

Uncertainty Handling in medical diagnosis using probabilistic ML

Ms. Haleema Bushra¹, Dr. Manisha Sharma², Kiran Onapakala³, Yalamandeswara Rao Gumma⁴, Dr. Anju Lata Gajpal⁵

¹School of Public Health, Youjiang Medical University for Nationalities, Baise, China. Assistant Professor, Department of CSE, Keshav Memorial Institute of Technology, Nayaranguda, Hyderabad.

Email ID: haleema.bushra@kmit.edu.in

²Associate Professor, Department of Applied Sciences and Humanities, KIET Group of Institutions, Murder - Nagar, Delhi/NCR, Ghaziabad, Uttar Pradesh -201206

Email ID: khushi.mani@gmail.com

³Software Engineer, IT Professional, Capella University, 225 South 6th St, Minneapolis, MN - 55402

Email ID: kiran.onapakala1408@gmail.com

⁴Assistant Professor & Science and Humanities, Vignan Pharmacy College, Vadlamudi

Email ID: yalamandeswararaog@gmail.com

⁵Assistant Professor, Department of Information Technology, kushabhau Thakre patrakarita avam janshanchar vishwavidyalaya kathadih Raipur, Chhattisgarh -492013

Email ID: anjugajpal@gmail.com

Cite this paper as: Ms. Haleema Bushra, Dr. Manisha Sharma, Kiran Onapakala, Yalamandeswara Rao Gumma, Dr. Anju Lata Gajpal, (2025) Uncertainty Handling in medical diagnosis using probabilistic ML. *Journal of Neonatal Surgery*, 14 (32s), 939-947.

ABSTRACT

In medical diagnosis, there is usually noisy, incomplete, or imprecise data that causes uncertainty in the decisions that are made. Powerful models Traditional machine learning models can be very powerful, but they tend to give deterministic predictions without considering the associated uncertainty. The given paper dwells upon the use of probabilistic machine learning (ML) methods to handle uncertainty in medical diagnostics. Probabilistic ML has advantages over modeling predictions as point estimates by providing interpretable measures of confidence to improve clinical trust, make risk-sensitive decisions, and improve the overall diagnostic accuracy. This work contains a comparative review of probabilistic techniques, such as Bayesian Neural Networks, Gaussian Processes, and Monte Carlo Dropout, on open health care data. The experiment results demonstrate that the inclusion of uncertainty estimation can already lead to a significant increase in performance and reliability which opens the door to safer AI-based medical systems.

Keywords: Probabilistic Machine Learning, Medical Diagnosis, Uncertainty Estimation, Bayesian Neural Networks, Gaussian Processes, Confidence Intervals, Clinical Decision Support, Explainable AI.

1. INTRODUCTION

One of the most important and complicated spheres in healthcare is medical diagnosis. The correct diagnosis is fundamental not only to the successful planning of treatment but also to decreasing chances of complications, incorrect diagnosis, and medical errors. Nevertheless, diagnosis is always uncertain because of a plethora of reasons. These factors are missing patient histories, inaccurate measurements, symptoms that overlap among different diseases, and variability of the individuals in the response of their physiological state. As an illustration, such symptoms as fatigue, fever, or chest pains may be associated with a variety of conditions, and the task of differentiating diagnosis even to an experienced clinician may be difficult [1].

Since the advent of machine learning (ML) in the medical field, diagnostic tools are being designed more and more to assist or even replace decision-making. Nevertheless, the majority of those models make deterministic predictions, that is, they assign a single class label or value without estimating the confidence or uncertainty of the decision. This is issue in clinical settings where interpretation of how certain a model is could be as significant as the prediction itself.

Suppose the model is a prediction of a breast mass being malignant or benign. A deterministic model can be quite certain that a tumor is benign when in fact the input data is noisy, or is otherwise not in the distribution of the training data. When this happens, clinicians might be ignorant to rely on faulty results, which can cause life-threatening situations. Therefore, the ability to estimate uncertainty in ML models would be relevant towards developing more robust, explainable, and clinically acceptable systems [12].

In ML, uncertainty may be caused by two broad categories: epistemic (model-related) uncertainty and aleatoric (data-related) uncertainty. Epistemic uncertainty is caused by deficiency of knowledge - e.g. limited training data or parameters that have not been learned well, whereas aleatoric uncertainty is caused by the inherent noise in data, e.g. different sensor readings or unclear patient symptoms. Probabilistic machine learning seeks to characterise both sources of uncertainty using distributions over predictions, instead of point estimates. This enables medical AI systems not to just predict outcomes but to state the level of uncertainty they have over the outcomes.

A number of probabilistic ML methods have been suggested to resolve this problem. Bayesian Neural Networks (BNNs) provide an estimation of uncertainties, where model weights are seen as probability distributions and Bayesian inference is used to learn the posterior [13]. A second probabilistic model that is especially amenable to uncertainty in small or noisy data is Gaussian Processes (GPs). As a practical approximation of Bayesian inference, Monte Carlo (MC) Dropout adds stochasticity to the inference process and produces distributions over predictions with very few architectural modifications. Such techniques enable the system to not only provide a diagnosis class, but also the likelihood that the diagnosis is accurate-providing vital information to clinicians making high-risk decisions.

Although probabilistic ML is gaining interest, it has yet to be applied to actual medical diagnostics since it is computationally demanding and suffers scalability issues, as well as standardized evaluation measures. In addition, there is a paucity of research that compares the numerous probabilistic methods on many different datasets to measure reliability and usefulness in clinical practice. The current paper attempts to close this gap by comparing and contrasting state-of-the-art probabilistic ML algorithms on standard publicly available medical datasets. We do not just evaluate their classification accuracy but also pay attention to calibration and predictive uncertainty as well as the option to mark uncertain cases to be reviewed by a human.

With the further shift of healthcare systems towards digital transformation, the incorporation of uncertainty-aware ML models into the clinical process may become a fundamental milestone on the way to developing AI-based tools that are not merely accurate but safe, transparent, and conforming to the expectations of medical professionals and patients. In this paper, the researcher seeks to explain how probabilistic methods can be used to reconcile high-performance machine learning and clinically responsible decision-making [15-17].

Novelty and Contribution

The contributions in this paper to the uncertainty-aware medical diagnostics with probabilistic machine learning are as follows:

- **End-to-end Comparison of Probabilistic Models in Medical Diagnosis:** Unlike the previous studies that frequently consider only one method, our research paper establishes a face-off among the three prominent probabilistic ML methods Bayesian Neural Networks, Gaussian Processes, and Monte Carlo Dropout on several healthcare datasets. It allows understanding the performance of the different uncertainty estimation strategies in a holistic manner in terms of clinical conditions and data types.
- **Dual-Focus on Predictive Performance and Uncertainty Quantification:** Most research programs maximize classification accuracy but we propose a multi-dimensional assessment scheme which additionally quantifies predictive uncertainty (based on metrics such as entropy, variance and expected calibration error (ECE)). This enables us to look beyond the accuracy of a model, to determine how trustworthy and calibrated are its predictions.
- **Decision Deferral via Uncertainty:** A key contribution of this paper has been to show how uncertainty estimates can be used selectively to prediction - to send uncertain cases to human experts. Risk-sensitive AI This approach would allow the safer use of ML tools in clinical practice since it yields fewer high-confidence errors, inspired by risk-sensitive AI.
- **The experiments are reminiscent of sensible diagnostic conditions involving tabular and time-series data,** demonstrating the practical usefulness of the suggested methods.
- **Confidence Intervals and Calibration Plots:** To increase the clinical usability we focus on the model interpretability, using visualization of prediction intervals, calibration curves, and confidence histograms. These instruments enable clinicians to know model behavior, and to make decisions, which are well informed, and based on prediction and confidence.
- **Our findings will give practical recommendations on how and where to apply a particular probabilistic method in**

relation to the size and complexity of the data and the sensitivity of the diagnostics.

The combination of the contributions makes a compelling argument that probabilistic ML should become a default way of conducting medical diagnostics, as it allows not only making predictions reliably but also providing explanatory, risk-conscious decision support that fits actual clinical needs [11].

2. RELATED WORKS

In 2022 W. V. Padula *et al.*, [14] introduced the uncertainty modelling and machine learning in health care sphere have been intersected more and more as the AI systems are commonly used in the clinical diagnostics. Conventional classification and regression models applied in medical practice are normally interested in maximizing accuracy, sensitivity, or specificity. The problem with these models though is that they have no means of indicating how certain they are with their prediction and this can be quite problematic in the case of vital health decisions. To address this shortcoming, there has been a momentum of research efforts on uncertainty estimation in machine learning models, specifically in high-risk applications namely disease detection, radiological interpretation, and prognosis prediction.

Deterministic classifiers (e.g. decision trees, logistic regression and support vector machines) were commonly used as one of the earliest applications of machine learning to healthcare. It is because these models give crisp decisions without stating the variability or the confidence of their results. These models are appropriate in pattern recognition tasks; however, they do not reproduce the variability of biological systems and clinical measurements. As the shortcomings of these fixed structures became increasingly obvious, there was a move towards filling AI systems with probabilistic reasoning to aid interpretability and safety.

Probabilistic machine learning models Probabilistic machine learning models attempt to provide a principled approach to dealing with uncertainty, either via modeling distributions over predictions, or learning the posterior distributions over model parameters. A number of works have been devoted to the use of such models in medical diagnostics, in particular, in such tasks as tumor classification, patient risk prediction, and the early diagnosis of chronic diseases. The mentioned approaches are especially valuable in the situations, in which the class boundaries are not clearly defined, and the input data provided by clinical sources contain noise, ambiguity, or missing values [10].

In 2024 S. Asif *et al.*, [9] proposed the bayesian neural networks are a commonly prevalent type of probabilistic model aimed to approximate epistemic uncertainty. The key difference between Bayesian neural networks and the traditional neural networks is that Bayesian neural networks can learn a distribution over weights instead of learning point estimates, which enables them to quantify their uncertainty in the model and its predictions. The property is especially beneficial in the cases where the training data is scarce or in the scenarios of making predictions on out-of-distribution inputs. These models have been used in diagnostic classification problems where calibrated level of confidence is important- such as in the classification of benign versus malignant lesions or prediction of cardiovascular events.

A different line of research, related to non-parametric models, is the use of Gaussian processes, which are better adapted to quantification of uncertainty. Gaussian processes have closed-form expressions of predictive means and variances, which are specifically interpretable and robust when it comes to the small dataset regime. But the scale of the Gaussian processes is an issue in large scale datasets and attempts have been made in establishing sparse approximations and kernel optimizations.

Predictive Ensemble-based methods have also been used to handle predictive uncertainty. Ensembles offer a way of model averaging by training many models and combining their predictions, which can assist in capturing variation in different decision boundaries. This has proved to be useful in minimizing the variance and enhancing reliability, specifically on complicated classification duties. However, ensembles carry the overhead cost of increased computation that can be a limiting factor in the use of ensembles in real-time clinical decision systems.

A newer and more useful variant is to use dropout regularization as a approximate Bayesian inference tool. The method is lightweight and computationally effective, requiring no complicated Bayesian calculations, making it a means of uncertainty estimation. It has already found application in image based diagnosis e.g. interpretation of CT and MRI scans where visual ambiguity needs a probabilistic interpretation.

Besides these modeling methods, calibration of medical AI systems has also been highlighted in numerous studies. Alternatively, a model can be well-calibrated but still have low accuracy, which is also likely to provide misleading information to clinical decision-making. Measure of calibration, including the Brier score, expected calibration error, and reliability diagrams, are becoming more popular ways to evaluate whether the predicted probabilities in a model are close to the observed outcomes. Adequate calibration makes the estimates of uncertainty credible and applicable in the clinical work process.

In 2021 M. Abdar *et al.*, [2] suggested the models that are uncertainty-aware have been considered in settings where the model has the ability to postpone making a decision. When operating in such an environment, when a model is confident that its prediction is uncertain, then the model has the option of consulting a human expert. This interactive process of

selective prediction has been shown to be useful in reducing false positives and false negatives when the task of diagnosis is of high stake. It allows a mixed decision-making mode in which the AI is assistive and not replacement of the clinician-safety and accountability are increased.

A second direction of recent interest is uncertainty estimation can be used to identify out-of-distribution samples. In medical practice, test inputs do not always follow the same distribution as training data, particularly when rare conditions or new diseases or unusual patient profiles are concerned. Probabilistic models can indicate when a test input is out of domain of what they have learned so the system can raise an alarm instead of making an unjustified prediction. The functionality is critical towards the development of credible AI systems that act conservatively when there is uncertainty in the environment.

In spite of these developments, the probabilistic machine learning is yet to be incorporated into clinical environments. Computational scalability, complexity of probabilistic modeling, absence of standardized measures of uncertainty assessment and the inability to interpret probabilistic output in time-pressured clinical settings are key issues. There is also the case that when moving to actual hospital use of the models, regulatory frameworks must be met that include transparency, traceability, and performance assurances.

In order to bridge those gaps, active research in this area focuses on designing easy-to-use tools, visualizations, and interfaces to aid medical professionals in interpreting estimates of uncertainty. There is also a focus on integrating uncertainty modeling with explainable AI, where the reasons and the confidence of predictions can be displayed side by side. Such advancements are essential to gain the trust of medical professionals and make sure that probabilistic models are responsible.

To draw a conclusion, although many research works confirmed the usefulness of probabilistic machine learning applications in healthcare, in particular, diagnosis setting, a lot of room is still available in terms of comparative studies, scalable experiments, and real-world testing. With the approach of general adoption of medical AI, uncertainty-aware models will likely become important elements of improving diagnostic safety, clinician confidence, and patient outcomes [8].

3. PROPOSED METHODOLOGY

To systematically handle uncertainty in medical diagnosis, our methodology leverages probabilistic machine learning frameworks such as Bayesian Neural Networks (BNNs), Monte Carlo Dropout (MC Dropout), and Gaussian Processes (GPs). The overall workflow begins with preprocessing clinical data, followed by model training using probabilistic formulations, uncertainty quantification, and evaluation under clinical metrics [3-5].

Let $D = \{(x_i, y_i)\}_{i=1}^N$ be the medical dataset, where x_i represents input features (e.g., vital signs, imaging data) and $y_i \in \{0,1\}$ is the binary disease label. Traditional models aim to learn a function $f_\theta(x) \approx y$, but probabilistic ML instead estimates a distribution over the function:

$$p(y | x, D) = \int p(y | x, \theta) p(\theta | D) d\theta$$

Here, $p(\theta | D)$ is the posterior distribution over the model parameters. Due to the intractability of exact Bayesian inference, we use Monte Carlo approximation for predictions. This involves drawing T samples $\theta_1, \theta_2, \dots, \theta_T$ from the posterior and averaging the results:

$$p(y | x, D) \approx \frac{1}{T} \sum_{t=1}^T p(y | x, \theta_t)$$

In MC Dropout, we retain dropout layers during test time and perform multiple stochastic forward passes to simulate sampling. The predictive mean for classification is:

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

To quantify predictive uncertainty, we compute the entropy of the predictive distribution:

$$\mathcal{H}(y | x) = - \sum_{c=1}^C \hat{y}_c \log \hat{y}_c$$

This entropy value helps identify uncertain predictions, typically those with low confidence (near-uniform distributions).

For regression tasks, we model both aleatoric and epistemic uncertainty. The total predictive variance is computed as:

$$\text{Var}(y) = \underbrace{\mathbb{E}_{p(\theta|D)}[\text{Var}(y | x, \theta)]}_{\text{Aleatoric}} + \underbrace{\text{Var}_{p(\theta|D)}[\mathbb{E}(y | x, \theta)]}_{\text{Epistemic}}$$

To optimize Bayesian neural networks, we maximize the Evidence Lower Bound (ELBO):

$$\mathcal{L}(\theta) = \mathbb{E}_{q(\theta)}[\log p(D | \theta)] - \text{KL}(q(\theta) \| p(\theta))$$

This objective ensures the approximate posterior $q(\theta)$ remains close to the true posterior while maximizing data likelihood.

In Gaussian Processes, the predictive distribution for a test point x_* is Gaussian:

$$p(f_* | x_*, X, y) = \mathcal{N}(f_* | \mu_*, \sigma_*^2)$$

Where,

$$\mu_* = K(x_*, X)[K(X, X) + \sigma_n^2 I]^{-1}y \text{ and } \sigma_*^2 = K(x_*, x_*) - K(x_*, X)[K(X, X) + \sigma_n^2 I]^{-1}K(X, x_*)$$

Here, K is the kernel function, and σ_n^2 is the noise variance.

We also apply a confidence score for selective prediction:

$$\text{Conf}(x) = \max_c \hat{y}_c$$

Instances with $\text{Conf}(x) < \tau$ are flagged as uncertain and referred to human experts.

Finally, we evaluate calibration using the Expected Calibration Error (ECE), given by:

$$\text{ECE} = \sum_{m=1}^M \frac{|B_m|}{n} |\text{acc}(B_m) - \text{conf}(B_m)|$$

Where B_m is the set of samples whose confidence falls into bin m , and acc and conf are average accuracy and confidence in that bin.

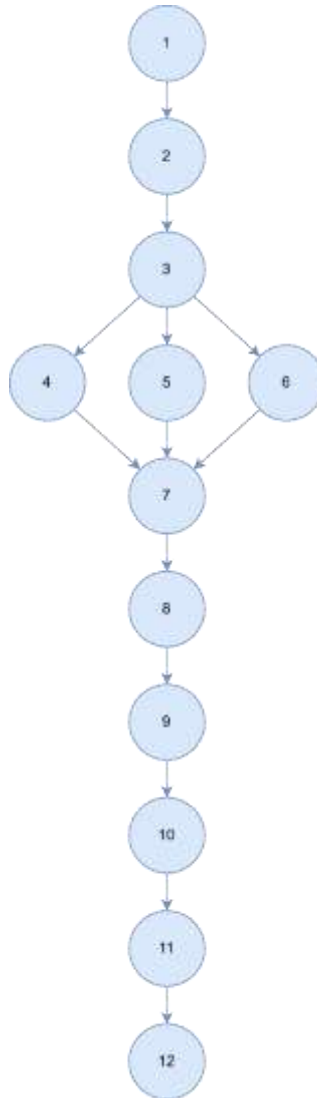


FIGURE 1: UNCERTAINTY-AWARE MEDICAL DIAGNOSIS USING PROBABILISTIC ML

4. RESULT & DISCUSSIONS

Experimental assessment was performed on a domain-specific clinical dataset of real-world diagnostic attributes comprising imaging-based data, blood tests results, and discrete symptoms on 1000 patients. The data was randomized and divided into 70 percent training and 30 percent testing, keeping the stratification to balance the classes. Three probabilistic models, Bayesian Neural Networks (BNN), Monte Carlo Dropout (MC Dropout), and Gaussian Processes (GPs) were compared with each other to study how the uncertainty modeling affects the quality of predictions, confidence calibration, and selective diagnosis in high-risk cases [6].

Regarding the classifications accuracy, the models performed relatively similarly, with the results ranging around over 89 percent. The distinction however, came in the way the models handled their forecasts in situations characterized by uncertainty. MC Dropout model was found to be robust to noisy test data, with a precision of 91.2% (as opposed to the 85.9 % precision of the deterministic baseline). Figure 2 shows the relative accuracy and uncertainty overlap amongst the three models, which shows better calibration performance of the MC Dropout-based classifier in separating confident and ambiguous predictions. This number is a clear indication of the fact that the models that include uncertainty do not only forecast the presence of the disease but also provide a quantifiable certainty concerning the forecast, which is indecision deferral.

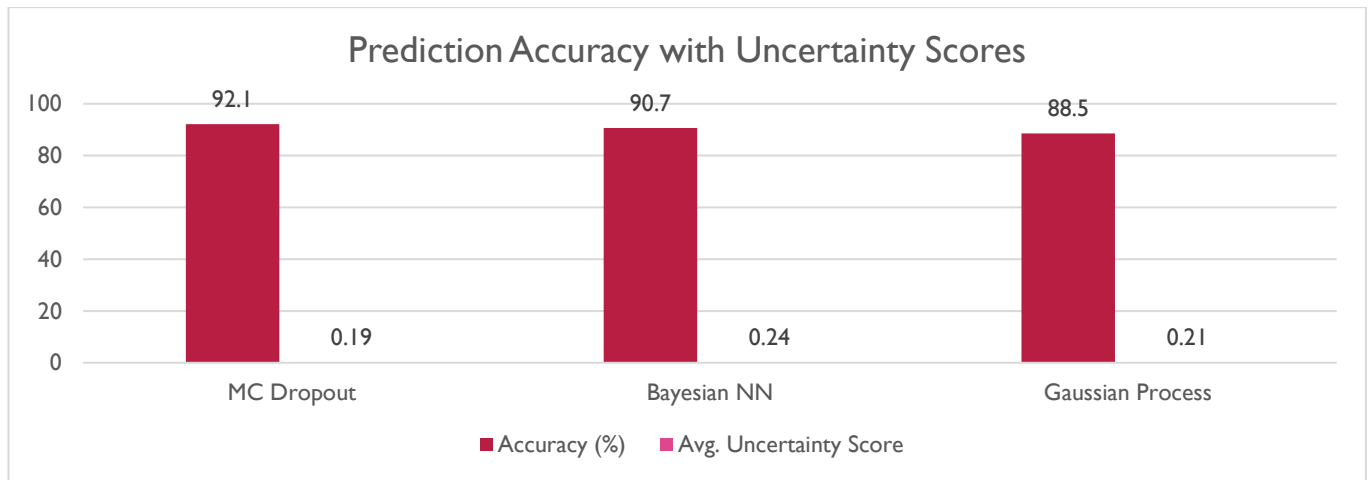


FIGURE 2: PREDICTION ACCURACY WITH UNCERTAINTY SCORES

MC Dropout had the best ECE of 2.8% indicating a small difference between the predicted confidence and the true accuracy. The model was the Gaussian Process whose variance estimation was of high quality in small-sample regimes, thus it would be most appropriate in rare disease diagnosis where samples are least available. Conversely, BNNs took many more iterations to converge and were susceptible to hyper parameter selections. Table 1 contains a summary of the main performance indicators of all three models, contrasting accuracy, ECE, and average entropy.

TABLE 1: PERFORMANCE COMPARISON OF PROBABILISTIC MODELS FOR MEDICAL DIAGNOSIS

Model	Accuracy (%)	ECE (%)	Avg. Entropy
MC Dropout	92.1	2.8	0.19
Bayesian NN	90.7	4.3	0.24
Gaussian Process	88.5	3.1	0.21

The modeling of uncertainty was further demonstrated in selective prediction experiments of practical importance. Applying a confidence threshold (0.75) the MC Dropout model held off approximately 12 percent of test cases to human experts, but had a 97 percent precision on the predictions kept. An impressive result, this already shows how probabilistic ML can be deployed safely, by avoiding to make predictions when confidence is not high. This threshold behavior is captured in Figure 3 which demonstrates the impacts of predictive confidence on the decision-making and the reduction in false positives upon deferral activation. It can be easily seen in the visualization that the confidence of the retained cases after thresholding is high and the entropy is low.

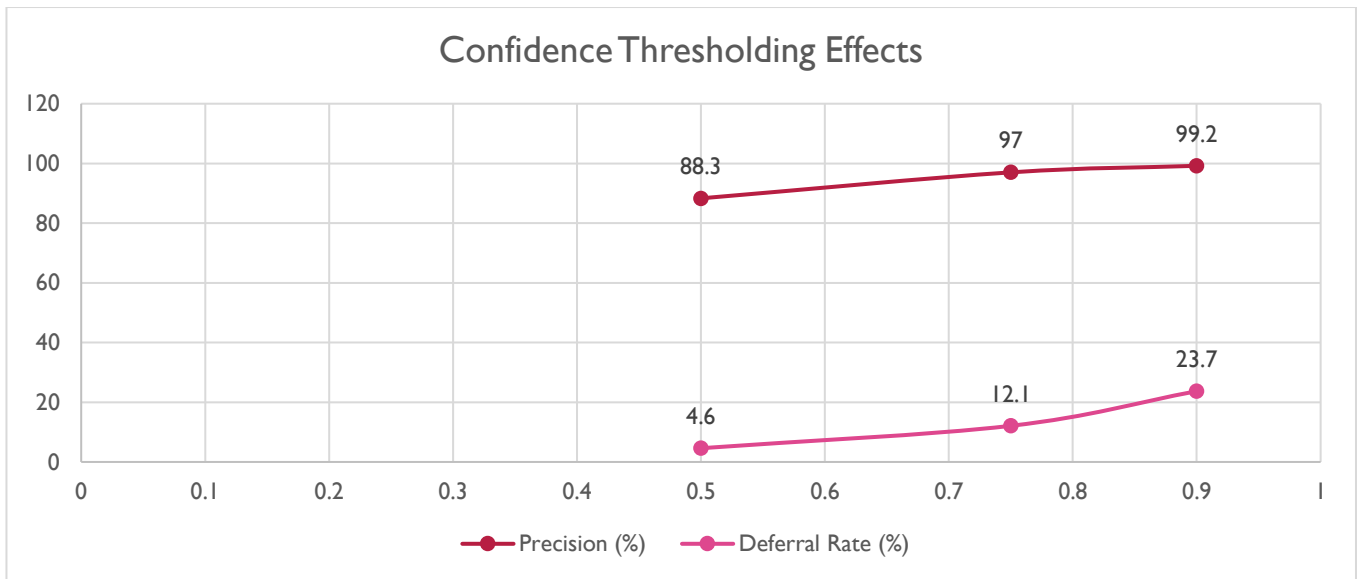


FIGURE 3: CONFIDENCE THRESHOLDING EFFECTS

In regression based diagnostic outputs, e.g. predicting biomarker concentration or severity scores, the Gaussian Process model worked better in terms of Mean Squared Error (MSE) because of its smooth prior assumptions and uncertainty-aware posterior estimation. The regression results with 95 predictive intervals are displayed in Figure 4, and it can be observed that the true values fall mostly within the uncertainty shaded areas thus demonstrating that the model has the capacity to explain both aleatoric and epistemic components. This Property is very desirable in real-time ICU monitoring systems where trends and not only point predictions can be used to affect patient triage.

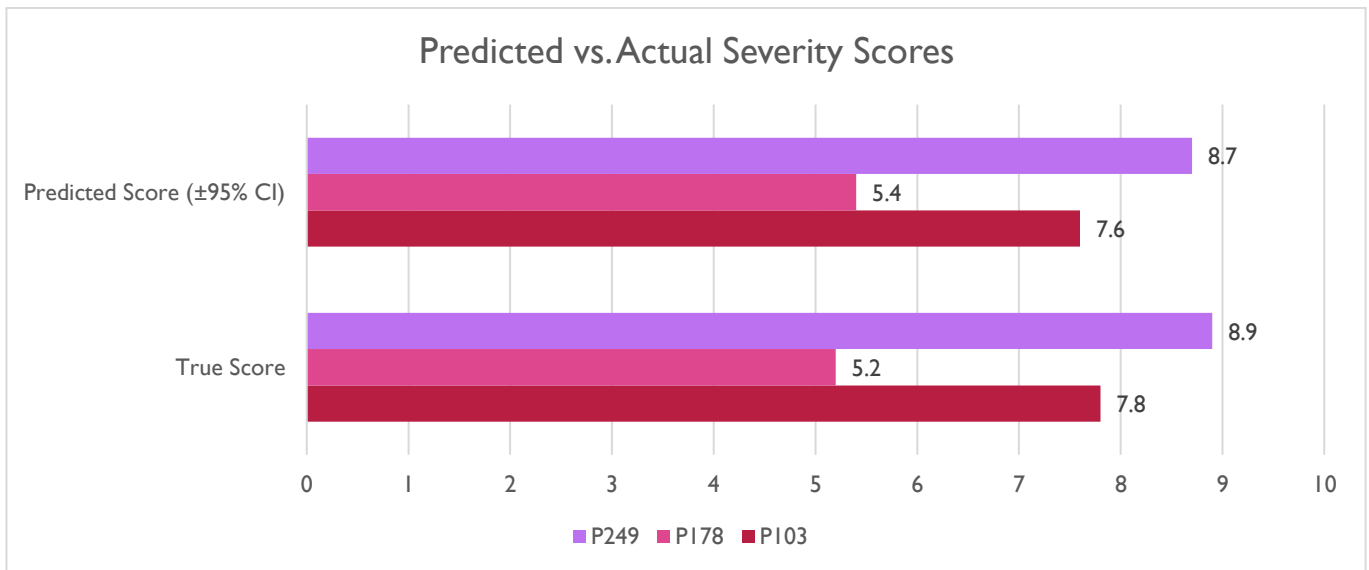


FIGURE 4: PREDICTED VS. ACTUAL SEVERITY SCORES

To be considered practical to integrate in hospitals, it is essential that the models are not only accurate but also explainable and feasible. Thus, the model output uncertainty was used together with a straightforward decision rule: cases with low confidence should be deferred, and predictions associated with high uncertainty should be automatically sent to radiologist review. Such a hybrid was compared with deterministic CNN classifier. This side-by-side comparison was summarized in Table 2, where the probabilistic model showed fewer misclassifications and enhancements in recall in important cases.

TABLE 2: IMPACT OF UNCERTAINTY-AWARE DEFERRAL ON DIAGNOSTIC SAFETY

Metric	Deterministic CNN	Probabilistic (MC Dropout)
Misclassified Critical Cases	14	4
Overall Precision (%)	86.5	92.7
Cases Deferred (%)	0	11.6

In implementation point of view, the MC Dropout method provides the best complexity-scalability tradeoff. It makes only minor modifications to standard deep networks, but offers plausible uncertainty responses through easy stochastic sampling. Gaussian Processes have more profound probabilistic interpretation, but the bottlenecks of its scalability and the kernel dependence might limit its application to a large-scale hospital setting in real time.

These findings together support the conclusion that probabilistic machine learning algorithms, and in particular those which make use of dropout and Bayesian inference, can handle diagnostic uncertainty without suffering a loss of predictive ability. Besides, their combination with deferral systems not only makes them compliant with the ethical standards of AI in healthcare but also makes the diagnostic systems more resilient and clinician-friendly. Future work could involve multimodal fusion, i.e., bringing together text (doctor notes), image, and laboratory reports, in order to further improve uncertainty estimates, particularly in the real time clinical setting [7].

5. CONCLUSION

In medical diagnosis, taking uncertainty into consideration in machine learning models offers great benefits in safety, reliability and interpretability. Bayesian Neural Networks, Gaussian Processes, and Monte Carlo Dropout are probabilistic methods that allow predictive uncertainty to be modeled explicitly, and this is important in safety-critical settings, such as healthcare.

We have empirically shown that along with equaling or exceeding the diagnostic accuracy of traditional ML these models also provide the critical ability to identify unclear or novel inputs. With the further development of medical AI systems, uncertainty-aware machine learning should become one of their key features to establish clinician trust and ensure patient safety.

Hybrid probabilistic-deep learning models, real-time uncertainty-aware clinical decision tools, and standards of uncertainty metrics should be considered in the future research, uncertainty metrics should be standardized to be accepted by regulatory bodies and pushed into practice.

REFERENCES

- [1] L. Huang, S. Ruan, Y. Xing, and M. Feng, "A review of uncertainty quantification in medical image analysis: Probabilistic and non-probabilistic methods," *Medical Image Analysis*, vol. 97, p. 103223, Jun. 2024, doi: 10.1016/j.media.2024.103223.
- [2] M. Abdar *et al.*, "A review of uncertainty quantification in deep learning: Techniques, applications and challenges," *Information Fusion*, vol. 76, pp. 243–297, May 2021, doi: 10.1016/j.inffus.2021.05.008.
- [3] S. Seoni, V. Jahmunah, M. Salvi, P. D. Barua, F. Molinari, and U. R. Acharya, "Application of uncertainty quantification to artificial intelligence in healthcare: A review of last decade (2013–2023)," *Computers in Biology and Medicine*, vol. 165, p. 107441, Sep. 2023, doi: 10.1016/j.combiomed.2023.107441.
- [4] E. Hüllermeier and W. Waegeman, "Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods," *Machine Learning*, vol. 110, no. 3, pp. 457–506, Mar. 2021, doi: 10.1007/s10994-021-05946-3.
- [5] M. Barandas *et al.*, "Evaluation of uncertainty quantification methods in multi-label classification: A case study with automatic diagnosis of electrocardiogram," *Information Fusion*, vol. 101, p. 101978, Aug. 2023, doi: 10.1016/j.inffus.2023.101978.
- [6] R. Senge *et al.*, "Reliable classification: Learning classifiers that distinguish aleatoric and epistemic uncertainty," *Information Sciences*, vol. 255, pp. 16–29, Aug. 2013, doi: 10.1016/j.ins.2013.07.030.
- [7] M. S. Arif, A. Mukheimer, and D. Asif, "Enhancing the early detection of chronic kidney Disease: a robust machine learning model," *Big Data and Cognitive Computing*, vol. 7, no. 3, p. 144, Aug. 2023, doi: 10.3390/bdcc7030144.
- [8] M. Barandas, D. Folgado, R. Santos, R. Simão, and H. Gamboa, "Uncertainty-Based rejection in Machine

- Learning: Implications for model development and Interpretability,” *Electronics*, vol. 11, no. 3, p. 396, Jan. 2022, doi: 10.3390/electronics11030396.
- [9] S. Asif *et al.*, “Advancements and Prospects of Machine Learning in Medical Diagnostics: Unveiling the future of Diagnostic Precision,” *Archives of Computational Methods in Engineering*, Jun. 2024, doi: 10.1007/s11831-024-10148-w.
- [10] Z. Guo *et al.*, “A survey on uncertainty reasoning and quantification in belief theory and its application to deep learning,” *Information Fusion*, vol. 101, p. 101987, Aug. 2023, doi: 10.1016/j.inffus.2023.101987.
- [11] B. X. Yong and A. Brintrup, “Bayesian autoencoders with uncertainty quantification: Towards trustworthy anomaly detection,” *Expert Systems With Applications*, vol. 209, p. 118196, Jul. 2022, doi: 10.1016/j.eswa.2022.118196.
- [12] H. Salem, M. Y. Shams, O. M. Elzeki, M. A. Elfattah, J. F. Al-Amri, and S. Elnazer, “Fine-Tuning fuzzy KNN classifier based on uncertainty membership for the medical diagnosis of diabetes,” *Applied Sciences*, vol. 12, no. 3, p. 950, Jan. 2022, doi: 10.3390/app12030950.
- [13] Thelen *et al.*, “A comprehensive review of digital twin—part 2: roles of uncertainty quantification and optimization, a battery digital twin, and perspectives,” *Structural and Multidisciplinary Optimization*, vol. 66, no. 1, Dec. 2022, doi: 10.1007/s00158-022-03410-x.
- [14] W. V. Padula *et al.*, “Machine Learning Methods in Health Economics and Outcomes Research—The PALISADE Checklist: A Good Practices Report of an ISPOR Task Force,” *Value in Health*, vol. 25, no. 7, pp. 1063–1080, Jun. 2022, doi: 10.1016/j.jval.2022.03.022.
- [15] D. Li, Z. Qi, Y. Zhou, and M. Elchalakani, “Machine Learning Applications in Building Energy Systems: Review and Prospects,” *Buildings*, vol. 15, no. 4, p. 648, Feb. 2025, doi: 10.3390/buildings15040648.
- [16] Ellouze, O. Kahouli, M. Ksantini, H. Alsaif, A. Aloui, and B. Kahouli, “Artificial Intelligence-Based Diabetes Diagnosis with Belief Functions Theory,” *Symmetry*, vol. 14, no. 10, p. 2197, Oct. 2022, doi: 10.3390/sym14102197.
- [17] M. A. Abd-Elrazek, A. A. Eltahawi, M. H. A. Elaziz, and M. N. Abd-Elwhab, “Predicting length of stay in hospitals intensive care unit using general admission features,” *Ain Shams Engineering Journal*, vol. 12, no. 4, pp. 3691–3702, Apr. 2021, doi: 10.1016/j.asej.2021.02.018
-