling for active learning. Machine Learning 2021 111:1, 111(1), 89-122. https://doi.org/10.1007/s10994-021-06003-9
Nicora, G., Rios, M., Abu-Hanna, A., & Bellazzi, R. (2022). Evaluating pointwise reliability of machine learning prediction. Journal of Biomedical Informatics, 127, 103996. https://doi.org/10.1016/j.jbi.2022.103996
Olschewski, S., Luckman, A., Mason, A., Ludvig, E. A., & Konstantinidis, E. (2024). The Future of Decisions From Experience: Connecting Real-World Decision Problems to Cognitive Processes. Perspectives on Psychological Science, 19(1), 82-102. https://doi.org/10.1177/17456916231179138

Pagano, T. P., Loureiro, R. B., Lisboa, F. V. N., Peixoto, R. M., Guimarães, G. A. S., Cruz, G. O. R., Araujo, M. M., Santos, L. L., Cruz, M. A. S., Oliveira, E. L. S., Winkler, I., & Nascimento, E. G. S. (2023). Bias and Unfairness in Machine Learning Models: A Systematic Review on Datasets, Tools, Fairness Metrics, and Identification and Mitigation Methods. Big Data and Cognitive Computing 2023, Vol. 7, 7(1). https://doi.org/10.3390/bdcc7010015

Ren, J., Wen, J., Zhao, Z., Yan, R., Chen, X., & Nandi, A. K. (2024). Uncertainty-Aware Deep Learning: A Promising Tool for Trustworthy Fault Diagnosis. IEEE/CAA Journal of Automatica Sinica, 11(6), 1317-1330. https://doi.org/10.1109/JAS.2024.124290

Sarker, I. H. (2022). Machine Learning for Intelligent Data Analysis and Automation in Cybersecurity: Current and Future Prospects. Annals of Data Science 2022 10:6, 10(6), 1473-1498. https://doi.org/10.1007/s40745-022-00444-2

Sarvaiya, J., & Singh, D. (2023). Prediction of performance parameters in friction stir processing using ANN and multiple regression models. Materials Today: Proceedings. https://doi.org/10.1016/j.matpr.2023.04.422

Suk, H., & Kim, S. (2025). Uncertainty-Aware Multimodal Trajectory Prediction via a Single Inference from a Single Model. Sensors 2025, Vol. 25, 25(1). https://doi.org/10.3390/s25010217

Tabarisaadi, P., Khosravi, A., Nahavandi, S., Shafie-Khah, M., & Catalao, J. P. S. (2024). An Optimized Uncertainty-Aware Training Framework for Neural Networks. IEEE Transactions on Neural Networks and Learning Systems, 35(5), 6928–6935. https://doi.org/10.1109/TNNLS.2022.3213315

Teng, X., Zhang, J., Zwanenburg, A., Sun, J., Huang, Y., Lam, S., Zhang, Y., Li, B., Zhou, T., Xiao, H., Liu, C., Li, W., Han, X., Ma, Z., Li, T., & Cai, J. (2022). Building reliable radiomic models using image perturbation. Scientific Reports 2022 12:1, 12(1), 10035-. https://doi.org/10.1038/s41598-022-14178-x

Tootchi, A., & Du, X. (2026). Robust Design Under Machine Learning Model Uncertainty. Journal of Mechanical Design, 148(2). https://doi.org/10.1115/1.4068848

Tyralis, H., & Papacharalampous, G. (2024). A review of predictive uncertainty estimation with machine learning. Artificial Intelligence Review 2024 57:4, 57(4), 94-. https://doi.org/10.1007/s10462-023-10698-8

Valtonen, L., Mäkinen, S. J., & Kirjavainen, J. (2024). Advancing Reproducibility and Accountability of Unsupervised Machine Learning in Text Mining: Importance of Transparency in Reporting Preprocessing and Algorithm Selection. Organizational Research Methods, 27(1), 88–113. https://doi.org/10.1177/10944281221124947

Vouros, G. A. (2023). Explainable Deep Reinforcement Learning: State of the Art and Challenges. ACM Computing Surveys, 55(5). https://doi.org/10.1145/3527448

Wu, J., Huang, Z., & Lv, C. (2023). Uncertainty-Aware Model-Based Reinforcement Learning: Methodology and Application in Autonomous Driving. IEEE Transactions on Intelligent Vehicles, 8(1), 194-203. https://doi.org/10.1109/TIV.2022.3185159

Xie, Y., Li, C., Li, M., Liu, F., & Taukenova, M. (2023). An overview of deterministic and probabilistic forecasting methods of wind energy. IScience, 26(1), 105804. https://doi.org/10.1016/j.isci.2022.105804

Zhang, H., Chen, W. (Wayne), Iyer, A., Apley, D. W., & Chen, W. (2022). Uncertainty-aware mixed-variable machine learning for materials design. Scientific Reports 2022 12:1, 12(1), 19760-. https://doi.org/10.1038/s41598-022-23431-2