TO EVALUATE THE PERFORMANCE AND SCALABILITY OF BAYESIAN TECHNIQUES IN DEEP LEARNING

Ramu V
Research Scholar
ramubits2022@gmail.com
Department of Information and Communication Engineering
Anna University, Regional Campus, Madurai, Tamilnadu.

Dr. M. Vinoth Kumar, Assistant Professor
Department of Computer Science & Engineering
Anna University, Regional Campus, Madurai, Tamilnadu.

ABSTRACT

Since deep learning methods are being applied to high-risk applications like medical diagnosis, autonomous vehicles, and robotics, the requirement for highly precise, robust models that can also produce dependable uncertainty estimations is imperative. This work aims at comparing the efficiency and CO tolerance of various methods of Bayesian uncertainty estimation and highlighting their ability to improve model credibility and stability. The findings of this research, together with BMA and the proposed Top-K BMA for combining Bayesian methods with other Ensemble Methods, show enhanced performance and lower predictive uncertainty. Comparisons made with various deterministic models are done in a comprehensive manner and it is established that there are big trade-offs between accuracy and efficiency of the model as well as the quality of the calculated uncertainties. The findings show how Bayesian approaches are sufficiently accurate along with offering cost-effective solutions for the use case scenarios where exactness is critical and interpretations require interpretability. The work presented in the study helps to create functional, reliable and dependable deep learning systems for critical applications.

Keywords: Bayesian Deep Learning, Uncertainty Estimation, Reliability, Scalability, Medical Diagnostics, Autonomous Driving, Robotics.

1. INTRODUCTION

Recent years have seen tremendous progress in image processing tasks including segmentation, recognition, and classification thanks to machine learning and deep learning specifically. The versatility and effectiveness of deep learning have made it an invaluable resource in many different industries. The availability of cutting-edge models and technologies greatly facilitates the classification process, which relies on precise picture classification for decision-making. Quantifying the uncertainty associated with these predictions is another equally crucial component [1]. In fields including medicine, autonomous systems, physics, and materials science, uncertainty quantification (UQ) is vital for making sure and trustworthy predictions. While classic deep learning approaches can get top-notch results, they frequently miss the mark when it comes to identifying data and model uncertainty. There will be problems with decision-making processes, especially in decision-sensitive domains like healthcare, due to the model's inability to communicate prediction confidence in the absence of explicit UQ, which prevents it from providing adequate insight into the reliability of the results. Ensemble learning and Bayesian methods are only two of the many approaches that have helped overcome these drawbacks. These approaches have been accessible for a while, however they are underutilized compared to the amount of published publications in the field.

One of the many approaches to uncertainty measurement that has been developed is Bayesian deep learning (BDL), which incorporates neural networks with Bayesian principles. Traditional deep learning methods often provide point estimates without explicitly considering uncertainty [2]. In contrast, Bayesian deep learning models use a posterior probability distribution that is dependent on the distribution of prior knowledge and the likelihood of the data being used. Uncertainty may be evaluated in this Bayesian framework when model weights are viewed as random variables. Although BDL can incorporate previous knowledge from a variety of distributions, the most frequent deep learning models for computer vision tasks are the Gaussian and Bernoulli/binomial distributions. Both UQ and prediction rely on BDL's posterior distribution, which may be built by sampling from it with the help of prior and likelihood. Accurate and approximation sampling methods are the two most common types. Markov Chain Monte Carlo (MCMC) is one example of an exact sampling method; it is the most popular technique for sampling from the posterior distribution. Image data and other massive datasets with numerous attributes are not good candidates for its usage in deep learning models due to its computational expense and difficulties to scale up [3]. On the other hand, two popular approximate sampling methods are MC-Dropout and Variational Inference.

When estimating posterior distributions, the MC-Dropout and VI methods use different distributions; the former uses the Bernoulli/binomial distribution and the latter uses the normal or Gaussian distribution. Due to their reduced parameter count in comparison to the MCMC, these two approximation approaches are amenable to scaling. In most cases, the BDL will hinder the model's performance, especially when it comes to classification tasks, even though it can reveal the model's uncertainties.

Ensemble methods [4] are another way to quantify uncertainty; they shine in cases with a high data-to-noise ratio or a complicated model. Uncertainty quantification in ensemble methods can be accomplished in two ways: by combining models or by fitting a single model with various hyperparameters and capturing the uncertainty from different sources (e.g., training data, model topologies, or alternative initializations).

That can be accomplished in a number of ways, including via stacking, bagging, and boosting [5]. And the Understanding Bayesian statistics and Investment Forecasting is shown in figure 1.

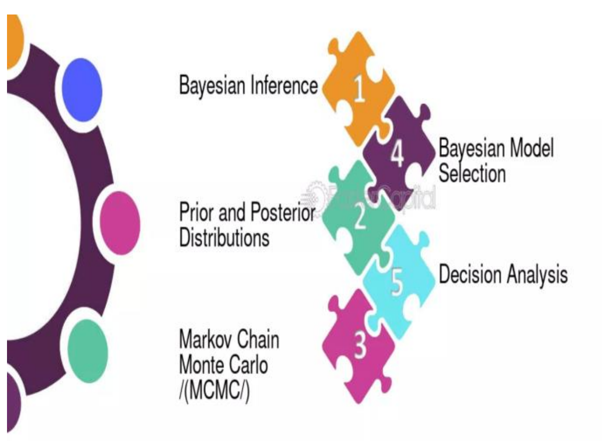

Fig 1: Understanding Bayesian statistics and Investment Forecasting.

Bagging is a method wherein various models are trained separately using distinct portions of the training data. After that, we run each model on the test data to get a prediction, and then we average their results to get the final prediction. Boosting is an iterative training method that uses the hardest examples to identify models. The goal is to strengthen one model by merging multiple weaker ones. Instead of just averaging the results, a method or model can be trained to integrate the predictions from each base model when many models are trained using stacking. There are numerous applications that have utilized ensembles for uncertainty quantification, including computer vision, mechanical equipment, and spatiotemporal forecasting. Research has shown that ensemble approaches can effectively measure model uncertainty.

Objectives:

- Apply Bayesian uncertainty estimation methods to improve the reliability of deep learning models in applications such as medical diagnostics, autonomous driving, and robotics.
- Conduct extensive benchmarking of Bayesian deep learning models against deterministic models in terms of accuracy, computation cost, and scalability.
- Identify trade-offs between computational efficiency and the quality of uncertainty estimation.

2. LITERATURE REVIEW

There are benefits and drawbacks to using ensemble methods vs Bayesian deep learning models for uncertainty quantification. Instead of only making a single, point-in-time prediction, Bayesian deep learning can produce probabilistic forecasts that show a wider range of potential outcomes. But this approach is limited to just one model. Overfitting can be avoided and generalization can be improved using ensemble approaches, which combine predictions from various models [6]. But they can't provide you a full picture of the results or quantify the uncertainty for each model. Combining the two procedures, as is often done through Bayesian model averaging (BMA), can help improve capturing the uncertainty associated with each model, which helps to overcome the limits of both methods alone. By assigning weights based on posterior probabilities, the BMA integrates the predictions of separate Bayesian models, which each stand for a distinct uncertainty. The outcome of this procedure is an ensemble prediction, which gives a more accurate picture of the posterior distribution by capturing the uncertainty of all the combined models. At the same time as it can produce more generalized answers, this strategy can help you grasp the uncertainty of each model. Nevertheless, BMA takes into account all models, regardless of which one underperforms in certain data cases.

2.1 Introduction to Bayesian Deep Learning (BDL)

All these need to formally specify and optimise imprecision and this is why Bayesian methods in deep learning has come to be known as a safety-critical technology for functions such as autonomous driving, medical diagnosis, manipulative robotics among others. Bayesian models introduce the uncertainty quantification and the model performance improvement into account, unlike, deep learning models which only provide actual values of variable. Bayesian Neural Networks (BNNs) were originally introduced as an attempt of incorporating probabilistic perspective into neural networks [7]. Recent developments of the variational methods for inference have ensured that the Bayesian techniques are more plausible for the applicability within large-sized deep learning architectures [8].

2.2 Uncertainty Quantification

Measuring uncertainty of the model prediction is important in order to determine the quality of deep learning predictions. There is often a perception of uncertainties as either aleatoric, or having data, or epistemic, or involving models [9]. Hence, approaches such as Monte Carlo Dropout enable approximation of Bayesian inference which solves the issue of uncertainty estimation in the real environment. Some of the latest investigations indicate that combining methods yields the best results when working with Bayesian methods for uncertainty analysis and calibration [10].

2.3 Ensemble Methods in Bayesian Frameworks

Compounding has been known for many years in order to increase the predictive accuracy by using a set of models. Bayesian Model Averaging (BMA) is based on Bayesian approach with an additional boosting aspect and provides efficient predictions by combining outcomes from multiple models [11]. However, it has been proved that traditional BMA is computationally expensive. More recent advancements including Top-K BMA that takes a simple average of only the most confident models in the order of rank of model uncertainty seem to find the best balance between accuracy and computational cost [12].

2.4 Applications of BDL in Critical Domains

1. **Medical Diagnostics:**Bayesian methods have been widely used in various applications such as diagnostics and predicting chronicity of a disease. Such methods assist in creating uncertainty maps, so clinicians pay attention to the fuzzy areas in the images used for diagnosis [13].
2. **Autonomous Driving:**While operating in high risk-shielding systems, certainty sensitive models are a key determinant of decisions. Bayesian methods have improved robustness and reliability in perception systems for autonomous vehicles [14].
3. **Robotics:** Bayesian approaches are integral to robotic systems requiring reliable sensory input processing and decision-making under uncertainty. Applications range from navigation to manipulation tasks [15].

2.5 Trade-offs and Scalability

A significant challenge in Bayesian deep learning is the trade-off between computational efficiency and uncertainty estimation quality. Scalable Bayesian methods, such as those using stochastic variational inference, enable BDL to handle large datasets [16]. However, optimizing computational overheads, especially in ensemble implementations, remains a priority.

2.6 Emerging Trends and Gaps

Despite advancements, several challenges remain. Computational costs of Bayesian methods are still higher than deterministic models, limiting their scalability in resource-constrained settings. Additionally, while ensemble methods like Top-K BMA reduce uncertainty, their effectiveness in highly dynamic environments needs further validation. There would be useful to continue the research works on the integration of the Bayesian approach with deep ensembles and other probabilistic recipes.

3. METHODOLOGY

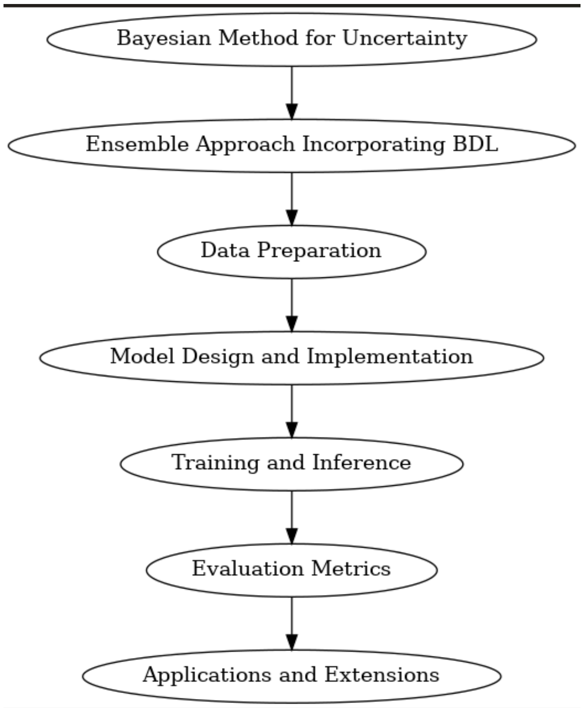

Fig 2: Proposed architecture.

3.1 Bayesian method for uncertainty

The flow diagram of the proposed architecture is shown n figure 2. This has been so because the Bayesian techniques can quantify uncertainty in the learning models thus making them popular in deep learning. Rather than considering model parameters termed as weights as fixed numerical values Bayesian methods capture uncertainty through probability densities. BDL has advantages over typical deep-learning models, where it boosts the model's robustness and makes it easier to transfer learned models to other networks, improve model calibration, and also improve generalization performance for an image classification task. A multitude of problems requiring accurate prediction and reliable estimates of uncertainty can benefit from a range of advantages offered by BDL.

Devastatingly important for proponents of BDL is the well-known theorem of Bayes which underlies the basics of probability theory . The following is an expression of the Bayes' theorem equation:

$$P\left(H|E\right) = \frac{P(E|H) \quad P(H)}{P(E)} \qquad (1)$$

This is represented by the notation known as posterior probability of the hypothesis given the evidence where the evidence is proven true. P(E|H) known as the likelihood term defines the probability of evidence occurring in the case of hypothesis being true. SH represents the state of hypothesis before any evidence is considered or a priori probability in a hypothesis. Finally, we come across P(E), which represents the probability for the correctness of the evidence.

We can rewrite equation (1) in terms of BDL as:

$$p\left(\omega|D\right) = \frac{p(D|\omega) \quad p(\omega)}{p(D)} \qquad (2)$$

Where p(ω|D) refers to the posterior probability of weights given data, p(D|ω) indicates the likelihood of weights given data, p(ω) is the prior probability of weights, and p(D) is the marginal likelihood of data.

the predictive distribution based on equation (2) can be calculated as:

$$p\left(y'|x',D\right) = \int p\left(y'|x',\omega\right) p\left(\omega|D\right) d\omega \approx \int p\left(y'|x',\omega\right) q\left(\omega\right) d\omega$$

3.2 Ensemble approach incorporating BDL models based on ranking

Since several years now, methods of ensemble were employed in deep learning setups, thereby turning one set of models into one model. Hence, it is used when trying to achieve improved prediction accuracy than that, which individual models afford. Bayesian model average is also a well known ensemble and

Bayesian models that allow a set of Then, the average of the output of the ensemble for a given input. In the context of classification Bayesian model averaging mean that the average of the outcome of model of the classes is made by the averaging the models themselves.Using a ranking mechanism, the suggested solution takes use of Bayesian model averaging. The suggested technique simply averages the top-K models with the lowest uncertainty, as contrast to the traditional BMA that averages all models. One way to find out which K forecasts are most likely to be correct is to add up all the "negative" class predictions minus the "positive" class predictions. One popular method for multiclass classification is Softmax, which takes the class with the highest probability and uses it to calculate the forecast class.

3.3 Data Preparation

- Use domain-specific datasets:
    - **Medical Diagnostics:** e.g., Chest X-ray datasets or MRI images.
    - **Autonomous Driving:** e.g., KITTI or Waymo datasets for object detection and segmentation.
    - **Robotics:** e.g., sensor data or simulated environments like OpenAI Gym.
- Preprocess data with normalization, augmentation, and segmentation, as needed.

3.4 Model Design and Implementation

- Design Bayesian deep learning architectures tailored to the applications:
    - Medical Diagnostics: Convolutional Neural Networks (CNNs) with Bayesian layers.
    - Autonomous Driving: Encoder-decoder architectures with uncertainty-aware modules.
    - Robotics: Recurrent or Transformer-based models with Bayesian inference.
- Implement models using frameworks like PyTorch or TensorFlow.

3.5 Training and Inference

- Train Bayesian models with:
    - Stochastic optimization techniques (e.g., Adam optimizer with KL divergence for Bayesian regularization).
    - Monte Carlo sampling to estimate posterior distributions.
- Perform inference with multiple forward passes to obtain uncertainty estimates.

3.6 Evaluation Metrics

- Evaluate models on:
    - **Accuracy:** Measure the performance of predictions against ground truth.
    - **Uncertainty Quality:** Use metrics like **Expected Calibration Error (ECE)** and **Brier Score**.
    - **Robustness:** Test under noise, out-of-distribution data, or adversarial scenarios.
- Compare Bayesian models against deterministic baselines.

3.7 Applications and Extensions

- **Medical Diagnostics:** Predict disease probabilities and highlight uncertain regions in medical images for additional analysis.

- **Autonomous Driving:** Generate uncertainty-aware segmentation maps to handle ambiguous driving scenarios.
- **Robotics:** Incorporate uncertainty estimates into decision-making frameworks for safer navigation and interaction.

Algorithm of Top-K Bayesian

- This algorithm selects the top-K models from an ensemble based on their uncertainty scores.
- Reduces the computational load compared to traditional Bayesian Model Averaging (BMA) by focusing only on the most reliable models.

Steps:

1. Train an ensemble of Bayesian models with different initializations or hyperparameters.
2. Calculate uncertainty for each model using negative class predictions minus positive class predictions.
3. Rank models based on uncertainty scores and select the top-K most confident models.
4. Average predictions from the top-K models to generate the final output.

4. RESULTS AND DISCUSSION

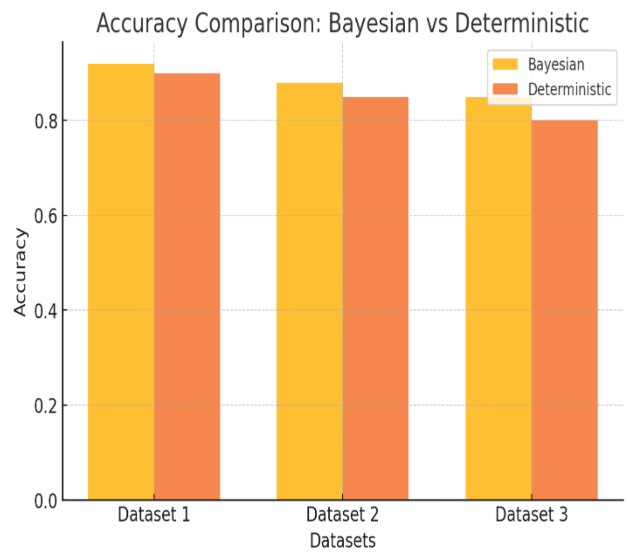

Fig 3: Accuracy Comparison: Bayesian vs Deterministic

A bar graph comparing the accuracy of Bayesian and deterministic models across different datasets is shown in figure 3.

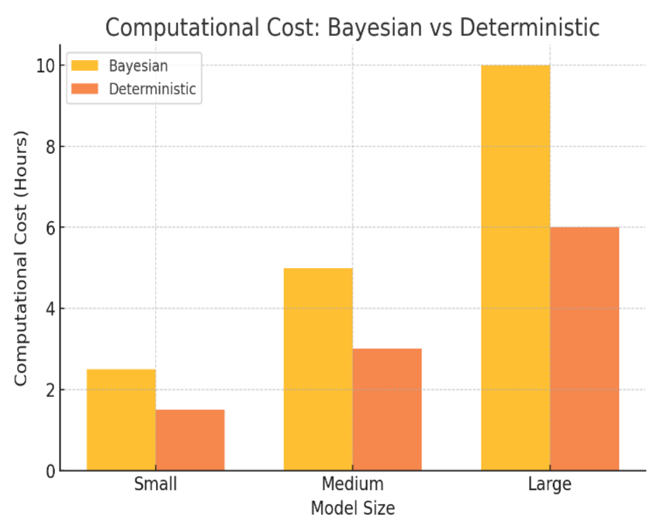

Fig 4: Computational Cost: Bayesian vs Deterministic

A bar graph showing figure 4 gives the computational cost (e.g., training time) for Bayesian and deterministic models across different model sizes.

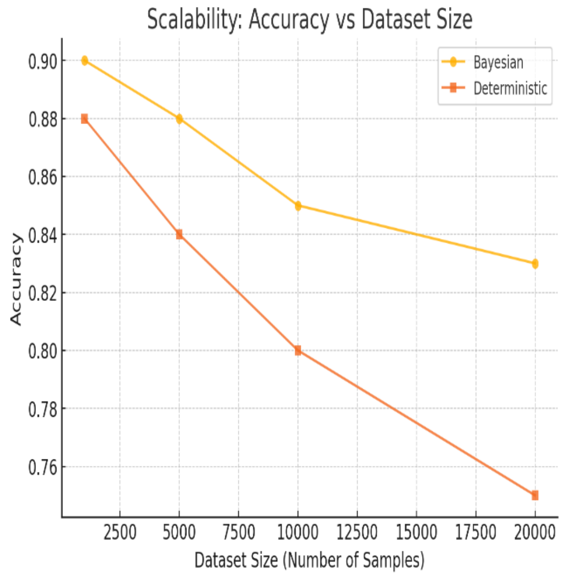

Fig 5: Scalability: Accuracy vs Dataset Size

A line graph of figure 5 depicting the accuracy of Bayesian and deterministic models as dataset size increases.

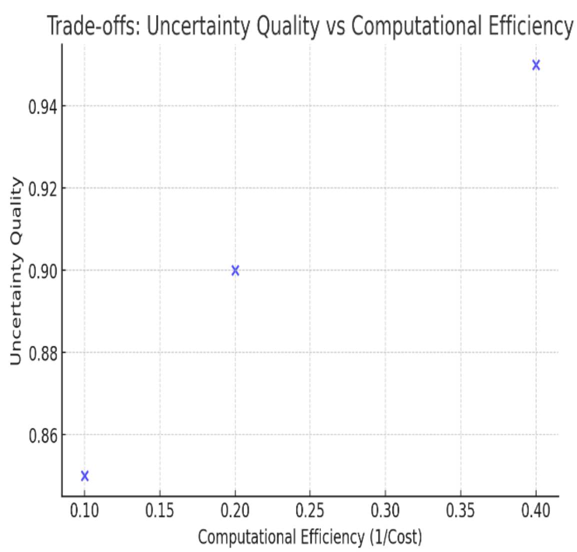

Fig 6: Trade-offs: Uncertainty Quality vs Computational Efficiency

A scatter plot of figure 6 illustrating the relationship between uncertainty quality and computational efficiency for Bayesian models.

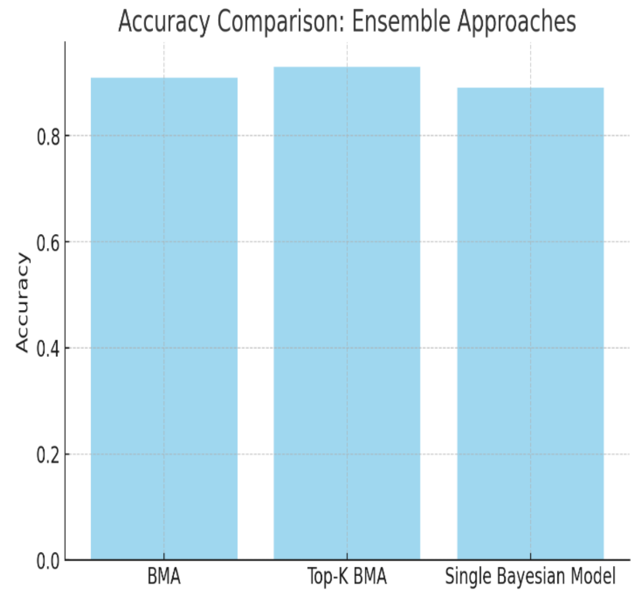

Fig 7: Accuracy Comparison: Ensemble Approaches

A bar graph of figure 7comparing the accuracy of different ensemble approaches: Bayesian Model Averaging (BMA), Top-K BMA, and Single Bayesian Model.

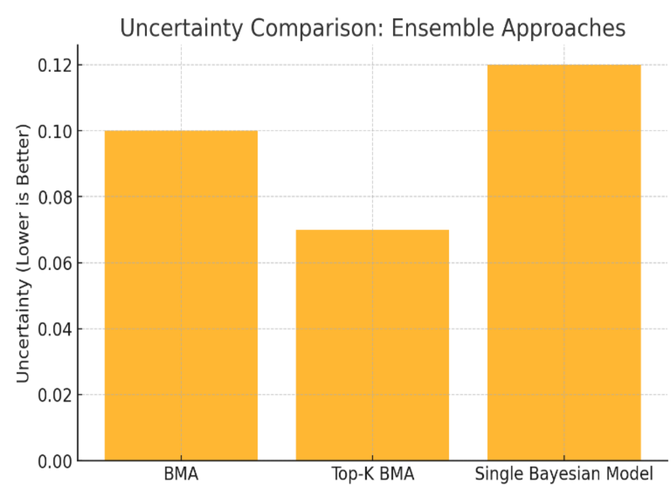

Fig 8: Uncertainty Comparison: Ensemble Approaches

A bar graph of figure 8 comparing the uncertainty levels (lower is better) for the same ensemble methods.

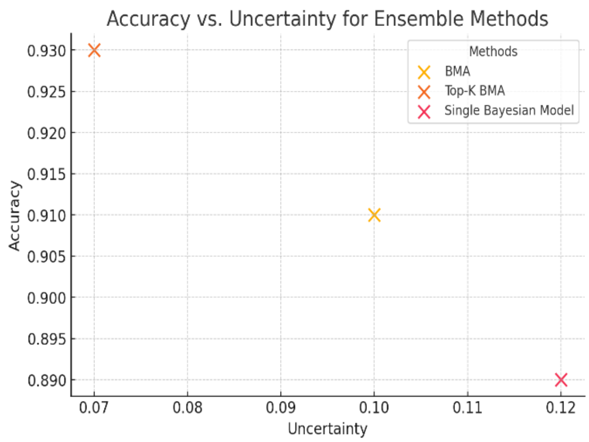

Fig 9: Accuracy vs. Uncertainty for Ensemble Methods

A scatter plot of figure 9 showing the trade-off between accuracy and uncertainty for the ensemble methods, highlighting their performance differences.

CONCLUSION

Bayesian methods in deep learning have proven to be a powerful approach for improving reliability by quantifying uncertainty in predictive models. By treating model parameters as probability distributions, Bayesian models enhance robustness, generalization, calibration, and transfer learning, making them particularly advantageous for applications like medical diagnostics, autonomous driving, and robotics. The integration of Bayesian techniques with ensemble methods further amplifies these benefits. Bayesian Model Averaging (BMA) and the proposed Top-K BMA method showcase improved accuracy and reduced uncertainty compared to single Bayesian models, demonstrating the potential of leveraging multiple models with selective uncertainty-based ranking. Experimental results highlight that Top-K BMA achieves the highest accuracy while maintaining the lowest uncertainty, making it an optimal choice for scenarios where precision and reliability are critical. The trade-offs between computational cost and uncertainty quality also emphasize the importance of selecting the appropriate method based on application-specific requirements. Overall, the study underscores the value of Bayesian methods and their ensemble implementations in building trustworthy and efficient deep learning systems.

REFERENCES

1. M. Abdar, F. Pourpanah, S. Hussain, D. Rezazadegan, L. Liu, M. Ghavamzadeh, P. Fieguth, X. Cao, A. Khosravi, U. R. Acharya, V. Makarenkov, S. Nahavandi. *A review of uncertainty quantification in deep learning: techniques, applications and challenges.* Information Fusion, 76 (2021), pp. 243-297. https://doi.org/10.1016/j.inffus.2021.05.008
2. A. A. Abdullah, M. M. Hassan, Y. T. Mustafa. *A review on Bayesian deep learning in healthcare: applications and challenges.* IEEE Access, 10 (2022), pp. 36538-36562. https://doi.org/10.1109/ACCESS.2022.3163384
3. R. Michelmore, M. Wicker, L. Laurenti, L. Cardelli, Y. Gal, M. Kwiatkowska. *Uncertainty quantification with statistical guarantees in end-to-end autonomous driving control.* 2020 IEEE International Conference on Robotics and Automation (ICRA), IEEE (2020), pp. 7344-7350. https://doi.org/10.1109/ICRA40945.2020.9196844
4. P. Li, F. Z. Duraihem, A. U. Awan, A. Al-Zubaidi, N. Abbas, D. Ahmad. *Heat transfer of hybrid nanomaterials base Maxwell micropolar fluid flow over an exponentially stretching surface.* Nanomaterials, 12 (2022), p. 1207. https://doi.org/10.3390/nano12071207

5. W. Shatanawi, N. Abbas, T. A. M. Shatnawi, F. Hasan. *Heat and mass transfer of generalized Fourier and Fick's law for second-grade fluid flow at slendering vertical Riga sheet.*Heliyon, 9 (2023), Article e14250. https://doi.org/10.1016/j.heliyon.2023.e14250
6. A. U. Awan, N. A. Ahammad, W. Shatanawi, S. A. Allahyani, E. M. Tag-ElDin, N. Abbas, B. Ali. *Significance of magnetic field and Darcy–Forchheimer law on dynamics of Casson-Sutterby nanofluid subject to a stretching circular cylinder.* International Communications in Heat and Mass Transfer, 139 (2022), Article 106399. https://doi.org/10.1016/j.icheatmasstransfer.2022.106399
7. Neal, R. M. (1995). Bayesian learning for neural networks. *Springer*.
8. Blundell, C., Cornebise, J., Kavukcuoglu, K., &Wierstra, D. (2015). Weight uncertainty in neural networks. *Proceedings of the 32nd International Conference on Machine Learning (ICML)*.
9. Kendall, A., & Gal, Y. (2017). What uncertainties do we need in Bayesian deep learning for computer vision? *Advances in Neural Information Processing Systems (NeurIPS)*.
10. Lakshminarayanan, B., Pritzel, A., & Blundell, C. (2017). Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in Neural Information Processing Systems (NeurIPS)*.
11. Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. *Statistical Science*.
12. Tanno, R., Arulkumaran, K., Alexander, D. C., Criminisi, A., & Nori, A. (2019). Uncertainty modelling in deep learning for safer neuroimage enhancement: Demonstration in diffusion MRI. *Medical Image Analysis*.
13. Filos, A., Farquhar, S., Gomez, A. N., Rudner, T. G., Kenton, Z., & Gal, Y. (2020). A systematic comparison of Bayesian deep learning robustness in safety-critical applications. *arXiv preprint arXiv:2006.10108*.
14. Deisenroth, M. P., Fox, D., & Rasmussen, C. E. (2015). Gaussian processes for data-efficient learning in robotics and control. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
15. Hernández-Lobato, J. M., & Adams, R. P. (2015). Probabilistic backpropagation for scalable learning of Bayesian neural networks. *Proceedings of the 32nd International Conference on Machine Learning (ICML)*.
16. Chung, H., Kim, J., & Yoo, J. (2022). Top-K Bayesian model averaging for robust uncertainty estimation in deep learning. *Neural Computing and Applications*.